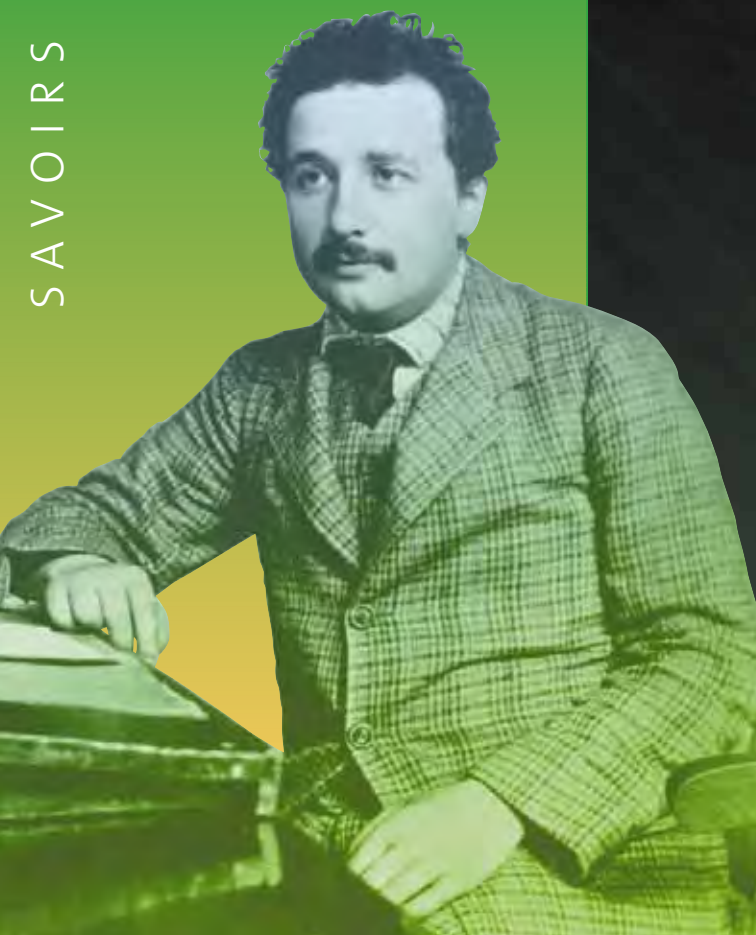


PHYSIQUE

# Einstein aujourd'hui

SAVOIRS ACTUELS



*Alain ASPECT*

*François BOUCHET*

*Éric BRUNET*

*Claude COHEN-TANNOUJI*

*Jean DALIBARD*

*Thibault DAMOUR*

*Olivier DARRIGOL*

*Bernard DERRIDA*

*Philippe GRANGIER*

*Franck LALOË*

*Jean-Paul POCHOLLE*

 CNRS EDITIONS

  
EDP  
SCIENCES



# Einstein aujourd'hui

S A V O I R S A C T U E L S

EDP Sciences/CNRS ÉDITIONS

*Illustration de couverture* : Photographie d'A. Einstein (1905). © Hebrew University of Jerusalem. Archives, California Institute of Technology.

© 2005, **EDP Sciences**, 17, avenue du Hoggar, BP 112, Parc d'activités de Courtabœuf, 91944 Les Ulis Cedex A

et

**CNRS ÉDITIONS**, 15, rue Malebranche, 75005 Paris.

Tous droits de traduction, d'adaptation et de reproduction par tous procédés réservés pour tous pays. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (art. L. 122-4, L. 122-5 et L. 335-2 du Code de la propriété intellectuelle). Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre français d'exploitation du droit de copie, 3, rue Hautefeuille, 75006 Paris. Tél. : 01 43 26 95 35.

**ISBN** EDP Sciences 2-86883-768-9

**ISBN** CNRS ÉDITIONS 2-271-06311-6

# Avant-propos

Par *Michèle Leduc et Michel Le Bellac*

EN 1905, Einstein publiait une série d'articles fondateurs sur la relativité, les quanta lumineux et le mouvement brownien. Un siècle après cet « *annus mirabilis* », il nous a paru intéressant de demander à quelques physiciens français de renom de donner leur point de vue sur l'apport d'Einstein à la physique de ce début du XXI<sup>e</sup> siècle. Nous avons choisi de ne pas nous limiter à l'année 1905, mais nous avons pris en compte l'ensemble de l'œuvre scientifique d'Einstein jusqu'au milieu des années 1930. En effet ses derniers travaux passés à la postérité sont d'une part celui sur les fondements de la mécanique quantique, effectué en collaboration avec Podolsky et Rosen et connu sous les initiales « EPR » (1935), et d'autre part celui sur les lentilles gravitationnelles (1936). Le panorama dressé dans ce livre permet de juger de la diversité impressionnante des sujets abordés par Einstein. Parmi les plus marquants, citons :

*Quanta* : quanta lumineux (1905) ; chaleur spécifique des solides (1907) ; émission stimulée de rayonnement (1917) ; condensation de Bose-Einstein (1923).

*Relativité* : relativité restreinte (1905) ; énoncé du principe d'équivalence (1907) ; relativité générale (1916).

*Fluctuations* : mouvement brownien (1905) ; théorie des fluctuations thermodynamiques (1910).

*Fondements de la mécanique quantique* : article EPR, débats avec Niels Bohr (1935).

*Cosmologie* : principe cosmologique et constante cosmologique (1917) ; modèle d'Univers en expansion d'Einstein et de Sitter (1932).

Ces travaux ont non seulement ouvert la voie à une bonne partie de la physique du XX<sup>e</sup> siècle, mais ils continuent encore aujourd'hui à avoir un impact dans des domaines parfois inattendus : ainsi il a fallu attendre près de 50 ans, avec les travaux de John Bell et les expériences qui ont suivi, pour que soit reconnue l'importance de l'intrication des systèmes quantiques. Ce n'est qu'en 1995 qu'a été vérifiée sans ambiguïté la prédiction de la condensation

de Bose-Einstein. Et Einstein n'avait sûrement pas envisagé que la relativité serait indispensable pour notre système de navigation GPS, un outil qui fait partie maintenant de notre vie quotidienne.

Ce livre contient sept contributions. La première est une introduction historique écrite par Olivier Darrigol, qui décrit en détail la genèse des articles fondateurs de 1905, et plus succinctement celle des travaux ultérieurs. On constatera combien Einstein, loin de l'image du savant solitaire isolé dans son bureau des brevets à Zürich, était parfaitement au courant des grandes interrogations de la physique de son époque et des travaux de ses grands contemporains (Lorentz, Poincaré, Planck, Mach, ...) tout en « les dépassant par l'ampleur et la radicalité de ses points de vue », comme le souligne Olivier Darrigol.

Les trois contributions suivantes concernent la physique quantique, en suivant l'ordre historique inversé des articles fondateurs. Alain Aspect et Philippe Grangier montrent comment Einstein, Podolsky et Rosen (EPR), parallèlement à Erwin Schrödinger, avaient mis en lumière une particularité unique de la mécanique quantique, qui est largement connue aujourd'hui sous le nom d'intrication des systèmes quantiques. Einstein, très opposé à l'interprétation de Copenhague de la mécanique quantique, s'était appuyé sur l'intrication afin d'essayer de prouver que la mécanique quantique constituait une description incomplète de la réalité physique. Niels Bohr avait jusque là répondu brillamment aux objections antérieures d'Einstein, mais il ne put donner cette fois qu'une réponse alambiquée. Ce débat de géants tomba dans l'oubli pendant une trentaine d'années, et il fallut attendre les travaux de John Bell en 1964 pour que l'article EPR revienne sur le devant de la scène. En effet, John Bell put montrer que les idées contenues dans cet article étaient susceptibles de vérification expérimentale. Alain Aspect et Philippe Grangier décrivent les expériences récentes qui ont permis d'invalider la conclusion immédiate du travail EPR (il faut compléter la mécanique quantique), tout en montrant en revanche combien la notion d'intrication est fondamentale : jointe aux progrès expérimentaux qui permettent de manipuler des objets quantiques individuels, elle a donné naissance à une « nouvelle révolution quantique », dont un exemple est l'information quantique.

Dans l'article suivant, Claude Cohen-Tannoudji, Jean Dalibard et Frank Laloë examinent la postérité du travail de 1923, qui prédisait le phénomène appelé aujourd'hui « condensation de Bose-Einstein », le fait que des atomes d'un certain type (obéissant à la statistique de Bose) ont tendance à s'accumuler dans un état quantique unique à très basse température. Einstein lui-même était assez sceptique sur la possibilité de mettre cet effet en évidence. Depuis les premières vérifications expérimentales de la condensation de Bose-Einstein en 1995, on assiste à une explosion de travaux sur ce nouvel état de la matière. L'intérêt du phénomène vient de la relation fascinante et controversée entre superfluidité et condensation de Bose-Einstein, et aussi des perspectives de transposer toute l'optique aux ondes de matière, et de fabriquer des « lasers à atomes ». Enfin les condensats de Bose-Einstein atomiques gazeux (et les gaz

d'atomes fermioniques à très basse température) permettent de réaliser des systèmes modèles qui servent à tester des idées de la physique de la matière condensée dans des conditions parfaitement contrôlées.

Dans la dernière contribution à cette série « quantique », Jean-Paul Pocholle fait le point sur ce qui est sans doute l'application la plus largement connue des travaux de 1917, le laser. L'apport conceptuel d'Einstein à la théorie de l'interaction lumière-matière est essentiel. Il a en effet introduit la notion d'émission stimulée de rayonnement, qui est à l'origine du processus d'émission laser ; celle-ci a vu le jour dans les laboratoires à la fin des années 1950, lorsque les physiciens ont été capables de réaliser des inversions de population de niveaux atomiques ou moléculaires. Depuis lors, la technologie du laser a enregistré des progrès spectaculaires, allant des diodes laser à semi-conducteur de taille micrométrique jusqu'aux sources laser à verre dopé surpuissantes pour la fusion contrôlée. Les applications des lasers sont aujourd'hui trop nombreuses pour être recensées de façon exhaustive ; ils sont devenus les compagnons de notre vie quotidienne tout autant que des outils incontournables pour la recherche.

L'apport d'Einstein à la théorie du mouvement brownien et aux fluctuations statistiques est sans doute moins connu que ses contributions à la physique quantique et à la relativité. Elle n'en est pas moins essentielle, car il fut le premier à saisir le lien quantitatif entre les fluctuations d'origine microscopique et les propriétés thermodynamiques, une idée à la base de la physique statistique moderne. Comme le montre Olivier Darrigol, sa compréhension profonde des fluctuations constitue aussi le fil directeur de ses travaux sur la physique quantique. Dans leur chapitre, Bernard Derrida et Eric Brunet donnent un aperçu de la variété des problèmes de physique allant de la théorie du mouvement brownien à celle des polymères, du théorème fluctuation dissipation aux systèmes hors équilibre, qui sont dans le prolongement des travaux d'Einstein.

Les deux derniers chapitres, par Thibault Damour et François Bouchet, sont des synthèses très documentées et actualisées sur la relativité générale et la cosmologie. Après la relativité restreinte en 1905, il fallut à Einstein une dizaine d'années de travail acharné pour mettre au point sa théorie relativiste de la gravitation, ou relativité générale. Cette extraordinaire construction intellectuelle connut des débuts très médiatisés en raison de l'expédition d'Eddington confirmant la déviation des rayons lumineux par le Soleil. Elle fut cependant assez largement ignorée des physiciens jusqu'au début des années 1960, en raison de ses difficultés mathématiques et du caractère limité des tests expérimentaux envisageables à l'époque. Comme le montre Thibault Damour, la perception de la relativité générale est totalement différente aujourd'hui. Elle est vérifiée dans un grand nombre de situations, qui vont de notre environnement immédiat (système GPS) jusqu'aux confins de l'Univers, avec une précision relative qui peut atteindre  $10^{-5}$ . Elle est devenue une théorie indispensable de l'astrophysique contemporaine : elle est à la base de notre

compréhension des étoiles massives, des trous noirs, des lentilles gravitationnelles et des ondes gravitationnelles.

La relativité générale est au fondement de la cosmologie moderne. C'est ce qu'Einstein avait compris dès 1917, comme en témoigne la saga de la constante cosmologique (notée  $\Lambda$ ), qu'il introduisit dans un premier temps afin de rendre compte d'un Univers supposé statique, pour la renier quinze ans plus tard quand s'imposa l'idée de l'expansion de l'Univers. François Bouchet décrit l'évolution fulgurante de la cosmologie depuis dix ans, qui aboutit aujourd'hui au modèle standard dit  $\Lambda$ CDM, où figurent  $\Lambda$  parce que la constante cosmologique a finalement refait surface et CDM, initiales de Cold Dark Matter, la matière sombre froide. Ce modèle standard rend compte de façon remarquable d'un grand nombre d'observations d'origine très diverse, tout en nous apprenant que la matière telle que nous la connaissons ne représente qu'une fraction infime (moins de 5 %) de l'Univers : les observations concordent pour suggérer qu'il est aussi constitué de 25 % de matière sombre et de 70 % d'énergie sombre, dont la nature reste aujourd'hui complètement mystérieuse.

L'année 2005 a été choisie comme « année mondiale de la physique » par l'ONU et l'UNESCO pour commémorer l'« annus mirabilis » dont c'est le centième anniversaire, et ce livre s'inscrit dans cette perspective. Nous remercions les auteurs, qui figurent parmi les meilleurs spécialistes au monde dans leurs disciplines respectives, d'avoir pris le temps d'apporter des contributions originales tout spécialement dédiées à ce centième anniversaire.





**Michèle Leduc**

Michèle Leduc est directrice de recherche au CNRS. Elle poursuit des recherches en physique des atomes refroidis par laser au Laboratoire Kastler Brossel à l'École Normale Supérieure (Paris). Elle dirige la collection « Savoirs Actuels » du CNRS depuis 1986.



**Michel Le Bellac**

Michel Le Bellac est professeur émérite à l'Université de Nice, chercheur au laboratoire INLN à Sophia Antipolis. Auteur de nombreux livres de physique, il est le coordinateur de l'ouvrage « Einstein aujourd'hui ».



# Table des matières

<b>Avant-propos</b> . . . . .	<b>iii</b>
<i>Michèle Leduc et Michel Le Bellac</i>	
<b>1 1905 : un nouvel élan</b> . . . . .	<b>1</b>
<i>Olivier Darrigol</i>	
<b>2 De l'article d'Einstein Podolsky et Rosen à l'information quantique : les stupéfiantes propriétés de l'intrication</b> . . .	<b>39</b>
<i>Alain Aspect et Philippe Grangier</i>	
<b>3 La condensation de Bose-Einstein dans les gaz</b> . . . . .	<b>87</b>
<i>Claude Cohen-Tannoudji, Jean Dalibard et Franck Lalö</i>	
<b>4 De l'émission stimulée d'Einstein aux lasers aujourd'hui</b>	<b>129</b>
<i>Jean-Paul Pocholle</i>	
<b>5 Le mouvement brownien et le théorème de fluctuation-dissipation</b> . . . . .	<b>203</b>
<i>Bernard Derrida et Éric Brunet</i>	
<b>6 Relativité générale</b> . . . . .	<b>267</b>
<i>Thibault Damour</i>	
<b>7 Cosmologie</b> . . . . .	<b>321</b>
<i>François R. Bouchet</i>	



# Chapitre 1

## 1905 : un nouvel élan

Par *Olivier Darrigol*

### Sommaire

1.1	Thème de recherche 1 : l'électrodynamique des corps en mouvement . . . . .	7
1.2	Thème de recherche 2 : microphysique théorique . . . . .	13
1.3	Thème de recherche 3 : le rayonnement du corps noir . . . . .	17
1.4	Les clefs du succès . . . . .	21
1.5	Trois réceptions . . . . .	23
1.6	L'esprit relativiste . . . . .	24
1.7	L'énigme quantique . . . . .	28
1.8	Après la mort . . . . .	34
1.9	Sources et références . . . . .	35



### **Olivier Darrigol**

Olivier Darrigol est directeur de recherche au CNRS, membre du groupe d'Histoire des sciences Rehseis (Paris). Après une formation initiale en physique théorique, il s'est orienté vers l'Histoire des sciences. Il est l'auteur d'ouvrages concernant l'Histoire de la physique quantique, celle de l'électrodynamique et celle de l'hydrodynamique. Il a contribué à l'édition des œuvres choisies d'Einstein en français.

# 1905 : un nouvel élan

C'ÉTAIT IL Y A CENT ANS. Albert Einstein publiait dans *Annalen der Physik* son fameux triptyque : quanta lumineux / mouvement brownien / relativité. Dans le présent ouvrage, quelques physiciens d'aujourd'hui témoignent de la fertilité du genre de théorie physique inauguré en ce temps. Le but de l'essai historique que voici est de montrer comment l'année 1905 marqua un tournant dans la quête intellectuelle d'Einstein, de la critique des fondements à la construction hésitante, paradoxale et inachevée d'une nouvelle physique.

Pour un adolescent de la fin du XIX<sup>e</sup> siècle, témoin d'un afflux de merveilles techniques, lecteur avide d'ouvrages de vulgarisation scientifique et fils d'un chef d'entreprise d'électrotechnique, il était bien naturel de se passionner pour les sciences physiques. Dès ses années de lycée, Einstein savait qu'il voulait devenir physicien théoricien ou mathématicien. En témoigne l'essai qu'il écrivit à l'occasion d'un examen de Français<sup>1</sup> :

## Mes projets d'avenir.

Un homme heureux est trop content de la présence [du présent] pour penser beaucoup à l'avenir. Mais de l'autre côté ce sont surtout les jeunes gens qui aiment à s'occuper de hardis projets. Du reste c'est aussi une chose naturelle pour un jeune homme sérieux, qu'il se fasse une idée aussi précise que possible du but de ses désirs.

Si j'avais le bonheur de passer heureusement mes examens, j'irai à l'école polytechnique de Zürich. J'y resterais quatre ans pour étudier les mathématiques et la physique. Je m'imagine de devenir professeur dans ces branches de la science de la nature en choisissant la partie théorique de ces sciences.

Voici les causes qui m'ont porté à ce projet. C'est surtout la disposition individuelle pour les pensées abstraites et mathématiques, le manque de la phantasie [imagination] et du talent pratique. Ce sont aussi mes désirs qui me conduisaient à la même résolution. C'est tout naturel; on aime toujours faire les choses pour lesquelles on a le talent. Puis c'est aussi une certaine indépendance de la profession scientifique qui me plaît beaucoup.

---

<sup>1</sup>A. Einstein, essai extrait de la Staatsarchiv Kanton Argau, reproduit dans [2] pp. 42–43 (fac-similé avec les corrections du professeur, ignorées dans la présente transcription). Cet essai valut à Einstein la note 3-4 sur 6. Voir [1–3].

À l'âge de seize ans, Einstein imagine un modèle de l'éther électromagnétique, un peu naïf mais cohérent. « *C'est le cœur lourd* », écrit-il sur un ton qu'il allait vite abandonner, « *que je concentre ces réflexions en un essai qui apparaît plus comme un programme que comme un mémoire. . . Fasse que l'indulgence du lecteur bien disposé soit à la mesure du sentiment d'humilité avec lequel je soumetts ces lignes.* » L'année suivante, à son entrée au *Polytechnikum* de Zurich, il décide fermement de se consacrer à la physique. Durant le reste de ses études, il partage son temps entre la fréquentation du laboratoire, les recherches personnelles et une liaison amoureuse avec son condisciple serbe Mileva Marić. Il ne suit qu'une partie des cours, qu'il juge trop dogmatiques, et préfère lire les grands auteurs de l'époque : Kirchhoff, Helmholtz, Hertz, Drude, Boltzmann, Mach et Poincaré<sup>2</sup>.

Ambitieux, à l'affût de la nouveauté, Einstein médite déjà sur les tensions de la physique contemporaine, sur les points d'achoppement signalés par les grands maîtres. Il n'en manquait pas. Dans les douze dernières années du siècle et particulièrement en Allemagne, la physique avait connu une avalanche de découvertes qui remettaient en question bien des idées reçues. En 1888, les expériences de Heinrich Hertz avaient imposé la théorie électrodynamique de James Clerk Maxwell au dépend des vieilles théories allemandes d'action directe à distance. Dans les années 1890, les découvertes des rayons X, de la radioactivité et de l'électron avaient nourri une nouvelle microphysique expérimentale. Aux yeux d'un nombre croissant de physiciens, atomes, ions et électrons n'étaient plus de simples fictions théoriques : il devenait possible d'étudier dans le laboratoire des conséquences assez directes de leur existence [5, 6].

En raison de ces développements, les jugements sur les théories existantes évoluèrent, de nouvelles théories apparurent, on débattit de leurs mérites et l'on s'interrogea sur les fondements de la théorie physique en général. Dans le domaine de l'électromagnétisme, Hertz proposa une version plus abstraite de la théorie de Maxwell, où les équations fondamentales étaient posées comme axiomes, sans aucun fondement mécanique. À l'instar de Maxwell, Hertz admettait toutefois l'existence d'un milieu de propagation, l'éther, éventuellement modifié par la présence de matière. Par souci de simplicité, il supposait que ce milieu était complètement entraîné par la matière dans ses mouvements. Mais comme il le soulignait lui-même, cette hypothèse était contredite par des expériences d'optique des corps en mouvement (par exemple, l'expérience de Hippolyte Fizeau de 1851, selon laquelle la lumière se propageant à travers un courant d'eau n'est que partiellement entraînée par ce courant). Einstein fut frappé par cette remarque de Hertz et esquissa à ce moment une approche de l'électromagnétisme semblable à celle du théoricien néerlandais Hendrik Lorentz, dont il prit connaissance un peu plus tard<sup>3</sup>.

---

<sup>2</sup>Einstein, « *Über die Untersuchung des Aetherzustandes in magnetischem Felde* », manuscrit de 1895 [4] (vol. 1 : pp. 6–9).

<sup>3</sup>Voir [6] et la lettre d'Einstein à Marić d'août 1899, [4] (vol. 1, traduit dans [10] vol. 2 : pp. 28–29).



Le but général de Lorentz était de résoudre les difficultés de la théorie électromagnétique de la lumière. Pour rendre compte de la dispersion optique et des effets du mouvement des corps transparents, il admettait, contrairement à Maxwell, que toutes les interactions entre l'éther et la matière s'effectuaient via des particules chargées (ions ou électrons) et que l'éther se laissait traverser par les particules de matière sans être aucunement entraîné. L'autorité de cette conception crut considérablement vers la fin du siècle, quand l'effet magnéto-optique de Pieter Zeeman, la découverte de l'électron et de nouvelles expériences d'optiques des corps en mouvement vinrent la confirmer. Les lettres d'Einstein de cette époque montrent qu'il la partageait [7].

Un peu plus tard, en 1901, Einstein lut les *Leçons sur la théorie des gaz* de Ludwig Boltzmann, avec un enthousiasme qui le poussa à s'exclamer : « *Der Boltzmann ist ganz grossartig!* » Comme Einstein le savait, la théorie de Boltzmann, développée de 1868 à 1895 à partir de la théorie cinétique des gaz de Maxwell, avait longtemps été négligée en Allemagne au profit de la thermodynamique macroscopique. Mais à la fin du siècle, la nouvelle microphysique expérimentale renforça les conceptions atomistes et permit de nouvelles applications ou transpositions de la théorie cinétique des gaz, par exemple les théories de l'ionisation par les rayons X ou la théorie électronique des métaux de Paul Drude. C'est à propos de cette dernière qu'Einstein s'exclama : « *Der Drude ist ein genialer Kerl.* » Sans doute était-il frappé par le succès d'une combinaison de deux innovations majeures, la théorie des électrons et la théorie cinétique des gaz, à l'intersection de deux domaines de la physique<sup>4</sup>.

Einstein suivit aussi, d'un œil critique, les travaux de Max Planck sur le rayonnement du corps noir, qui comme ceux de Drude mêlaient électromagnétisme et phénomènes thermiques. Stimulé par les recherches de Boltzmann, Planck cherchait à comprendre l'irréversibilité thermodynamique à partir de la micro-dynamique sous-jacente. Mais il ne tolérait pas l'interprétation statistique de la loi de croissance de l'entropie, et rejetait par conséquent la théorie cinétique moléculaire qui y conduisait. La véritable cause de l'irréversibilité, selon lui, était à chercher dans les interactions entre matière et rayonnement électromagnétique. Après maints efforts et diverses péripéties, il en vint à adopter une version formelle de la relation de Boltzmann entre entropie et probabilité et obtint ainsi la formule du rayonnement noir qui porte son nom. Nous y reviendrons plus loin. Mais notons déjà que pour Planck l'intérêt principal de cette dérivation était d'établir un pont entre théorie des gaz et théorie du rayonnement, et de permettre une nouvelle estimation du nombre d'Avogadro, en fait la plus précise de l'époque [4 ; vol. 1, 8].

De ses lectures au front des théories de l'époque, Einstein tira trois thèmes de recherche :

(1) Clarifier les relations entre l'éther et la matière, en rapport avec les difficultés de l'optique des corps en mouvement.

---

<sup>4</sup>Einstein à Marić, 13 septembre 1900 [4] (vol. 1 : pp. 259–261) ; 4 avril 1901 [4] (vol. 1 : pp. 284–286).

(2) Donner des moyens plus directs et plus précis d'accéder aux grandeurs moléculaires et affermir les fondements de la théorie de Boltzmann.

(3) Éclaircir les mécanismes responsables du rayonnement noir.

Einstein fut aussi sensible aux commentaires méthodologiques dont les grands physiciens de l'époque accompagnaient leurs travaux théoriques. Hertz insistait sur le caractère symbolique de la théorie physique, sur la nécessité de définitions opérationnelles des grandeurs fondamentales, et sur l'élimination des redondances dans la représentation des phénomènes. Boltzmann défendait la liberté de construction des théories physiques, et critiquait ceux qui croyaient ne point faire d'hypothèses en se confinant à la phénoménologie mathématique. Helmholtz et Planck promouvaient un genre de physique dans lequel des principes généraux tels que la conservation de l'énergie ou le principe de moindre action devaient guider et contraindre la construction des théories. Ils étaient convaincus d'une unité profonde de la physique et tentaient d'établir de nouveaux ponts entre des domaines jusque là séparés. Les adeptes d'une telle physique des principes, particulièrement Wilhelm Wien et Henri Poincaré, se servaient volontiers d'expériences de pensée pour explorer le jeu combiné des principes [1, 6].

Einstein prit ainsi conscience de diverses stratégies dont il pouvait s'inspirer au gré des circonstances. Et il s'imprégna d'une conception de la physique théorique qui intégrait la critique des fondements dans l'élaboration des théories. Sa lecture des écrits philosophiques de David Hume, Ernst Mach et Henri Poincaré renforça cette dimension de sa pensée. Pour ces auteurs comme pour les grands théoriciens de l'époque, il n'était plus question de concevoir une genèse purement inductive de la théorie physique. Certes, l'expérience restait le point de départ et le point d'arrivée de toute théorie. Mais l'activité intellectuelle du théoricien avait sa propre dynamique, ses propres outils critiques et constructifs. C'est à ce moment que la physique théorique devint une discipline autonome, surtout en Allemagne [5, 9].

Porté par cet élan de la physique théorique, Einstein mûrit ses réflexions sur les trois thèmes du rapport éther/matière, des grandeurs moléculaires et du rayonnement. Au printemps 1905, il annonce à son ami Conrad Habicht une série de progrès sur ces trois fronts, d'un ton jovial et badin<sup>5</sup> :

Cher Habicht,

Il règne entre nous un silence sacré, au point que j'ai un peu le sentiment de commettre une profanation en le brisant aujourd'hui par un bavardage sans grand intérêt. Mais toutes les choses élevées ne finissent-elles pas ainsi en ce monde ?

Que faites-vous donc, espèce de baleine congelée, espèce de quartier d'âme fumée en conserve, ou que sais-je encore que j'aimerais vous jeter à la figure, emporté que je suis par 70 % de colère et 30 % de pitié ! Vous pouvez remercier ces 30 % qui m'ont empêché que je ne vous expédie l'autre jour — alors que, plus silencieux qu'une tombe, vous n'étiez pas apparu à Pâques — une boîte de conserve remplie d'oignons et d'ail coupés en rondelles.

<sup>5</sup>Einstein à Habicht, printemps 1905, [4] vol. 3, traduit dans [10] vol. 1 : p. 36.

Mais pourquoi donc ne m'avez-vous toujours pas envoyé votre thèse? Ne savez-vous point, affreux personnage, que je fais partie des une personne et demie qui la parcourraient avec plaisir et intérêt? Je vous promets en échange quatre travaux, dont je pourrai prochainement vous envoyer le premier, car je dois recevoir très bientôt les exemplaires d'auteur. Il y est question de rayonnement et des propriétés énergétiques de la lumière, d'une façon tout à fait révolutionnaire, ainsi que vous le verrez si vous m'envoyez *auparavant* votre thèse.

Mon deuxième travail est une détermination de la taille véritable des atomes à partir de la diffusion et de la viscosité des solutions liquides diluées de substances neutres. Dans le troisième, je démontre que, d'après la théorie moléculaire de la chaleur, des corps d'un ordre de grandeur de 1/1000 mm en suspension dans un fluide doivent déjà effectuer un mouvement désordonné perceptible, lui-même provoqué par l'agitation thermique; et, de fait, les physiologistes ont observé des mouvements affectant de petits corps inanimés en suspension, mouvements auxquels ils donnent le nom de « mouvement moléculaire brownien. »

Le quatrième travail est encore à l'état d'ébauche; il s'agit d'une électrodynamique des corps en mouvement qui repose sur des modifications de la théorie de l'espace et du temps. La partie purement cinématique de ce travail vous intéressera sûrement.

Solo[vine] continue à donner des cours particuliers et ne se décide pas à passer l'examen; je le plains beaucoup, car il mène une bien triste existence. Il a d'ailleurs l'air bien éprouvé. Mais je ne crois pas qu'il soit possible de l'amener à avoir des conditions de vie plus supportables. Vous savez comment il est!

Bien cordialement,

Votre A.E.

Vous avez également toutes les amitiés de ma femme et du petit diable, qui a maintenant un an.

Envoyez vite votre travail!

Dans ce qui suit, nous allons voir comment Einstein aboutit aux quatre travaux auquel il fait allusion dans cette lettre, en suivant l'ordre historique d'apparition de ses thèmes de recherche plutôt que l'ordre des publications de 1905. Puis nous examinerons les clefs de son succès, la réception immédiate de ses travaux et les prolongements qu'il leur donna lui-même. Nous concluons par quelques indications de l'efflorescence actuelle de la physique einsteinienne.

## 1.1 Thème de recherche 1 : l'électrodynamique des corps en mouvement

La correspondance d'Einstein des années 1899–1901 porte plusieurs traces de son intérêt pour l'électrodynamique des corps en mouvement et pour les phénomènes optiques apparentés. Une lettre à Marić d'août 1899 témoigne de l'avancement de ses réflexions<sup>6</sup> :

<sup>6</sup>Voir note 3.

J'étudie une nouvelle fois avec la plus grande attention *La propagation de la force électrique* de Hertz. . . Je suis de plus en plus convaincu que l'électrodynamique des corps en mouvement, telle qu'elle se présente actuellement, ne correspond pas à la réalité, et qu'il doit être possible de la présenter de façon plus simple. L'introduction du mot « éther » dans les théories de l'électricité a conduit à l'idée d'un milieu du mouvement duquel il est possible de parler, sans que l'on soit capable, à mon avis, d'associer une signification physique à ce que l'on dit. . . Il faudra considérer les courants électriques non comme « une polarisation électrique s'annulant dans le temps » [point de vue de Maxwell] mais comme un mouvement de masses électriques vraies, dont les équivalents chimiques semblent prouver l'existence [allusion à la loi de Faraday]. . . L'électrodynamique serait alors la théorie des déplacements de [masses] électriques et magnétiques en mouvement.

Comme Lorentz et avant de l'avoir lu, Einstein pensait que tous les phénomènes électromagnétiques impliquant la matière devaient pouvoir se ramener aux interactions entre des particules chargées (les ions) et un éther stationnaire, dépourvu de toute matérialité. Ainsi un courant de conduction devait-il être compris comme une circulation d'ions ; une polarisation comme un déplacement d'ions élastiquement liés. La stationnarité rigoureuse de l'éther suggérait un effet du mouvement des dispositifs optiques terrestres par rapport à l'éther, qu'Einstein pensait pouvoir démontrer en comparant les intensités de la lumière émise par une source terrestre dans deux directions opposées [11].

Einstein pensait aussi que le mouvement d'un corps transparent par rapport à l'éther devait affecter la propagation de la lumière. En effet, en se propageant dans un corps transparent, une onde électromagnétique engendrait une vibration sympathique des ions et un rayonnement secondaire de ces ions. L'interférence des rayonnements primaire et secondaire permettait de prévoir, en sus de l'effet d'indice optique, un effet du mouvement de ce corps sur la vitesse de la lumière. Einstein proposa à son professeur de physique, Heinrich Weber, une expérience qui permit de tester cette conséquence de la stationnarité de l'éther<sup>7</sup>.

Sans qu'Einstein en fût encore conscient, Lorentz avait déjà calculé ce dernier effet, et Hippolyte Fizeau l'avait expérimentalement mesuré. Le coefficient théorique d'entraînement de la lumière,  $1 - 1/n^2$  si  $n$  est l'indice optique, était celui qu'Augustin Fresnel avait jadis introduit afin que les lois de la réfraction dans un prisme terrestre fussent insensibles au vent d'éther. L'expérience en confirmait la valeur. Weber, qui était conscient de ces développements, traita Einstein de haut (*stiefmütterlich*, selon l'intéressé) et le renvoya à la lecture d'un mémoire de Wien de 1898 récapitulant les progrès récents des physiciens sur la question du mouvement de l'éther. Wien discutait l'expérience de Fizeau, la théorie de Lorentz et maintes tentatives, toutes vaines, de détecter des effets du vent d'éther créé par le mouvement de la terre. Parmi celles-ci se trouvait l'expérience de 1887 par laquelle Albert Michelson et Edward Morley comparaient interférométriquement les vitesses de la lumière dans deux directions orthogonales. Le résultat positif de l'expérience

<sup>7</sup>Voir Einstein à Marić, 28 (?) septembre 1899, [4] vol. 1 : pp. 233–235.

de Fizeau semblait confirmer la stationnarité de l'éther, alors que le résultat négatif de celle de Michelson et Morley suggérait un entraînement total de l'éther par la matière ! [6, 7, 13, 14]

On imagine aisément la perplexité d'Einstein face à ce verdict contradictoire de l'expérience. Il continua quelque temps à imaginer de nouvelles expériences de détection du vent d'éther. Comme il l'apprit au début de 1902, Lorentz était parvenu, dans un ouvrage de 1895, à rendre compte du résultat négatif de telles expériences, dans la mesure où leur précision n'excédait pas le carré du rapport de la vitesse  $v$  de la terre (par rapport à l'éther) à celle,  $c$ , de la lumière. Au premier ordre, sa démonstration s'appuyait sur un passage du repère de l'éther à celui de la terre par la transformation  $x' = x - vt$ , suivi de l'introduction du « temps local »  $t' = t - vx'/c^2$ , qui permettait de retrouver la forme originale des équations de Maxwell (supposées valides dans le repère de l'éther). Au second ordre, Lorentz obtenait la même invariance en tenant compte d'une contraction longitudinale des corps solides lors de leur déplacement par rapport à l'éther, dans un rapport  $\sqrt{1 - v^2/c^2}$ . Cette contraction était nécessaire à l'équilibre interne d'un solide en mouvement, dans la mesure où ce mouvement affectait les forces de cohésion moléculaire comme des forces électromagnétiques [14].

Ainsi la théorie de Lorentz pouvait-elle rendre compte de la totalité des phénomènes électromagnétiques et optiques connus, y compris les expériences de Fizeau et de Michelson-Morley. Elle ne faisait pourtant pas l'unanimité. Au nom d'un positivisme Machien, l'électrodynamicien strasbourgeois Emil Cohn préférait s'abstenir complètement de l'éther, traiter les champs électriques et magnétiques comme des entités physiques indépendantes et modifier les équations de Maxwell macroscopiques dans la matière de telle sorte qu'elles rendent compte à la fois des expériences de Fizeau et de Michelson-Morley (1900–1904). À Bonn, Alfred Bucherer partit du principe, vérifié par les vieilles théories allemandes de l'électrodynamique, que les interactions électromagnétiques ne pouvaient dépendre que du mouvement de la matière par rapport à la matière ; il élimina l'éther ; et il suggéra de n'utiliser les équations de Maxwell que formellement, comme un moyen de calculer la force s'exerçant sur une particule chargée dans le repère tangent à son mouvement (1903–1906) [6].

Mais le critique le plus influent de Lorentz fut Henri Poincaré. Le mathématicien français ne voyait dans l'éther qu'une métaphore commode, un résidu de l'analogie fondatrice entre lumière et vibration d'un milieu élastique. Par conséquent, il jugeait extrêmement improbable que l'éther puisse être traité comme un autre corps matériel dans l'application des principes de la mécanique. En particulier, il fut le premier à penser, dès 1899, que le principe de relativité devait s'appliquer à la matière seule, sans considération de l'éther. Autrement dit, il jugeait que l'impossibilité de détecter le vent d'éther devait être érigée en principe. Il pensait aussi, en 1900, que le principe de l'égalité de l'action et de la réaction devait s'appliquer à la matière seule. Or la théorie de Lorentz violait ces deux principes entendus en ce sens. En effet, elle impliquait

le recul d'un émetteur unidirectionnel de rayonnement ; elle prévoyait un effet du premier ordre du vent d'éther sur les forces électromagnétiques ; et elle ne parvenait à expliquer l'absence d'autres effets de ce vent qu'au pris d'une accumulation d'hypothèses [6, 9, 14].

Dans un mémoire capital de 1900, Poincaré accompagna ces critiques d'une interprétation nouvelle des changements de champs et de coordonnées que Lorentz utilisait de manière purement formelle dans ses démonstrations d'indétectabilité du vent d'éther. Selon Poincaré, les champs et coordonnées transformés étaient ceux que des observateurs en mouvement (par rapport à l'éther) mesureraient en suivant certaines conventions naturelles. En particulier, le « temps local »  $t' = t - vx'/c^2$  de Lorentz devait être interprété (au premier ordre) comme le temps mesuré par des observateurs se mouvant à la vitesse  $v$  et ayant synchronisé leurs horloges par l'échange de signaux lumineux sans tenir compte de leur mouvement par rapport à l'éther (et faisant donc comme si la lumière se propageait à la vitesse constante  $c$  dans leur propre repère) [6, 15, 16].

Lorentz ne réagit pas à cette remarque et continua d'interpréter ses transformations comme relevant d'états fictifs. En 1904, il parvint cependant à les généraliser à tous les ordres en  $v/c$ . À une transformation galiléenne près, ces nouvelles transformations correspondaient à peu près à ce que Poincaré allait appeler « les transformations de Lorentz ». Mais l'invariance formelle était encore imparfaite (pour les termes de couplage) et du point de vue de Lorentz elle n'impliquait pas directement l'invariance des phénomènes. Il fallait encore supposer que la contraction de Lorentz s'appliquait aux électrons et ne retenir que la composante dipolaire du rayonnement. Au-delà de cette dernière approximation, l'éther restait en principe détectable [17].

À la lecture de ce nouveau mémoire, Poincaré comprit que de légères modifications des transformations de Lorentz conduisaient à un groupe d'invariance exacte des équations de Maxwell-Lorentz. Il affirma que cette invariance formelle impliquait une invariance exacte des phénomènes, conformément au principe de relativité. Il jugea que le même groupe d'invariance devait s'appliquer à tout type de force, y compris les forces de cohésion de l'électron et les forces gravitationnelles. Il obtint ainsi un modèle d'électron contractile et la dynamique relativiste correspondante. Enfin, il proposa une modification covariante de la loi de Newton grâce à un formalisme quadri-dimensionnel semblable à celui ultérieurement développé par Hermann Minkowski. Ces résultats parurent sous forme de résumé dans les *Comptes rendus* de juin 1905, et en 1906 dans un long et savant mémoire sur « la dynamique de l'électron » [2, 7, 9, 12, 18, 19].

Lors de leçons données à la Sorbonne en 1906, Poincaré montra de surcroît que les transformations de Lorentz étaient compatibles avec sa procédure optique de synchronisation des horloges, compte tenu de la contraction de Lorentz. Et il mit à profit l'invariance de Lorentz pour établir directement les équations du mouvement de l'électron, sans aucun modèle. Tout cela nous rapproche de la relativité d'Einstein. Il importe toutefois de souligner que

Poincaré, jusqu'à la fin de sa vie, maintint une nécessité psychologique de l'éther et continua de discerner l'espace et le temps « vrais », liés à l'éther, de l'espace et du temps « apparents », mesurés par des observateurs en mouvement. En vertu du principe de relativité, le choix du référentiel de l'éther ne pouvait être que purement conventionnel, mais il permettait de maintenir les concepts ordinaires d'espace et de temps, que Poincaré jugeaient trop enracinés dans l'usage courant pour que l'étude de situations physiques extrêmes justifîât leur remplacement [6, 20].

Revenons en 1901, à Einstein et à sa perplexité face au conflit entre le résultat de Fizeau et celui de Michelson-Morley. À un moment bien antérieur à 1905 mais que les documents connus ne permettent pas de dater précisément, Einstein cessa de s'interroger sur la possibilité éventuelle d'effets du vent d'éther et admit le principe de relativité. Contrairement à Poincaré (mais comme Bucherer), il décida de rejeter l'éther devenu complètement indétectable et d'admettre que la vitesse de la lumière dépendait de la vitesse de sa source, comme elle l'aurait fait dans une théorie corpusculaire newtonienne. Ce point de vue, développé par Walter Ritz en 1908, satisfait immédiatement au principe de relativité et permet de retrouver la plupart des résultats de l'électrodynamique de Lorentz, dans la mesure où les actions d'une source sont décrites par une expression modifiée des potentiels retardés tenant compte du mouvement de la source. Mais au bout d'un certain temps, Einstein se persuada que ce point de vue conduisait à des effets paradoxaux d'auto-couplage. Le dilemme de l'électrodynamique des corps en mouvement restait donc entier : si la vitesse de la lumière avait une valeur constante, indépendante de sa source (comme le suggérait la théorie de Lorentz), alors d'après le principe de relativité cette constante devait être la même dans tout référentiel inertiel, contrairement à la loi galiléenne d'addition des vitesses<sup>8</sup>.

Einstein rapporte qu'au début de 1905, lors d'une conversation avec son ami Michele Besso [1], il eut soudain l'idée que le temps pouvait être redéfini de telle manière que la vitesse de la lumière soit effectivement la même dans tout référentiel inertiel. Très vite, utilisant une synchronisation optique des horloges semblable à celle de Poincaré, il put montrer le caractère relatif de la simultanéité et dériver la forme exacte des transformations de Lorentz. Il examina ensuite les effets physiques de contraction des longueurs et de dilatation des temps ; il obtint la nouvelle loi d'addition des vitesses ; il démontra l'invariance des équations de Maxwell-Lorentz sous le groupe de transformation de Lorentz, avec quelques applications à l'optique des corps en mouvement ; et il donna la forme invariante relativiste des équations de la dynamique [2, 9, 12, 21].

Tel est en bref le contenu du célèbre article reçu le 30 juin 1905 par *Annalen der Physik* [21]. Les sources de l'inspiration soudaine d'Einstein

---

<sup>8</sup>Voir [4] vol. 2 : pp. 263–264. Le 17 décembre 1901, Einstein confiait à Marić (vol. 1 : pp. 325–326) qu'il travaillait à un « article capital » sur l'électrodynamique des corps en mouvement. Rien n'indique s'il pensait encore à une théorie du type de celle de Lorentz ou s'il s'agissait déjà d'une théorie de l'émission.

restent à ce jour mystérieuses. De façon certaine, on sait qu'Einstein avait lu le *Versuch* de Lorentz de 1895 (qui contient le temps local), qu'il était au courant des expériences sur l'entraînement de l'éther et qu'il avait lu *La science et l'hypothèse* de Poincaré, qui contenait une défense du principe de relativité et quelques considérations sur la mesure du temps et sur le caractère conventionnel de la simultanéité. Par ailleurs, il était probablement conscient de la transformation de Lorentz de 1904 et il avait probablement lu le mémoire de Poincaré de 1900 qui contient l'interprétation du temps local<sup>9</sup>. Il est donc possible que l'inspiration lui soit venue en méditant cette interprétation.

Si toutefois Poincaré avait été sa seule source, il n'aurait point éprouvé le besoin de renverser les concepts classiques d'espace et de temps et d'éliminer complètement l'éther. Il semblerait que son passage antérieur par une théorie de l'émission et peut-être aussi le concept de quanta lumineux aient facilité cette étape radicale. Mais la clef de l'originalité d'Einstein se trouve sans doute dans l'introduction qu'il donne à son mémoire de juin 1905 : il y dénonce, dans la théorie de Lorentz, l'existence d'asymétries dépourvues de contrepartie phénoménale. Ainsi la théorie de Lorentz donne-t-elle deux explications différentes du phénomène d'induction électromagnétique lors du mouvement relatif d'un conducteur et d'un aimant : par un champ électrique ( $\mathbf{E}$ ) si c'est l'aimant qui se meut à travers l'éther, par un champ électromoteur ( $\mathbf{v} \times \mathbf{B}$ ) dépourvu d'énergie si c'est le conducteur qui se meut. Einstein, contrairement à Poincaré, refusait une telle redondance dans la représentation théorique des phénomènes et exigeait que tous les référentiels inertiels soient traités sur le même pied, puisque qu'aucun effet physique ne permettait de les discerner. Il se situait ainsi dans la lignée de Mach et de Hertz, qui favorisaient l'économie et l'univocité de la représentation au dépend des intuitions acquises [9].

Non content de réformer les concepts d'espace et de temps, quelques mois plus tard Einstein remit en question le dogme classique de la conservation de la masse. En 1900, Poincaré avait noté que l'application de la conservation de l'énergie à l'émission de rayonnement conduisaient à des estimations différentes de l'impulsion de recul de la source, selon que le bilan était effectué par un observateur lié à la source ou par un observateur en mouvement dans la direction d'émission. Dans sa théorie de la relativité, Einstein se heurta à un paradoxe similaire. Supposons avec lui qu'une quantité  $J/2$  d'énergie rayonnante soit émise simultanément par une source immobile dans deux directions opposées (et donc sans recul). Pour un observateur se mouvant à la vitesse  $u$  sur la ligne d'émission, l'énergie émise peut se calculer en appliquant une transformation de Lorentz aux champs électromagnétiques correspondants, ce qui donne :

$$J' = (1 - u^2/c^2)^{-1/2}(1 + u/c)J/2 + (1 - u^2/c^2)^{-1/2}(1 - u/c)J/2 \approx J(1 + u^2/2c^2) \quad (1.1)$$

---

<sup>9</sup>Il le cite en effet dans un article de 1906.



Autrement dit, l'énergie consommée par l'émetteur semble dépendre du mouvement de l'observateur qui l'apprécie. Einstein ne vit qu'une manière d'éviter cette violation du principe de relativité : admettre une diminution  $J/c^2$  de la masse de la source lors du processus d'émission. En effet, une telle variation de masse implique que l'énergie cinétique de l'émetteur diminue de  $-(1/2)(J/c^2)u^2$ , conformément à l'excédent d'énergie rayonnante donné par la formule précédente [6, 15, 22].

De ce raisonnement particulier, Einstein inféra que la masse d'un corps dépendait de son contenu en énergie de manière tout à fait générale. Il nota aussi que des défauts de masse appréciables devaient correspondre aux énormes énergies impliquées dans les transformations radioactives. « *La chose est plaisante à considérer* », confia-t-il alors à son ami Conrad Habicht, « *mais Dieu n'est-il pas en train d'en rire et me mène-t-il par le bout du nez ? Ça, je suis incapable de le savoir.* »<sup>10</sup>

## 1.2 Thème de recherche 2 : microphysique théorique

Alors même qu'il abordait, autour de 1900, la question des rapports de la matière et de l'éther, Einstein s'enthousiasmait pour les théories moléculaires. Sa correspondance amoureuse des années 1899–1901 porte la trace de tentatives d'explications ioniques ou électroniques des relations entre la chaleur, l'électricité et le rayonnement. Il cherchait à comprendre les effets thermo-électriques comme un couplage entre le mouvement thermique et les forces électriques ; il évaluait la contribution du gaz d'électrons d'un métal à sa chaleur spécifique — d'où son intérêt pour la théorie de Drude ; et il imaginait que la chaleur spécifique d'isolants dépendit des vibrations d'électrons liés, ce qui la relierait au spectre d'absorption infrarouge et au rayonnement thermique.

En proposant des vérifications expérimentales de ces idées, Einstein espérait sans doute attirer l'attention de son professeur de physique, Heinrich Weber, qui était spécialiste du type pertinent de mesures thermiques. Il envoya aussi une lettre critique à Drude. N'obtenant point le soutien espéré, il se détourna temporairement de cette ligne de recherche et s'intéressa, de 1900 à 1902, à la détermination expérimentale de constantes moléculaires à partir des phénomènes de capillarité, d'électrochimie des solutions salines, et de transport dans les gaz. Dans ses deux premières publications, en 1901 et 1902, et dans une première thèse doctorale jamais soutenue, il identifiait les forces s'exerçant entre deux molécules à la superposition de forces centrales de la forme  $c_i c_j F(r_{ij})$ , où  $i$  et  $j$  sont deux atomes constituants de ces molécules,  $r_{ij}$  la distance qui les sépare,  $c_i$  et  $c_j$  des constantes caractéristiques de ces atomes, et  $F$  une fonction universelle. Cette hypothèse impliquait des relations entre les tensions superficielles de diverses substances (cas des liquides),

<sup>10</sup>Einstein à Habicht, été 1905, [4] vol. 3, extrait traduit dans [10] vol. 2 : pp. 59–60.

entre leurs potentiels électrochimiques (cas des solutés) ou leurs coefficients de transport (cas des gaz)<sup>11</sup>.

Quelques spéculatives que fussent ces théories du jeune Einstein, elles lui permirent de mêler astucieusement des raisonnements thermodynamiques macroscopiques et des raisonnements mécaniques microscopiques. Ce jeu d'échelles le poussa sans doute à approfondir la *Théorie des gaz* de Boltzmann, qu'il avait lue en septembre 1900. Dans ses trois publications suivantes (1902 à 1904), il proposait de fonder toute la thermodynamique statistique sur une base incontestable et très générale. Selon lui, Boltzmann s'était abondamment servi de probabilités sans pour autant donner à ce concept une signification physique précise. Pour combler ce déficit, Einstein définissait la probabilité d'un état d'un système par la fraction de temps passée par le système dans cet état au bout d'une évolution très longue [2, 23–25].

Pour rendre compte de l'uniformité du comportement thermodynamique d'un système macroscopique fermé, Einstein admettait que la probabilité ainsi définie était la même pour tout état initial compatible avec une valeur donnée des invariants globaux (quantité de mouvement, moment cinétique et énergie). Autrement dit, il supposait que la moyenne temporelle  $\bar{\Omega}$  d'une propriété quelconque  $\Omega$  du système était la même pour tout état initial ainsi choisi. Pour un ensemble stationnaire de copies du système soumis à la même restriction, cette hypothèse (ergodique faible) implique

$$\bar{\Omega} = \langle \bar{\Omega} \rangle = \overline{\langle \Omega \rangle} = \langle \Omega \rangle \quad (1.2)$$

où  $\langle \ \rangle$  désigne la moyenne d'ensemble. Ainsi la moyenne temporelle peut-elle être remplacée par une moyenne sur un ensemble stationnaire. De surcroît, l'unicité supposée de la moyenne temporelle implique l'unicité de la moyenne d'ensemble.

Comme Einstein souhaitait le fondement le plus général possible, il n'exigeait pas la structure hamiltonienne de la dynamique du système. Il se contentait d'admettre l'existence d'un choix de coordonnées tel que le volume d'un élément d'espace des phases soit invariant dans le temps. Il est alors évident que l'ensemble défini par une distribution uniforme sur l'hypersurface associée aux invariants globaux est stationnaire. Dans le cas d'un système contenu dans un récipient rigide, le seul invariant global est l'énergie, et l'ensemble stationnaire pertinent correspond à la distribution microcanonique de Gibbs. Einstein montrait plus loin qu'un petit sous-ensemble d'un tel ensemble était distribué canoniquement, avec la densité  $\rho = e^{-\beta H}/Z$ ,  $H$  étant la fonction énergie du système,  $Z$  un facteur de normalisation, et  $\beta$  un coefficient identifié à  $1/kT$  (où  $T$  désigne la température du gaz parfait, et  $k$  le rapport de la constante des gaz parfaits au nombre d'Avogadro). Il définissait travail, chaleur et entropie à partir des déformations lentes d'un tel ensemble. Il reliait l'entropie d'un macro-état à sa probabilité microcanonique, et déduisait la loi de croissance de l'entropie d'un système isolé de l'hypothèse de la probabilité toujours croissante de ses états.

<sup>11</sup>Voir [4] vol. 1 : pp. 264–266 ; vol. 2 : pp. 3–8.

Mis à part l'admission de microdynamiques non hamiltoniennes, rien de tout cela n'était original. Sans qu'Einstein le sût, des considérations très similaires se trouvaient disséminées dans divers écrits de Boltzmann ; et Josiah Willard Gibbs avait offert, dans sa *Mécanique statistique* de 1902, une magnifique étude systématique des ensembles statistiques. Einstein se démarquait cependant de ses prédécesseurs en retournant constamment à l'interprétation physique des probabilités thermodynamiques comme fraction de temps passée par le système dans tel ou tel de ses états. Cette interprétation mettait l'accent sur les fluctuations du système autour de son état d'équilibre, et faisait ainsi des écarts au comportement thermodynamique l'essence de la mécanique statistique, alors que Boltzmann et Gibbs insistaient sur le caractère négligeable de tels écarts [26].

À la fin de son mémoire de 1904, Einstein se servit de l'expression

$$\overline{(E - \bar{E})^2} = kT^2 d\bar{E}/dT \quad (1.3)$$

des fluctuations quadratiques moyennes de l'énergie  $E$  d'un système thermostaté pour faire de la constante  $k$  la mesure de la « stabilité thermique » d'un système. Il jugea « très intéressant » le fait que cette expression ne fasse intervenir que des grandeurs purement thermodynamiques (l'énergie interne  $\bar{E}$  et la température  $T$ ). Plus généralement, il interprétait la relation de Boltzmann

$$S = k \ln W \quad (1.4)$$

entre entropie et probabilité comme un pont entre une grandeur thermodynamique mesurable (ici l'entropie) et la probabilité  $W$  de fluctuations reflétant la microdynamique du système.

Pour déterminer la constante  $k$  et le nombre d'Avogadro, il suffisait donc d'imaginer une situation physique dans laquelle les fluctuations fussent accessibles à l'expérience. En 1904, Einstein n'en connaissait point. Mais il imagina que dans un corps noir de dimensions  $L$  de l'ordre de la longueur d'onde moyenne  $\lambda_m$  du rayonnement thermique, les fluctuations quadratiques relatives devaient être de l'ordre de l'unité en raison des ondulations de ce rayonnement. Compte tenu de la loi de Stefan  $\bar{E} = L^3 \sigma T^4$ , cette condition conduit à la relation  $\lambda_m \approx (4k/\sigma)^{1/3} T^{-1}$ , qu'Einstein jugea compatible avec les mesures du rayonnement noir et la valeur alors admise de  $k$ . Le succès de cet argument le convainquit du rôle que pouvaient jouer les fluctuations dans l'exploration microphysique, pour la matière comme pour le rayonnement.

En 1905, Einstein prit conscience de l'existence et de l'observabilité des fluctuations de particules suspendues dans un liquide. Ayant lu *La Science et l'hypothèse*, il avait peut-être noté l'allusion que faisait Poincaré aux travaux du Lyonnais Louis-George Gouy, qui tendaient à démontrer l'origine thermocinétique du mouvement brownien. Mais l'intérêt d'Einstein pour ce problème semble plutôt dériver de sa thèse doctorale d'avril 1905, dans laquelle il se proposait de déterminer des dimensions moléculaires et le nombre d'Avogadro à partir des propriétés de viscosité et de diffusion de l'eau sucrée. Par un

raisonnement purement hydrodynamique, il dérivait d'abord la relation

$$\mu^* = \mu(1 + f) \quad (1.5)$$

entre la viscosité  $\mu^*$  de la solution, la viscosité  $\mu$  du solvant, et la fraction de volume  $f$  occupée par le soluté. Puis il établissait, comme on va le voir, la relation

$$D = RT/6\pi\mu Na \quad (1.6)$$

entre le coefficient  $D$  de diffusion du soluté, la constante  $R$  des gaz parfaits, la viscosité  $\mu$  du solvant, le nombre d'Avogadro  $N$  et le rayon  $a$  des molécules du soluté. De ces deux relations et des données empiriques existantes, il extrayait ensuite les valeurs de  $N$  et  $a$  [2, 27]<sup>12</sup>.

Pour dériver la seconde relation, Einstein équilibrait la résistance visqueuse d'une molécule sphérique avec l'action de la pression osmotique sur la surface de cette molécule :

$$6\pi\mu av = -(1/\nu N)d(\nu RT)/dx \quad (1.7)$$

où  $v$  est la vitesse de la molécule,  $\nu$  la concentration molaire de soluté, et  $x$  la coordonnée par rapport à laquelle la concentration varie. Comme Einstein le notait, cette équation ne va pas de soi, car du point de vue de la théorie cinétique la pression osmotique n'est qu'une force apparente, résultant d'actions moléculaires stochastiques et donc bien différente d'une force extérieure réelle (électrique ou gravifique par exemple). Sa justification ultime repose sur la mécanique statistique. En effet, suivant une loi de Boltzmann, la concentration d'une substance soumise à un champ de force de potentiel  $V$  est proportionnelle à  $e^{-V/kT}$ , ce qui revient formellement à dire que la force est opposée au gradient de la pression osmotique.

Dans ce raisonnement, Einstein traitait les molécules du soluté comme des corps macroscopiques soumis à des forces apparentes. Il était donc naturel qu'il s'interrogeât sur le cas d'une suspension. Du point de vue de la thermodynamique macroscopique, notait-il, la seule force agissant sur les particules en suspension est la force visqueuse (en sus de la gravité). Du point de vue de la thermodynamique statistique, une suspension ne diffère d'une solution que par l'échelle des particules suspendues. Elle doit donc être soumise elle aussi à des effets de diffusion, toujours donnés par la formule  $D = RT/6\pi\mu Na$ . Cette diffusion, tout comme dans le cas du soluté, doit être comprise comme un effet stochastique des collisions des molécules du fluide environnant (toute interaction directe entre les particules est négligée).

Le but de l'article d'Einstein reçu le 11 mai 1905 par *Annalen der Physik* était d'explicitier l'interprétation stochastique de la diffusion des particules en suspension, et d'en déduire une nouvelle manière d'atteindre le nombre d'Avogadro. Considérons, avec Einstein, un temps  $\tau$  très bref devant le temps caractéristique de diffusion, mais suffisamment long pour que les déplacements

<sup>12</sup>Comme Ludwig Hopf le découvrit en 1911, la formule (1.5) comporte une erreur de calcul et doit être remplacée par  $\mu^* = \mu(1 + 5f/2)$ .

d'une des particules au cours d'intervalles successifs de durée  $\tau$  soient statistiquement indépendants. Soit  $\varphi(\Delta)d\Delta$  la probabilité pour que le déplacement d'une particule au bout du temps  $\tau$  soit compris entre  $\Delta$  et  $\Delta + d\Delta$  ; et  $f(x, t)$  la distribution spatiale (unidimensionnelle) des particules à l'instant  $t$ . La formule

$$f(x, t + \tau) = \int_{-\infty}^{+\infty} f(x + \Delta, t)\varphi(\Delta)d\Delta \quad (1.8)$$

développée au premier ordre en  $\tau$  et au second ordre en  $\Delta$ , conduit à l'équation de diffusion

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2} \quad (1.9)$$

avec

$$D = \frac{1}{2\tau} \int_{-\infty}^{+\infty} \Delta^2 \varphi(\Delta)d\Delta \quad (1.10)$$

Le déplacement quadratique moyen des particules au bout du petit temps  $\tau$  se trouve ainsi relié au coefficient de diffusion. Cette relation reste vraie pour des temps arbitrairement longs, comme l'indique la solution

$$f(x, t) = \frac{n}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (1.11)$$

de l'équation de diffusion (9) pour la condition initiale  $f(x, 0) = n\delta(x)$ . La mesure du déplacement quadratique moyen d'une particule brownienne permettrait donc d'accéder au nombre d'Avogadro  $N$  via l'expression  $D = RT/6\pi\mu Na$  du coefficient de diffusion. En conclusion de son article, Einstein exhortait les expérimentateurs à étudier cette « question importante pour la théorie de la chaleur » [2, 28].

On connaît l'écho de ce mémoire d'Einstein (et des travaux indépendants de Marian von Smoluchowski), particulièrement en France, avec les expériences qui valurent à Jean Perrin le prix Nobel et avec l'importante théorie stochastique de Paul Langevin (1908). Le physicien d'aujourd'hui y trouve en germe la relation fluctuation-dissipation et la notion de processus Markovien. Pour les contemporains d'Einstein, le mouvement brownien offrait non seulement une nouvelle méthode de mesure du nombre d'Avogadro, mais encore il donnait aux fluctuations thermodynamiques une réalité physique qu'elles n'avaient point auparavant. Un peu de fumée et un microscope suffisaient à révéler d'incessantes violations de la thermodynamique classique, pour la plus grande joie du jeune iconoclaste qu'était alors Einstein.

### 1.3 Thème de recherche 3 : le rayonnement du corps noir

De l'opinion d'Einstein, ses mémoires de 1905 sur le mouvement brownien et sur la relativité n'étaient pourtant pas les plus révolutionnaires.

Seul le premier mémoire du triptyque, celui sur les quanta lumineux, méritait ce qualificatif. Le principal contexte en était le problème du rayonnement thermique, qui se situait à la croisée de l'intérêt d'Einstein pour les phénomènes thermiques, l'électromagnétisme, leurs interrelations et la microphysique sous-jacente. Dès 1901, Einstein s'intéressa à la théorie que Planck prétendait donner de ce phénomène<sup>13</sup>.

Vers la fin du XIX<sup>e</sup>, le rayonnement thermique était l'objet de nombreuses études expérimentales et théoriques, en raison du développement des lampes à incandescence, de soucis métrologiques et de l'attrait théorique qu'exerçait un phénomène mêlant chaleur et lumière. On savait, depuis Gustav Kirchhoff (1860), que les lois de la thermodynamique impliquaient l'universalité du spectre du rayonnement émis par un « corps noir », c'est-à-dire un corps totalement absorbant et maintenu à température constante. Concrètement, ce spectre est celui du rayonnement contenu dans une cavité absorbante thermostatée. Son universalité attirait naturellement les métrologues et les théoriciens. Max Planck espérait le dériver théoriquement au bout d'une étude de l'évolution irréversible d'un système d'oscillateurs électriques (« résonateurs ») en interaction avec le rayonnement confiné dans une cavité aux parois idéalement réfléchissantes. L'échec de cette approche le conduisit, vers la fin de 1900, à se tourner vers la relation de Boltzmann entre entropie et probabilité [8, 26].

Planck définissait le macro-état d'un ensemble de  $N$  résonateurs de fréquence  $\nu$  par leur énergie moyenne  $U$ , et un micro-état par la donnée des énergies de chaque résonateur. Comme Boltzmann, il donnait à ces énergies des valeurs discrètes, multiples entiers de l'élément  $\varepsilon$ , afin d'obtenir une valeur finie et définie du nombre  $W$  de micro-états compatibles avec un macro-état donné. Posant  $P = NU/\varepsilon$ , cela donne

$$W = (N + P - 1)! / (N - 1)! P! \quad (1.12)$$

Dans la limite de grand  $N$ , la formule de Boltzmann  $S = k \ln W$  et la relation thermodynamique  $dS/dU = 1/T$  donnent alors

$$U = \frac{\varepsilon}{e^{\varepsilon/kT} - 1} \quad (1.13)$$

Par ailleurs, en appliquant la théorie de Maxwell aux interactions entre résonateur et ondes électromagnétiques de phase aléatoire et de fréquence accordée, Planck avait obtenu la relation

$$u_\nu = (8\pi\nu^2/c^3)U \quad (1.14)$$

entre l'énergie moyenne (temporelle)  $U$  du résonateur et la densité spectrale énergétique  $u_\nu$  du rayonnement noir. Il savait aussi, d'après un raisonnement thermodynamique dû à Wien, que la fonction  $u_\nu(T)$  était nécessairement de

---

<sup>13</sup>Voir Einstein à Marić, 4 avril 1901, [4] vol. 1 : pp. 284–286.

la forme  $\nu^2 f(\nu/T)$ . Il fallait donc prendre  $\varepsilon = h\nu$ , ce qui conduit à la loi de Planck [29]

$$u_\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1} \quad (1.15)$$

Planck se réjouit de l'excellent accord de cette formule avec les mesures les plus récentes de ses amis spectroscopistes berlinois. Il se félicita aussi d'établir, grâce à la formule de Boltzmann et à la constante  $k$ , un pont entre théorie des gaz et théorie du rayonnement, et il en tira une nouvelle détermination du nombre d'Avogadro. Sa démonstration restait cependant obscure. Pourquoi les éléments d'énergie persistaient-ils dans l'expression finale de l'entropie d'un résonateur, alors que dans les considérations analogues de Boltzmann (pour les gaz) toute trace de discrétisation s'évanouissait au bout des calculs ?

Planck avait certes des raisons, liées à une réinterprétation très particulière de la théorie de Boltzmann, d'attendre une telle asymétrie. Mais il était le seul à les comprendre. Einstein, qui ne se laissait impressionner par aucune autorité, jugea que Planck avait erré dans son application de la méthode de Boltzmann. Selon lui, cette méthode devait nécessairement conduire à la valeur  $kT$  de l'énergie moyenne d'un résonateur, et donc à la formule spectrale

$$u_\nu = (8\pi\nu^2/c^3)kT \quad (1.16)$$

Le succès de Planck dans la détermination de la valeur de  $k$  ne dépendait que de la validité approximative de cette dernière formule dans le domaine des basses fréquences. Pour rendre compte du reste du spectre, il fallait nécessairement sortir du cadre de l'électrodynamique classique.

Einstein exprima publiquement cette opinion dans l'introduction de l'article reçu le 18 Mars 1905 par *Annalen der Physik* [30] et intitulé : « Sur un point de vue heuristique concernant la production et la transformation de la lumière. » Il n'y mentionnait toutefois pas le nom de Planck, ayant appris à ses dépens qu'un peu de prudence diplomatique pouvait aider à sa carrière. Une fois constaté l'échec de la détermination de la probabilité thermodynamique  $W$  à partir de la théorie classique du rayonnement, il se proposait ensuite de faire l'inverse, c'est-à-dire d'inférer les éléments d'une nouvelle théorie du rayonnement à partir de l'expression de  $W$  requise par l'expérience [2, 26, 30, 31].

Dans le domaine des hautes fréquences, le spectre du rayonnement noir est approximativement représenté par la loi de Wien

$$u_\nu = \alpha\nu^3 e^{-\beta\nu/T} \quad (1.17)$$

Pour un volume  $V$  et une énergie  $E = u_\nu V d\nu$ , l'entropie d'un tel rayonnement est

$$S = \int_0^E dE'/T' = -\frac{E}{\beta\nu} \ln(E/\gamma V) \quad (1.18)$$

avec  $\gamma = e\alpha\nu^3 d\nu$ . Si le volume passe de  $V_0$  à  $V$ , l'entropie varie de

$$S(V) - S(V_0) = \ln(V/V_0)^{E/\beta\nu} \quad (1.19)$$

Grâce à la formule de Boltzmann  $S = k \ln W$ , Einstein en déduit que la probabilité  $W$  d'une fluctuation du rayonnement par laquelle le rayonnement d'une cavité de volume  $V$  se trouve confiné dans le volume partiel  $V_0$ , est donnée par

$$W = (V/V_0)^{E/k\beta\nu} \quad (1.20)$$

Cette expression, note Einstein, est celle qu'on attendrait si le rayonnement était constitué d'un nombre  $E/k\beta\nu$  de quanta d'énergie indépendants  $k\beta\nu$ <sup>14</sup>.

D'autres n'auraient vu là que simple coïncidence formelle. Mais Einstein, sur la base de son interprétation physique des probabilités thermodynamiques, croyait fermement que les fluctuations révélaient la structure dynamique sous-jacente. Il n'hésita donc pas à supposer qu'un rayonnement dilué se comportait, dans son interaction avec la matière, comme s'il était constitué de quanta indépendants. Dans le cas de l'effet photoélectrique, l'éjection d'un électron du métal par la lumière incidente n'est alors possible que si l'énergie des quanta incidents excède le travail d'extraction  $P$ . Et l'énergie cinétique des photo-électrons est la différence  $k\beta\nu - P$  entre cette énergie et ce travail. Dans des expériences qu'Einstein admirait, Philipp Lenard avait bel et bien observé l'existence d'un seuil de fréquence et avait établi que l'énergie des électrons émis au-delà de ce seuil ne dépendait que de la fréquence du rayonnement incident.

L'année suivante, Einstein montra que la loi de Planck résultait d'une extension naturelle de la thermodynamique statistique, à condition de supposer que l'énergie des résonateurs de Planck ne pouvait être qu'un multiple entier du quantum  $h\nu$ , conformément à l'hypothèse des quanta lumineux. Pour qui croit que Planck admettait la quantification discrète des résonateurs dès 1900, cette remarque d'Einstein manque d'intérêt. Mais en réalité Planck n'avait introduit les éléments d'énergie que dans un calcul d'entropie combinatoire, sans ce prononcer sur la signification dynamique de cette étape. Comme Boltzmann, il n'envisageait les énergies discrètes qu'en tant que fiction, et préférerait imaginer des « domaines élémentaires » équiprobables sur l'axe des énergies. Dans sa conférence Nobel de 1920, ce savant intègre attribua à Einstein la paternité d'une quantification vraiment *physique* [8, 32, 33] :

Bien que [la constante  $h$ ] fût indispensable – car elle seule permettait de déterminer la taille des « domaines élémentaires » de la probabilité – elle s'avéra récalcitrante à toute tentative d'intégration dans la théorie classique. . . Bientôt, l'échec de telles tentatives. . . ne laissa plus de doute : soit le quantum d'action n'était qu'une quantité fictive, soit la dérivation de la loi du rayonnement [noir] reposait sur une idée vraiment physique. . . L'expérience a décidé en faveur de la seconde alternative. Mais la science ne doit pas cette décision prompte et indubitable aux tests de la loi de distribution du rayonnement thermique, et encore moins à ma dérivation particulière de cette loi ; elle le doit aux progrès incessant des chercheurs qui se sont servi du quantum d'action dans leurs recherches. A. Einstein fit la première percée dans ce domaine.

<sup>14</sup>Dans les notations de Planck, qu'Einstein évite soigneusement,  $k\beta\nu = h\nu$ .



## 1.4 Les clefs du succès

Quelle que fût la réception immédiate des mémoires d'Einstein de 1905, nous pouvons rétrospectivement juger qu'ils inauguraient trois domaines de la physique : la théorie de la relativité, la théorie des processus stochastiques et la physique quantique. Einstein était-il seul sur ce triple front ? Se démarquait-il vraiment de la physique antérieure ? L'*annus mirabilis* 1905 est-elle l'année d'Einstein ?

Quand on y regarde de près, aucune des innovations d'Einstein n'était complètement isolée et elles s'appuyaient toutes sur des avancées antérieures de ses aînés. Cependant, dans chaque cas, Einstein dépassait ses contemporains par l'ampleur et la radicalité de ses points de vue. Bien que Poincaré proposât, en même temps qu'Einstein, une théorie de la relativité empiriquement et formellement équivalente à la théorie d'Einstein, seul Einstein se permit de redéfinir les concepts d'espace et de temps. Bien que Smoluchowski aboutît à une analyse du mouvement brownien voisine de celle d'Einstein, seul Einstein poursuivait une stratégie générale d'exploration du microcosme à l'aide des fluctuations thermodynamiques ; les « révolutionnaires » quanta lumineux en sont un des remarquables produits. Même si, à cette époque des théoriciens comme Lorentz, Ehrenfest et Planck commençaient à douter de la capacité de la physique classique à rendre compte du rayonnement noir, Einstein fut le premier à afficher une certitude à cet égard. Pendant longtemps, il fut seul à croire que la nature continue de l'énergie rayonnante était en question. L'originalité et la profondeur de ses vues ne fait donc pas de doute, même si l'on fait la part du mythe.

Il est naturel de s'interroger sur les clefs d'un tel succès. Si tant est que l'on puisse les identifier, la meilleure façon de le faire est sans doute de comparer les approches d'Einstein et de ses contemporains sur des problèmes similaires. Ainsi se dégagent quelques singularités.

Il y a d'abord, de la part du jeune Einstein, un certain refus du formalisme et une volonté de donner un sens physique à tous les éléments formels d'une théorie. À une époque où l'électrodynamique de Lorentz tendait à devenir, dans les mains de son auteur et dans celles des mathématiciens de Göttingen, un lieu d'exhibition de virtuosité mathématique, Einstein préférait s'interroger sur la signification physique des concepts de base. De même, Einstein donnait un sens physique aux fluctuations thermodynamiques que ses prédécesseurs préféraient noyer dans un formalisme probabiliste partiellement interprété.

On est aussi frappé par la conscience épistémologique du jeune Einstein. Il n'était certes pas seul, au tournant du siècle, à avoir lu Hume, Mach, Boltzmann et Poincaré ; et il n'était pas seul à comprendre la faillibilité de nos concepts les mieux établis. Mais nul ne se défiait plus que lui des dogmes de la physique contemporaine. Il s'interrogeait sans cesse sur l'ancrage empirique des concepts fondamentaux de la physique, tantôt pour les réviser (l'espace et le temps), tantôt pour les consolider (les atomes). Il exigeait des théories physiques qu'elles représentent les phénomènes de manière univoque,

sans redondance inutile, avec le maximum d'homogénéité et de cohésion structurelle. C'est sur ce point que sa version de la théorie de la relativité diffère de celles de Poincaré et de Lorentz [9].

Einstein se singularisait aussi par sa volonté de se placer systématiquement dans les zones les plus turbulentes de la physique de son époque. Il méprisait toute routine et cherchait à trancher les grandes controverses à propos de l'existence des atomes et du rapport entre l'éther et la matière. Il était attentif aux faits étranges comme les expériences de Lenard sur l'effet photoélectrique, les violations de la loi de Dulong et Petit pour les chaleurs spécifiques des solides, l'expérience de Michelson et Morley, ou encore le mouvement brownien. Il suivait de près les innovations théoriques comme celles de Drude pour la conduction métallique, de Lorentz pour la théorie de l'électron et de Planck pour le corps noir. Il n'avait pas peur de mener simultanément des réflexions sur plusieurs fronts apparemment éloignés les uns des autres.

Enfin, Einstein recherchait une unité et une simplicité fondamentales de la physique. C'est sans doute en cela qu'il était le moins original. Les grands maîtres de cette époque partageaient cette quête. Presque tous approuvaient l'idée de Maxwell et de Helmholtz de soumettre l'ensemble de la physique au principe de moindre action. Certains, comme Lorentz et surtout Max Abraham cherchaient à donner un fondement électromagnétique à l'ensemble de la physique, y compris la mécanique et la théorie de la gravitation. D'autres, comme Wilhelm Ostwald, pensaient tout pouvoir déduire du principe de conservation de l'énergie. Drude et Planck se flattaient d'établir des ponts entre l'optique, l'électromagnétisme et la théorie de la chaleur, l'un grâce à l'idée du gaz d'électron dans les métaux, l'autre par une étude thermostatique du couplage entre oscillateurs et rayonnement. Suivant une distinction de Poincaré reprise par Einstein, certains préféraient unifier la physique par des principes généraux issus de l'expérience, d'autres par la construction d'une image hypothétique du monde [5].

Ce qui singularise Einstein dans sa quête unitaire, c'est plutôt la liberté de choix du mode d'unification. Dans le cas de l'électrodynamique des corps en mouvement, où l'approche constructive dominait alors, il préféra une théorie encadrée par de nouveaux principes. Pour la thermodynamique, il défendit la démarche constructive de Boltzmann, tout en envisageant une révision radicale de la dynamique sous-jacente. Soulignons aussi la tendance d'Einstein à penser l'unité de manière concrète, en imaginant des processus physiques qui mêlent diverses forces et diverses échelles. Par exemple, dans ses mains les résonateurs idéaux de Planck devenaient des oscillateurs moléculaires réels susceptibles d'entrer en collision avec les molécules d'un gaz environnant, ou encore d'interagir dans un réseau cristallin. Cette vision concrète excluait le genre de violation de l'équipartition de l'énergie admis par Planck ; et elle permettait de mettre en rapport chaleurs spécifiques et spectres d'absorption. Afin de mêler les échelles microscopique et macroscopique, Einstein introduisait d'autres dispositifs concrets : le corps noir de dimensions comparables à la longueur d'onde moyenne, et la suspension brownienne.

Ces quelques remarques et l'histoire qui précède permettent de saisir l'unité du triptyque de 1905. L'article sur les quanta lumineux et celui sur le mouvement brownien procèdent évidemment du même principe de la réalité physique des fluctuations thermodynamiques. Dans le premier cas, une estimation thermodynamique de telles fluctuations permet d'inférer la structure du rayonnement responsable ; dans le second cas l'observation des fluctuations permet d'accéder au nombre d'Avogadro, tout en validant l'interprétation statistique du second principe de la thermodynamique. En tête de l'article sur les quanta lumineux, Einstein critiquait l'asymétrie classique entre rayonnement ondulatoire et matière corpusculaire, et voyait dans l'introduction d'aspects corpusculaires du rayonnement un premier pas vers une image plus homogène du monde. Cette belle unité s'étend plus difficilement à l'article sur la relativité, qui répondait à des préoccupations de nature différente. On peut cependant remarquer que dans ce travail Einstein ouvrait la porte aux conceptions quantiques, car sa nouvelle cinématique était indépendante de la théorie de Maxwell-Lorentz et sa dynamique relativiste s'affranchissait de tout modèle constitutif de l'électron [34].

Cette unité des travaux de 1905 prit une forme concrète en 1909, quand Einstein présenta l'analyse du mouvement brownien d'un miroir immergé dans du rayonnement thermique. Cette étude impliquait l'analyse des processus stochastiques inaugurée dans l'article sur le mouvement brownien ; elle faisait aussi intervenir la relativité et les transformations de Lorentz dans le calcul du freinage radiatif du miroir ; et enfin elle conduisait à attribuer une partie des fluctuations de quantité de mouvement du miroir à l'émission et à l'absorption de quanta lumineux [35].

## 1.5 Trois réceptions

L'unité voulue du triptyque d'Einstein n'empêcha pas ses lecteurs d'en juger les composantes séparément. L'article sur le mouvement brownien connut un succès presque immédiat et il suscita les travaux de Perrin qui levèrent les derniers doutes sur la réalité des atomes. L'article sur la relativité conquiert assez rapidement les spécialistes de l'électrodynamique des corps en mouvement, grâce au soutien de Planck et grâce à l'élégance mathématique des contributions d'un ancien professeur d'Einstein, Hermann Minkowski. Mais certains experts, dont Lorentz, jugèrent inutile l'interprétation radicale que donnait Einstein des transformations de Lorentz. La relativité ne fut vraiment populaire qu'après que l'expédition d'Arthur Eddington permit, à la faveur d'une éclipse totale du soleil en 1919, de vérifier la déviation des rayons lumineux prédite par la relativité générale [2, 3].

Le plus contesté des articles d'Einstein fut sans doute celui sur les quanta lumineux, comme en témoigne le jugement formulé par Planck à l'occasion de la candidature d'Einstein à l'Académie de Berlin [36] :

En bref, on peut dire que, parmi les grands problèmes dont la physique moderne abonde, il n'en est guère qu'Einstein n'ait marqué de sa contribution. Il est vrai qu'il a parfois manqué le but lors de ses spéculations, par exemple avec son hypothèse des quanta lumineux ; mais on ne saurait lui en faire le reproche, car il n'est pas possible d'introduire des idées réellement nouvelles, même dans les sciences les plus exactes, sans parfois prendre des risques.

Les arguments d'Einstein en faveur des quanta lumineux n'avaient pas le pouvoir de conviction que les physiciens d'aujourd'hui leur attribuent. Ses plus éminents collègues, parmi lesquels Planck, Lorentz et Bohr, crurent y détecter des failles. L'optique ondulatoire était alors la mieux corroborée des théories physiques et il leur semblait difficile, voire impossible, de rendre compte du phénomène d'interférence dans une théorie corpusculaire de la lumière. Einstein ne savait lui-même résoudre les paradoxes liés à la double nature, corpusculaire et ondulatoire, du rayonnement. Jusqu'au début des années vingt, il fut le seul physicien de renom à croire aux quanta lumineux, et non sans quelques moments de doute (sur lesquels nous reviendrons dans un moment) [8, 26, 31].

Einstein joua pourtant un rôle capital dans l'émergence d'une nouvelle physique quantique : en contribuant à la conviction croissante d'un échec de l'électrodynamique classique dans le problème du corps noir, en établissant les lois de l'effet photoélectrique et en donnant, en 1907, une théorie quantique des chaleurs spécifiques des solides. Ce dernier point est particulièrement important, car il suscita un important programme expérimental de la part du thermochimiste Walther Nernst et aussi parce qu'Einstein y offrait une conception claire et radicale de la quantification de la dynamique moléculaire. Lors du premier congrès Solvay, dont Nernst détermina le thème quantique, bien des discussions tournèrent autour des idées lancées par le jeune Einstein. Nul n'approuva les quanta lumineux, mais tous admirent l'existence d'un nouveau type de discontinuité en physique, dont la nature restait fort obscure mais dont la nécessité générale ne faisait plus de doute [26, 31, 37, 38].

En somme, les travaux d'Einstein de 1905 inauguraient une nouvelle physique relativiste, atomiste et quantique. Ils marquaient aussi un tournant dans la carrière de ce jeune physicien, de l'identification des grands problèmes de son époque à la mise en place de puissantes stratégies de résolution. Au cours du demi-siècle qu'il lui restait à vivre, il déploya ces nouveaux moyens pour faire reculer aussi loin que possible les frontières de l'inconnu. Deux grandes questions s'imposèrent à son esprit : Comment construire une théorie relativiste de la gravitation ? Comment comprendre les mystérieux quanta ?

## 1.6 L'esprit relativiste

Quand Einstein et Poincaré, en 1905, traitaient de l'électrodynamique des corps en mouvement, ils rivalisaient avec les partisans d'une réduction de toute la physique, y compris la gravitation, à l'électromagnétisme. Et ils exigeaient

tous deux que toute théorie respectât la symétrie du groupe de Lorentz. Il ne faut donc pas s'étonner que Poincaré ait aussitôt tenté une modification covariante de Lorentz de la loi de Newton et qu'Einstein l'ait suivi de peu sur ce terrain, dans un article de synthèse de 1907. Einstein pensa d'abord à une théorie dans laquelle les graves seraient soumis à des forces dérivant d'un potentiel retardé proportionnel à leur masse (au repos). Mais il se rendit vite compte que dans ce cadre la composante verticale de l'accélération d'un grave en chute libre dépendait de sa vitesse horizontale, puisque la masse inertielle augmentait avec la vitesse sans que la masse gravitationnelle en dépendît. Cela contredisait un fameux principe de Galilée ainsi que l'égalité jusque là admise des masses inertielles et gravitationnelles [1, 2, 39].

Einstein eut alors « l'idée la plus heureuse de toute [sa] vie »<sup>15</sup> : il admit de manière générale que tous les corps éprouvaient la même accélération dans un champ de gravitation et en inféra que dans un champ homogène tous les mouvements se déroulaient comme ils le feraient en l'absence de champ par rapport à un système de référence en mouvement uniformément accéléré. À travers cette hypothèse, qu'Einstein nomma plus tard « principe d'équivalence », tout champ de gravitation peut être localement compris comme un effet de perspective lié au mouvement (accéléré) du système de référence, de même qu'en relativité restreinte tout champ électromoteur d'induction peut-être compris comme un effet de perspective lié au mouvement (uniforme) du système de référence. Le principe de relativité se trouvait ainsi étendu à des systèmes de référence non galiléens.

En 1907, dans une éblouissante série d'expériences de pensée, Einstein se servit du principe d'équivalence pour montrer qu'une horloge plongée dans un potentiel gravitationnel  $\Phi$  marchait  $1 + \Phi/c^2$  fois plus vite qu'une horloge éloignée de toute masse. Il en conclut que la lumière provenant de la surface du soleil devait subir un décalage vers le rouge de l'ordre de deux millièmes. Il montra aussi que la vitesse de la lumière était affectée du même facteur  $1 + \Phi/c^2$  et en conclut, par analogie avec un milieu transparent d'indice variable, que les rayons lumineux devaient être déviés lors de leur passage à proximité d'une masse importante. R.V. Pound et ses collaborateurs ne vérifièrent la première prédiction qu'en 1965 (par spectroscopie Mössbauer). Bien plus tôt, en 1919, l'expédition susmentionnée d'Arthur Eddington permit de confirmer la seconde et fit d'Einstein une vedette des médias [1, 2].

Pendant les quatre années qui suivirent les prédictions de 1907, Einstein cessa de publier sur la gravitation, sans doute parce que les quanta le préoccupaient alors davantage. Quant il y revint, en 1911, ce fut d'abord pour constater un certain vide conceptuel. Le principe d'équivalence suggérait une grande liberté dans le choix des coordonnées d'espace-temps. Mais alors ces coordonnées semblaient perdre toute signification physique : quel rapport pouvaient-elles avoir avec les mesures d'espace et de temps ? Par ailleurs, Einstein avait appris, dans le contexte des modèles rigides de l'électron, qu'une rotation

---

<sup>15</sup>Cette expression, extraite d'un manuscrit d'Einstein de 1919, est citée dans [2], p. 178.

uniforme du système de référence détruisait le caractère euclidien de la géométrie. En effet, le rapport de la circonférence au diamètre n'est plus égal à  $\pi$ , du fait que les règles tangentes à la circonférence sont soumises à la contraction de Lorentz alors que celles placées le long du diamètre ne le sont pas. Einstein savait aussi que dans le formalisme Minkowskien de la relativité restreinte, l'intervalle  $ds^2 = dx^\mu dx_\mu$  associé à une variation  $dx_\mu$  des coordonnées généralisait la distance euclidienne et jouait un rôle essentiel dans l'écriture lagrangienne des équations du mouvement d'un point matériel [2, 10 ; vol. 2].

En 1912, inspiré par une analogie avec la théorie des surfaces de Carl Friedrich Gauss, il eut l'idée d'identifier l'espace temps à une variété de métrique localement minkowskienne

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (1.21)$$

Il apprit alors de son ami mathématicien Marcel Grossmann qu'il existait une théorie mathématique de telles variétés, due à Bernhard Riemann. Grâce au calcul différentiel absolu de Tullio Levi-Civita, les deux hommes transposèrent les équations minkowskiennes de la relativité restreinte en équations covariantes sur une variété riemannienne. Puis ils cherchèrent à deviner les équations différentielles régissant la métrique  $g_{\mu\nu}$ . En raison du lien déjà connu entre le coefficient  $g_{00}$  de dilatation du temps et le potentiel de gravitation  $\Phi$ , ils pensaient à une généralisation covariante de l'équation de Poisson  $\Delta\Phi + \kappa\rho = 0$ , faisant intervenir des dérivées du deuxième ordre de  $g_{\mu\nu}$  et le tenseur d'énergie-impulsion  $T_{\mu\nu}$  (dont la composante  $T_{00}$  est une densité d'énergie qui, divisée par  $c^2$ , donne une masse gravitationnelle) [1, 2, 10 ; vol. 2, 40].

Dès 1913, ils considérèrent la possibilité

$$R_{\mu\nu} = \frac{\kappa}{c^4} T_{\mu\nu} \quad (1.22)$$

où  $R_{\mu\nu}$  et  $R$  sont obtenus par contraction du tenseur de courbure de Riemann  $R_{\mu\nu\rho\sigma}$ .

Mais ils pensaient, à tort, que la limite newtonienne de cette équation ne redonnait pas l'équation de Poisson. « Le cœur lourd », ils levèrent cette difficulté en adoptant une équation qui n'était plus covariante que sous des transformations linéaires de coordonnées. Einstein comprenait combien cette restriction s'écartait de l'esprit du principe d'équivalence. Mais il crut pouvoir la justifier par des arguments physiques fondés sur son imparfaite compréhension du rôle des coordonnées dans la description d'une variété. Ce n'est qu'à la fin de 1915 qu'il comprit qu'une légère modification de l'équation (22),

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{\kappa}{c^4} T_{\mu\nu} \quad (1.23)$$

convenait parfaitement à son problème. En 1916, il publia la version définitive de la relativité générale ainsi que la preuve que cette théorie rendait compte d'une partie longtemps inexpliquée de l'avance du périhélie de Mercure [41].

Ce succès et plus encore la beauté de la théorie, convainquirent Einstein de sa vérité. À partir des années vingt, la plupart de ses recherches s'en inspirèrent. Sur la trace d'Eddington et de Hermann Weyl, il tenta diverses extensions, à cinq dimensions ou à connexion non métrique, en vue d'une théorie unitaire de la gravitation et de l'électromagnétisme. Il chercha aussi des solutions singulières des équations du champ qui puissent représenter des particules élémentaires. Jamais satisfait par la mécanique quantique, il espéra un temps expliquer la quantification par une surdétermination des équations des champs gravifiques et électromagnétiques. Jusqu'à la fin de sa vie, il rêva de réduire toutes les incongruités quantiques à une théorie des champs dans l'esprit de la relativité générale. Aucune de ses nombreuses tentatives n'aboutit au résultat désiré, malgré l'aide de puissants mathématiciens [2, 10 vol. 2].

À peine la théorie de la relativité générale fut-elle achevée, qu'Einstein comprit son potentiel cosmologique. Les équations d'Einstein introduisent un couplage global entre la géométrie de l'univers et la distribution de ses masses. Elles posent donc spontanément la question cosmologique de la structure d'ensemble de l'Univers. Dès 1917, Einstein se prononça pour un modèle statique, clos et homogène. L'homogénéité (à grande échelle) de la distribution de masses s'imposait par sa simplicité. La clôture de l'espace sur lui-même résultait du principe de Mach, selon lequel l'inertie ne pouvait dépendre que de la distribution de masses (dans un univers infini, l'inertie dépendrait aussi des propriétés asymptotiques de l'espace). Quant au caractère statique de l'Univers, il était compatible avec les données observationnelles de l'époque. Mais Einstein se rendit vite compte que ses équations n'admettaient pas de solution non triviale satisfaisant ces trois conditions. Il se résigna donc à introduire un nouveau terme  $-\lambda g_{\mu\nu}$ , dit cosmologique, dans le premier membre de l'équation (23) [10 ; vol. 2, 42].

La suite des événements ne fut guère favorable à cette hypothèse. En 1922, le mathématicien russe Alexander Alexandrovitch Friedman découvrit des solutions cosmologiques dynamiques des équations d'Einstein sans constante cosmologique. En 1923, Edwin Powel Hubble établit la loi de récession des galaxies qui porte son nom. En 1927, Georges Lemaître proposa un modèle d'univers en expansion. Enfin, en 1930 Eddington démontra l'instabilité de la solution statique d'Einstein. Secoué par cette fatale série, en 1931 Einstein se rétracta [44] :

Dans l'état actuel de nos connaissances, le fait d'une densité de matière différente de zéro ne doit pas être mis en relation avec une courbure spatiale, mais avec une expansion spatiale. Naturellement, nous ne voulons pas dire par là qu'une telle courbure (positive ou négative) n'existe pas. Mais nous n'avons, pour le moment, aucun indice de son existence.

Par là-même, Einstein renonçait au principe de Mach. Comme il le reconnut plus tard, ce principe n'a vraiment de sens que dans la mesure où l'espace n'a pas de réalité indépendante des corps qui le meublent. Mais la relativité générale affirme justement le caractère physique et dynamique de l'espace [43].

Le sort de la constante cosmologique illustre bien la difficulté des applications cosmologiques de la relativité générale, même pour son génial fondateur. L'impénétrabilité des mathématiques impliquées poussa souvent Einstein à injecter une part d'intuition dans ses déductions, ou à généraliser un peu hâtivement des résultats partiels. Ainsi, en 1939, il crut pouvoir affirmer l'impossibilité de l'effondrement gravitationnel d'un nuage de poussières (de symétrie sphérique) au-delà du rayon de Schwarzschild. Il croyait par là démontrer l'impossibilité de construire physiquement la « singularité » de la solution sphériquement symétrique que Karl Schwarzschild avait donné aux équations d'Einstein dès 1915. Mais l'analyse d'Einstein admettait le caractère statique du champ engendré par les poussières. Dans la même année, John Robert Oppenheimer et Hartland Snyder, qui ne faisaient pas cette supposition gratuite, aboutirent à la conclusion contraire en montrant que les particules du nuage traversaient le rayon de Schwarzschild lors de l'effondrement. Il fallut néanmoins attendre les années soixante pour que se développe l'étude des « trous noirs ». Einstein n'était plus là ([10] vol. 3 : p. 58).

## 1.7 L'énigme quantique

Dès 1905, Einstein pensait faire porter la quantification sur le rayonnement-même, plutôt que sur le mécanisme de son interaction avec la matière. Néanmoins, dans les trois années suivantes, il préféra explorer les conséquences de la quantification de la matière en développant une nouvelle théorie des chaleurs spécifiques des solides. Il revint à la structure du rayonnement en 1909, à l'occasion d'une conférence prononcée à Salzbourg devant l'élite de la physique théorique de l'époque. Rappelons qu'en 1905 sa suggestion des quanta lumineux s'appuyait sur une étude des fluctuations d'un rayonnement noir de faible densité (obéissant à la loi de Wien). En 1909, il s'affranchit de cette limitation en appliquant à un rayonnement distribué suivant la loi de Planck la formule (3) pour les fluctuations quadratiques moyennes de l'énergie d'un système thermostaté. Cela donne

$$\overline{(E - \bar{E})^2} / \bar{E}^2 = 1/(\bar{E}/h\nu) + 1/(8\pi V\nu^2 d\nu/c^3) \quad (1.24)$$

où  $E$  désigne l'énergie du rayonnement contenu dans le volume  $V$  et dans l'intervalle spectral  $d\nu$  [26, 31, 45, 46].

Le premier terme du second membre est l'inverse du nombre moyen de quanta  $h\nu$  dans ce rayonnement, et s'explique aisément selon l'image d'un gaz de quanta lumineux indépendants. En revanche, le second terme, où l'on reconnaît l'inverse du nombre de modes propres d'une cavité de volume  $V$  dans l'intervalle spectral  $d\nu$ , correspond aux interférences d'un rayonnement ondulatoire dont les phases sont aléatoirement distribuées. Plus concrètement, dans l'argument du « miroir brownien » mentionné plus haut, Einstein montrait que les fluctuations de vitesse d'un miroir plongé dans du rayonnement



noir se composaient encore de deux termes, l'un ondulatoire, l'autre corpusculaire. Ce résultat lui permit d'augurer comme suit [46] (p. 817) :

Je pense que la prochaine phase du développement de la physique théorique nous apportera une sorte de fusion de la théorie ondulatoire et de la théorie de l'émission [corpusculaire].

Einstein venait lui-même d'explorer une théorie des champs dans laquelle électrons et quanta lumineux devait apparaître comme des singularités influencées par le champ environnant. L'échec de telles tentatives, et les nombreuses objections reçues d'autres théoriciens (surtout de Lorentz), l'amènèrent à douter des quanta lumineux pour quelques temps. En mai 1911, il écrivit à son ami Michele Besso : « *Je ne me demande plus si ces quanta existent réellement. Et je n'essaie plus de les construire car je sais maintenant que mon cerveau n'en est pas capable* ». En décembre 1911, une lettre à Ludwig Hopf le montre encore plus sceptique : « *Les quanta certes marchent comme ils devraient, mais ils n'existent pas plus... que l'éther immobile*<sup>16</sup> ».

Dans les deux années suivantes, Einstein travailla surtout à la théorie de la gravitation. Dans le domaine quantique, il se contenta de montrer que certains résultats réputés quantiques pouvaient s'obtenir sans aucune quantification. C'est probablement le succès de la théorie atomique de Niels Bohr, publiée en 1913, qui le fit retourner aux quanta purs et durs. Bohr fondait sa théorie sur l'hypothèse qu'un atome ou une molécule ne pouvait durablement exister que dans une série discrète d'états stationnaires et qu'il ne pouvait interagir avec le rayonnement que par des sauts brusques d'un tel état à un autre. De surcroît, il admettait qu'en première approximation la mécanique ordinaire s'appliquait à l'interaction coulombienne entre les électrons et le noyau d'un atome. Afin de retrouver le spectre de l'atome d'hydrogène, il eut enfin l'audace d'admettre que la fréquence du rayonnement émis ou absorbé dans un saut quantique n'était plus donnée par la fréquence de rotation de l'électron dans l'état initial mais s'obtenait en divisant la variation d'énergie de l'atome par la constante de Planck [8, 31, 48].

Einstein vit dans cette conception des interactions entre atome et rayonnement une opportunité pour une théorie statistique qui lui permit de résoudre une contradiction persistante de la théorie du rayonnement noir. Les dérivations antérieures de la loi de Planck maintenaient la relation classique  $u_\nu = (8\pi\nu^2/c^3)U$  entre l'énergie moyenne  $U$  d'un résonateur électrique et la densité spectrale énergétique  $u_\nu$  du rayonnement noir, tout en quantifiant l'énergie de ce résonateur pour les besoins du calcul d'entropie. Pour lever cette contradiction, il fallait une nouvelle théorie quantique de l'interaction entre résonateur et rayonnement. C'est ce qu'Einstein se proposa de faire en 1916, dans le cas plus général d'un atome de Bohr émettant et absorbant du rayonnement [8, 31, 49, 50].

---

<sup>16</sup>Einstein à Besso [mai 1911], dans « Einstein à Hopf » [décembre 1911], Einstein Archive, Jerusalem.

Par analogie avec un oscillateur classique rayonnant spontanément, ou interagissant avec un rayonnement accordé en phase ou en opposition de phase, Einstein admit trois processus quantique d'interaction : émission spontanée, émission induite et absorption, de probabilités par unité de temps données par les coefficients  $A_m^n$ ,  $\rho_\nu B_n^m$ ,  $\rho_\nu B_m^n$ , où  $n$  et  $m$  dénotent les deux niveaux d'énergie concernés, et  $\rho_\nu$  la densité d'énergie du rayonnement de fréquence  $\nu$  accordée à la transition entre ces deux niveaux. L'équilibre cinétique d'un ensemble d'atomes interagissant avec du rayonnement suivant ces lois statistiques n'est alors compatible avec leur équilibre thermodynamique que si  $B_n^m = B_m^n$ ,  $A_m^n/B_m^n = 8\pi h\nu^3/c^3$ ,  $E_n - E_m = h\nu$  et  $\rho_\nu = u_\nu$  (loi de Planck). Einstein obtenait ainsi une dérivation purement quantique de la loi de Planck, avec en prime deux relations entre les coefficients d'émission et d'absorption, et surtout une dérivation générale de la règle de fréquence  $\Delta E = h\nu$ , que Bohr n'avait cru vraie que pour des mouvements électroniques strictement périodiques.

Ces résultats jouèrent un rôle important dans l'évolution ultérieure de la théorie quantique, et ils continuent aujourd'hui de former la base de notre description des échanges entre atomes et rayonnement. Einstein n'y voyait pourtant qu'un pis-aller, en l'attente d'une théorie plus détaillée et non statistique des sauts quantiques. Pour lui, l'essentiel était plutôt dans leurs implications sur la structure du rayonnement. Dans une seconde partie de son étude, il montra en effet que les fluctuations de vitesse d'un atome interagissant avec du rayonnement noir n'étaient compatibles avec la valeur moyenne  $(3/2)kT$  de son énergie cinétique de translation que si le processus d'émission de rayonnement était orienté dans une direction bien définie de l'espace et comportait un échange de quantité de mouvement  $h\nu/c$ . Cette nouvelle preuve des quanta lumineux était bien plus contraignante que celles antérieurement données par Einstein, car elle ne présupposait guère que la quantification de la matière et la conservation de l'énergie et de la quantité de mouvement. De fait, Bohr et Schrödinger ne virent d'autres moyens d'y échapper que d'admettre une violation de la conservation de l'énergie dans les processus quantiques.

L'immense majorité des physiciens continuèrent d'ignorer les quanta lumineux, jusqu'à ce que Maurice de Broglie et Arthur Holly Compton, au début des années vingt, démontrent que l'énergie et la quantité de mouvement étaient bien conservées lors de processus individuels d'interaction entre atomes (ou électrons) et rayons X. D'après les expériences de Compton, lors de l'interaction entre un rayonnement monochromatique et un électron (quasi-libre) la fréquence du rayonnement secondaire dépend de la direction d'émission exactement comme le prévoit l'hypothèse des quanta lumineux. Ce résultat convainquit un nombre croissant de physiciens de la réalité des quanta lumineux, malgré l'absence d'une conciliation plausible avec la théorie ondulatoire.

Cependant, Bohr et ses collaborateurs crurent pouvoir échapper aux quanta lumineux par une étrange théorie de 1924, fondée sur l'idée d'oscillations virtuelles violant la conservation de l'énergie. Au début de 1925, d'insurmontables difficultés les conduisirent à renoncer à cette tentative.

Mais plutôt que d'accepter les quanta lumineux, Bohr décréta un échec général du mode causal et spatio-temporel de représentation dans le domaine atomique. La mécanique quantique naquit peu de temps après, d'abord sous la forme matricielle ou opératoire de Werner Heisenberg, Max Born, Pascual Jordan et Paul Dirac, puis sous la forme ondulatoire élaborée par Erwin Schrödinger à partir d'une idée de Louis de Broglie. Jordan en 1925, puis Dirac en 1927 quantifièrent le rayonnement, donnant ainsi une justification formelle aux quanta lumineux [8, 31].

Il serait naïf de croire que ces développements confirmèrent l'idée que se faisait le jeune Einstein des quanta lumineux. Pour les fondateurs de la nouvelle mécanique quantique, le concept classique de corpuscule devait être à jamais rejeté du domaine quantique. Suivant la « complémentarité » de Bohr (1927), la lumière pouvait certes manifester un comportement corpusculaire, mais seulement dans des contextes expérimentaux physiquement incompatibles avec les dispositifs susceptibles de révéler un comportement ondulatoire. De plus, la position et la quantité de mouvement d'un corpuscule ne pouvaient être simultanément déterminés dans un même dispositif. Il était donc vain de se représenter un électron ou un photon comme un corpuscule de trajectoire bien définie. La plupart des physiciens approuvèrent ce point de vue. Mais Einstein n'y trouva aucune réponse à ses vieilles inquiétudes sur la nature de la lumière. Peu de temps avant sa mort, il confiait à l'ami Besso<sup>17</sup> :

Ces cinquante ans de rumination consciente ne m'ont en rien rapproché d'une réponse à la question : « Que sont les quanta lumineux ? » Aujourd'hui, le premier fripon venu croit savoir ce qu'ils sont, mais il se leurre.

Revenons au début des années vingt, au moment où les quanta lumineux commençaient à être pris au sérieux. Einstein reçut alors d'un certain Satiendra Nath Bose un manuscrit contenant une nouvelle dérivation de la loi de Planck, entièrement corpusculaire mais promptement rejetée par la revue *Nature*. Bose traitait le rayonnement thermique comme un gaz de quanta lumineux à distribuer sur des cellules quantiques de taille  $h^3$  dans l'espace des configurations. Pour compter le nombre de distributions, il se servait d'une vieille formule combinatoire de Boltzmann, mais en intervertissant inconsciemment le rôle des cellules et celui des corpuscules. Cela conduisait à la loi de Planck via la relation de Boltzmann entre entropie et probabilité combinatoire. Einstein, peu inquiet de l'étrangeté du procédé, s'avisa que son caractère purement corpusculaire permettait une transposition au cas d'un gaz de matière. Ce faisant, et tenant compte de la constance du nombre total de molécules, il obtint la théorie dite de Bose-Einstein des gaz quantiques [31, 51, 52].

Cette théorie n'était certes pas la première qui prédit une dégénérescence quantique des gaz parfaits (dont les molécules interagissent peu). De manière générale, le théorème de Nernst (1905) impliquait que l'entropie d'un corps s'annule au zéro absolu de température. Pour les solides, la théorie d'Einstein

<sup>17</sup>Einstein à Besso, 12 décembre 1951, dans [47].

des chaleurs spécifiques satisfaisait cette contrainte. Pour des corps susceptibles de rester gazeux à très basse température, Planck pensait que la quantification du mouvement pouvait jouer un rôle similaire, à condition que les états quantiques du système global soient peu nombreux pour les énergies les plus basses. En 1916, il était parvenu à satisfaire à cette condition, et Schrödinger l'avait fait d'une manière différente en 1924. Mais leurs règles de quantification avaient quelque chose d'*ad hoc*, alors qu'Einstein pouvait justifier la sienne au nom de l'analogie entre matière et lumière.

Dans un second mémoire, Einstein décrit la condensation dite « de Bose-Einstein » et il relia le comptage de Bose à l'indiscernabilité des particules distribuées, qui supposait une « mystérieuse influence mutuelle ». Pour avoir une idée de la dynamique sous-jacente, il eut une nouvelle fois recours à un argument de fluctuation. Partant de la distribution d'énergie des molécules du gaz quantique, il obtint l'expression

$$\overline{(n_E - \bar{n}_E)^2} / \bar{n}_E^2 = 1/\bar{n}_E + 1/2\pi V h^{-3} (2m)^{3/2} E^{1/2} dE \quad (1.25)$$

pour les fluctuations quadratiques moyennes du nombre  $n_E$  de molécules (de masse  $m$ ) comprises dans le volume  $V$  et dans l'intervalle d'énergie cinétique  $dE$ . Cette formule a la même structure que la formule analogue de 1909 pour les fluctuations du rayonnement noir. Comme Einstein venait justement de lire la thèse de Louis de Broglie (que son ami Langevin avait soumise à son jugement), il reconnut dans le second terme du second membre l'inverse du nombre de modes stationnaires d'ondes brogliennes dans une cavité de volume  $V$  pour un intervalle d'énergie  $dE$ . Frappé par cette coïncidence, il écrivit à Langevin :

Le travail de Louis de Broglie m'a grandement impressionné. Il a soulevé un coin du grand voile. Dans un nouvel article, j'obtiens des résultats qui semblent confirmer les siens. Si vous le voyez, veuillez lui communiquer toute mon estime et ma sympathie.

Le jugement d'Einstein attira l'attention sur le travail de Louis de Broglie, qui eût sinon passé pour une spéculation purement formelle. L'année suivante, Schrödinger établit la mécanique ondulatoire sur cette base<sup>18</sup> [53–55].

Résumons. Bien qu'Einstein n'ait pas directement contribué à la construction de la mécanique quantique, il a fourni plusieurs des éléments de cette construction. Il a d'abord été le premier à reconnaître la nécessité d'une réforme radicale des théories classiques. Il a introduit la discontinuité quantique sous la forme utile à la théorie de Bohr. Il a donné des lois quantitatives – quoique statistiques – de l'interaction entre atomes quantifiés et rayonnement. Son insistance sur une analogie profonde entre matière et lumière a inspiré les ondes de matière de Louis de Broglie. À travers sa théorie quantique des gaz, il a identifié un trait caractéristique de la physique quantique, l'indiscernabilité

<sup>18</sup>Einstein à Langevin, 16 décembre 1924, Einstein Archive, Jerusalem.

de particules identiques, et reconnu son lien avec les ondes brogliennes. Enfin, il eut très tôt l'audace d'accorder à la lumière des propriétés corpusculaires et anticipa ainsi la notion de photon.

Et pourtant, Einstein ne se satisfait jamais de la mécanique quantique. Il reconnut certes la justesse et la puissance explicative de cette théorie, mais il ne put jamais l'admettre comme ultime représentation des phénomènes quantiques. Selon lui, le caractère probabiliste de cette théorie impliquait qu'elle fût incomplète. De même qu'il avait réduit les probabilités thermodynamiques à des fréquences temporelles, il espérait attribuer les probabilités quantiques à notre ignorance de la dynamique complète des systèmes quantiques. De 1927 à 1930, il tenta même de violer les relations d'incertitudes de Heisenberg par des expériences de pensée, dont la plus célèbre est le pèse-photon du congrès Solvay de 1930. Après que Bohr eut spectaculairement réfuté cet argument, Einstein admit l'impossibilité de mesurer des grandeurs conjuguées avec une précision qui enfreignît les relations de Heisenberg. Mais il continua de croire en la possibilité de les définir simultanément et exactement dans une théorie plus complète<sup>19</sup>.

Le célèbre argument formulé par Einstein, Podolsky et Rosen en 1935 montrait en effet que la mécanique quantique violait le critère de complétude suivant :

Chaque élément de la réalité physique doit avoir un correspondant dans la théorie physique ; [étant entendu que] si sans perturber le système en aucune façon, nous pouvons prédire avec certitude... la valeur d'une grandeur physique, alors il existe un élément de la réalité physique correspondant.

Il suffit, pour voir la violation, d'imaginer un système de deux particules 1 et 2 ayant chacune les propriétés conjuguées  $q$  et  $p$ , et prises dans un état propre des observables  $q_1 - q_2$  et  $p_1 + p_2$ . Une mesure de  $q_1$  permettrait de déterminer  $q_2$  sans aucunement perturber la particule 2, à condition que celle-ci soit suffisamment éloignée de la particule 1. De même, une mesure de  $p_1$  permettrait de déterminer  $p_2$  sans perturber la particule 2. Le critère de réalité d'EPR implique donc que  $q_2$  et  $p_2$  aient toutes deux des valeurs définies, bien que la mécanique quantique ne les définisse que dans la limite des relations d'incertitude de Heisenberg [31, 56].

Bohr opposa à Einstein un critère de complétude moins exigeant, selon lequel le degré de définition des grandeurs théoriques devait être exactement conforme aux possibilités de mesure effective de ces grandeurs. Le dispositif EPR ne violait nullement ce critère, car seule une mesure *effective* de  $q_1$  et de  $p_1$  aurait permis d'inférer conjointement les valeurs de  $q_2$  et de  $p_2$ . Or une telle mesure était inconcevable. Autrement dit, Bohr présentait que l'incompatibilité physique de dispositifs de mesure de grandeurs mesurées en un lieu donné, impliquait l'incompatibilité d'inférences sur des propriétés corrélées

<sup>19</sup>Bohr, « Discussions with Einstein on epistemological problems in atomic physics », dans [1] pp. 199–242.

d'un objet situé loin de ce lieu. Mais il n'avait que son instinct épistémologique pour rejeter le point de vue adverse. C'est à John Bell que nous devons la preuve que le critère EPR de réalité est incompatible avec les prédictions empiriques de la mécanique quantique.

Einstein ne vécut pas assez longtemps pour connaître ce revers. Jusqu'à sa mort, il martela son rejet de la « philosophie tranquillisante » qui émanait de Copenhague. Il multiplia les exemples d'incomplétude supposée de la mécanique quantique, se trompant d'ailleurs parfois sur les conséquences de cette dernière. Plutôt que d'explorer la nouvelle physique atomique et nucléaire, il cherchait à faire surgir les quanta d'une théorie des champs plus conforme à son instinct épistémologique. Il pressentait certes de grandes difficultés et rejetait toute tentative de théorie quantique déterministe qui maintint l'équation de Schrödinger et le principe de superposition. Mais il préférait viser un horizon lointain que sombrer dans ce qu'il percevait comme un positivisme « pour les boutiquiers et les ingénieurs ». Alors que le jeune Einstein recherchait l'équilibre entre théorie et expérience et se défiait de la virtuosité mathématique, « l'ermite de Princeton » plaçait ses espoirs dans des mathématiques si difficiles qu'il ne pouvait les contrôler seul<sup>20</sup>.

## 1.8 Après la mort

Les grandes questions qui agitaient encore Einstein à la veille de sa mort, l'unification des forces fondamentales et les fondements de la physique quantique, n'ont pas reçu le type de réponse qu'il souhaitait. Mais elles n'ont cessé d'inspirer des avancées de la physique fondamentale. C'était hier que la théorie des cordes promettait enfin d'unifier gravitation et quanta d'une manière satisfaisante. C'était hier que les inégalités de John Bell, inspirées par le paradoxe d'Einstein, Podolsky et Rosen, se trouvaient violées dans le laboratoire et qu'une non-localité incompatible avec le critère de réalité d'Einstein recevait une preuve expérimentale directe. C'était hier que le chat mort-vivant de Schrödinger, inspiré par une lettre d'Einstein, hantait certains laboratoires d'optique quantique sous la forme de délicats systèmes mésoscopiques. Et c'est bien aujourd'hui que certains physiciens rêvent d'exploiter l'intrication quantique pour construire un nouveau type d'ordinateur.

Les prolongements contemporains de la physique d'Einstein ne se limitent pas à ces grandes questions. La théorie quantique du rayonnement, telle qu'Einstein la formula dans ses mémoires de 1905 et de 1916 n'a cessé de recevoir des applications dont son auteur ne pouvait même rêver. L'émission stimulée est le fondement théorique des lasers et des masers. Les échanges de quantité de mouvement entre matière et rayonnement, tels qu'Einstein les décrivait en 1916, sont à la base du refroidissement par laser, dont dépendent bien des progrès de la physique atomique récente. Enfin, c'est très récemment

---

<sup>20</sup> « Boutiquiers et ingénieurs » est dans la lettre d'Einstein à Schrödinger du 22 décembre 1950, Einstein Archive, Jerusalem, traduit dans [10] vol. 1 : pp. 250–251.

que la condensation de Bose-Einstein a reçu une confirmation expérimentale directe. Dans le registre relativiste, l'astrophysique et la cosmologie qui prennent la relativité générale pour fondement n'ont cessé de se développer, dans des directions qu'Einstein n'a pas toujours pu prévoir. En témoigne la richesse actuelle de la physique des trous noirs. Dans le registre brownien, l'étude des phénomènes stochastiques et des violations, apparentes ou réelles, du second principe de la thermodynamique, continue de s'approfondir et de se diversifier.

Cent ans après l'« *annus mirabilis* » d'Einstein, nous continuons donc de cueillir les fruits des théories qu'il contribua à développer sur le sol mouvant mais fertile de la physique du début du XX<sup>e</sup> siècle. Le bon dieu (*das liebe Gott*) qu'il évoquait volontiers quand le monde se révélait trop étrange, a peut-être eu le mauvais goût de « jouer au dés » mais ne s'est pas trop moqué de lui. Il lui a même permis de relativiser le temps de sa propre mort, comme le suggèrent ces mots écrits à la disparition, presque simultanée, de son ami Michele Besso<sup>21</sup> :

Il est vrai qu'il se sera séparé de cet étrange monde peu de temps avant moi. Cela ne signifie rien. Pour nous physiciens croyants, la distinction entre passé, présent et futur n'est guère qu'une illusion, si tenace soit-elle.

## 1.9 Sources et références

*La liste qui suit est très limitative. Pour se faire une idée plus juste des études einsteiniennes, le lecteur pourra se référer aux bibliographies des quelques ouvrages historiques cités.*

- [1] A. Einstein, *Autobiographisches, Albert Einstein: Philosopher-scientist*, P.A. Schilpp éd., Evanston, 1–94 (1949).
- [2] A. Pais, « *Subtle is the Lord...* »: *The science and the life of Albert Einstein*, Oxford (1982).
- [3] A. Fölsing, *Albert Einstein: Eine Biographie*, Frankfurt (1993).
- [4] J. Stachel, M. Klein et al. (éds.), *The collected papers of Albert Einstein*, Princeton, 1987. Voir en particulier les introductions et commentaires aux deux premiers volumes, rédigés par J. Stachel, D. Cassidy, R. Schulmann, J. Renn, D. Howard, A.J. Kox.
- [5] C. Junknickel, R. McCormach, *Intellectual mastery of nature: Theoretical physics from Ohm to Einstein*, 2 vols., Chicago (1986).
- [6] O. Darrigol, *Electrodynamics from Ampère to Einstein*, Oxford (2000).
- [7] E. Whittaker, *A history of the theories of aether and electricity*, 2 vols., London (1951).

---

<sup>21</sup>Einstein à Vero a Bice Besso, 21 mars 1955, dans [47].

- [8] O. Darrigol, *From c-numbers to q-numbers: The classical analogy in the history of quantum theory*, Berkeley (1992).
- [9] M. Paty, *Einstein philosophe : La physique comme pratique philosophique*, Paris (1993).
- [10] F. Balibar et al. (éds.), *Albert Einstein : Œuvres choisies*, 6 vols., Paris, 1989–1993. Voir en particulier les introductions rédigées par J. Stachel, J. Eisenstaedt et J. Ritter pour les volumes sur la relativité.
- [11] A. Einstein, *How I created the theory of relativity?* (conférence de Kyoto du 14 décembre 1922), reconstruit d'après les notes de Jun Ishiwara, *Physics Today* **35** (8), 45–47 (1982).
- [12] A. Miller, *Albert Einstein's special relativity: Emergence and early interpretation (1905-1911)*, Reading (1981).
- [13] W. Wien, *Über die Fragen, welche die translatorische Bewegung des Lichtäthers betreffen*, *Annalen der Physik* **65** (1898), I-XVIII (appendice).
- [14] H. Lorentz, *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*, Leiden (1895).
- [15] H. Poincaré, *La théorie de Lorentz et le principe de la réaction*, Recueil de travaux offerts par les auteurs à H.A. Lorentz à l'occasion du 25<sup>e</sup> anniversaire de son doctorat le 11 décembre 1900, *Archives Néerlandaises* **5**, 252–278 (1900).
- [16] P. Galison, *Einstein's clocks, Poincaré's maps: Empires of time*, New York (2003).
- [17] H. Lorentz, *Electromagnetic phenomena in a system moving with any velocity smaller than light*, Koninklijke Akademie van Wetenschappen te Amsterdam, section of sciences, *Proceedings* **6**, 809–831 (1904).
- [18] H. Poincaré, *Sur la dynamique de l'électron*, *Académie des sciences de Paris, Comptes-rendus hebdomadaires des séances* **140**, 1504–1508 (1905).
- [19] H. Poincaré, *Sur la dynamique de l'électron*, *Rendiconti del Circolo Matematico di Palermo* **21**, 129–175 (1906).
- [20] H. Poincaré, *Les limites de la loi de Newton*, leçons à la Sorbonne (1906–1907), H. Vergne éd., *Bulletin astronomique publié par l'Observatoire de Paris* **17**, 121–365 (1953).
- [21] A. Einstein, *Zur Elektrodynamik bewegter Körper*, *Annalen der Physik* **17**, 891–921 (1905).
- [22] A. Einstein, *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?*, *Annalen der Physik* **18**, 639–641 (1905).
- [23] A. Einstein, *Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik*, *Annalen der Physik* **9**, 417–433 (1902).
- [24] A. Einstein, *Eine Theorie der Grundlagen der Thermodynamik*, *Annalen der Physik* **11**, 170–187 (1903).
- [25] A. Einstein, *Zur allgemeinen molekularen Theorie der Wärme*, *Annalen der Physik* **14**, 354–362 (1904).



- [26] M.J. Klein, nombreux articles dont les références se trouvent dans les bibliographies de [4, 6, 8, 10].
- [27] A. Einstein, *Eine neue Bestimmung der Moleküldimensionen*, Inaugural-Dissertation, Universität Zürich, Bern (1905).
- [28] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, *Annalen der Physik* **17**, 549–560 (1905).
- [29] M. Planck, *Über das Gesetz der Energieverteilung im Normalspektrum*, *Annalen der Physik* **4**, 553–563 (1901).
- [30] A. Einstein, *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, *Annalen der Physik* **17**, 132–148 (1905).
- [31] M. Jammer, *The conceptual development of quantum mechanics*, New York (1966).
- [32] A. Einstein, *Zur Theorie der Lichterzeugung und Lichtabsorption*, *Annalen der Physik* **20**, 199–206 (1906).
- [33] M. Planck, *Die Entstehung und bisherige Entwicklung der Quantentheorie* (conférence Nobel, 1920) dans *Physikalische Abhandlungen und Vorträge*, Braunschweig, 1958, vol. 3, 121–136, à la p. 127.
- [34] G. Holton, *Thematic origins of scientific thought: Kepler to Einstein*, Cambridge (1973).
- [35] A. Einstein, *Zum gegenwärtigen Stand des Strahlungsproblem*, *Physikalische Zeitschrift* **10**, 185–193 (1909).
- [36] C. Kirsten, H.J. Treder (éds.), *Albert Einstein in Berlin 1913–1934, Darstellungen und Dokumente*, Berlin (1979), p. 96.
- [37] A. Einstein, *Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme*, *Annalen der Physik* **22**, 180–190 (1907).
- [38] P. Langevin, L. de Broglie (éds.), *La théorie du rayonnement et les quanta. Rapports et discussions de la réunion tenue à Bruxelles, du 30 octobre au 3 novembre 1911*. Sous les auspices de M.E. Solvay, Paris (1912).
- [39] A. Einstein, *Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen*, *Jahrbuch der Radioaktivität und Elektronik* **4**, 411–462 (1907).
- [40] A. Einstein, M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*, *Zeitschrift für Mathematik und Physik* **62**, 225–261 (1913).
- [41] A. Einstein, *Grundlage der allgemeinen Relativitätstheorie*, *Annalen der Physik* **49**, 769–822 (1916).
- [42] A. Einstein, *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*, *Preussische Akademie der Wissenschaften zu Berlin, Mathematisch-Physikalische Klasse, Sitzungsberichte*, 142–152 (1917).
- [43] A. Einstein, *Zum kosmologischen Problem der allgemeinen Relativitätstheorie*, *Preussische Akademie der Wissenschaften zu Berlin, Mathematisch-Physikalische Klasse, Sitzungsberichte*, 235–237 (1931).

- [44] A. Einstein, *Les fondements de la théorie de la relativité générale*, Paris (1933), pp. 99–109.
- [45] A. Einstein, *Zum gegenwärtigen Stand des Strahlungsproblem*, *Physikalische Zeitschrift* **10**, 185–193 (1909).
- [46] A. Einstein, *Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung*, *Physikalische Zeitschrift* **10**, 817–825 (1909).
- [47] P. Speziali (éd.), *Albert Einstein-Michele Besso : Correspondance 1903–1955*, Paris, Hermann (1972).
- [48] N. Bohr, *On the constitution of atoms and molecules*, *Philosophical magazine* **26**, 1–25, 476–502, 857–875 (1913).
- [49] A. Einstein, *Zur quantentheorie der Strahlung*, *Physikalische Gesellschaft (Zürich), Mitteilungen* **16**, 47–62 (1916).
- [50] A. Einstein, *Physikalische Zeitschrift* **18**, 121–128 (1917).
- [51] A. Einstein, *Quantentheorie des einatomigen Gases*, *Preussische Akademie der Wissenschaften zu Berlin, Mathematisch-Physikalische Klasse, Sitzungsberichte*, 261–267 (1924).
- [52] O. Darrigol, *Statistics and combinatorics in early quantum theory II: Early symptoms of indistinguishability and holism*, *HSPS* **21**, 237–298 (1991).
- [53] A. Einstein, *Quantentheorie des idealen Gases*, *Preussische Akademie der Wissenschaften zu Berlin, Mathematisch-Physikalische Klasse, Sitzungsberichte*, 18–25 (1925).
- [54] L. de Broglie, *Recherches sur la théorie des quanta*, Paris, Masson (1924).
- [55] O. Darrigol, *Strangeness and soundness in Louis de Broglie's early works*, *Physis* **30**, 303–372 (1993).
- [56] A. Einstein, B. Podolsky, N. Rosen, *Can quantum-mechanical description of physical reality be considered complete*, *Physical review* **47**, 777–780 (1935).

# Chapitre 2

## De l'article d'Einstein Podolsky et Rosen à l'information quantique : les stupéfiantes propriétés de l'intrication

Par *Alain Aspect* et *Philippe Grangier*

### Sommaire

<b>2.1</b>	<b>Introduction : Einstein et la révolution quantique . . . . .</b>	<b>43</b>
<b>2.2</b>	<b>L'émergence de l'intrication quantique : de l'article EPR à la violation des inégalités de Bell . . . . .</b>	<b>44</b>
2.2.1	Le débat Bohr-Einstein et l'article EPR . . . . .	44
2.2.2	Paires de photons EPR corrélés en polarisation . . . . .	48
2.2.3	Le théorème de Bell . . . . .	52
2.2.4	Condition de localité et expérience avec polariseurs variables . .	56
2.2.5	Le recours à l'expérience . . . . .	58
2.2.6	Les expériences de deuxième génération . . . . .	60
2.2.7	Vers une expérience idéale : la chasse aux échappatoires . . . .	63
2.2.8	Conclusion : la non localité quantique . . . . .	67
<b>2.3</b>	<b>L'information quantique . . . . .</b>	<b>70</b>
2.3.1	Introduction . . . . .	70
2.3.2	Paires de photons et cryptographie quantique . . . . .	70
2.3.3	Expériences de cryptographie quantique . . . . .	73
2.3.4	L'ordinateur quantique . . . . .	76
2.3.5	Conclusion . . . . .	81

<b>2.4</b>	<b>Appendices</b> . . . . .	<b>81</b>
2.4.1	Le théorème de non-clonage . . . . .	81
2.4.2	L'algorithme de Shor . . . . .	82
2.4.3	La téléportation quantique . . . . .	83
<b>2.5</b>	<b>Références</b> . . . . .	<b>85</b>



### Alain Aspect

Alain Aspect est directeur de recherche au CNRS, au laboratoire Charles Fabry de l'Institut d'Optique à Orsay, et professeur à l'École Polytechnique. Il est membre de l'Académie des sciences et de l'Académie des technologies. Il a effectué des expériences sur les fondements de la mécanique quantique (tests des inégalités de Bell), puis travaillé avec Claude Cohen-Tannoudji sur le refroidissement d'atomes par laser. Il dirige actuellement le groupe « Optique Atomique » de l'Institut d'Optique, dont les recherches portent sur les condensats de Bose-Einstein et les lasers à atomes.



### **Philippe Grangier**

Philippe Grangier est directeur de recherche au CNRS, responsable du groupe « Optique Quantique » au laboratoire Charles Fabry de l'Institut d'Optique, Orsay. Il a réalisé de nombreux travaux sur les propriétés quantiques de la lumière, concernant par exemple la réalisation expérimentale d'états comprimés, d'états à un seul photon, et de mesures quantiques non destructives. Il poursuit actuellement des recherches dans le domaine de l'information quantique et de la cryptographie quantique, en utilisant en particulier des pièges optiques microscopiques permettant de manipuler des atomes individuels.



# De l'article d'Einstein Podolsky et Rosen à l'information quantique : les stupéfiantes propriétés de l'intrication

## 2.1 Introduction : Einstein et la révolution quantique

Le développement de la physique quantique, au XX<sup>e</sup> siècle, est l'une des aventures intellectuelles les plus extraordinaires de l'humanité. Cette théorie physique a modifié de fond en comble notre conception du monde, puisqu'elle nous force, par exemple, à accepter qu'une particule puisse se trouver à la fois ici et là, où qu'une porte quantique puisse être à la fois ouverte et fermée. Et elle a tout autant bouleversé notre mode de vie avec ses innombrables applications : lasers, transistors, circuits intégrés, qui sont les composants de base des technologies de l'information et de la communication.

On connaît le rôle séminal d'Einstein dans l'émergence de la théorie quantique puisque dès 1905, il proposait d'interpréter l'effet photoélectrique en introduisant explicitement la quantification de la lumière : en la considérant comme formée de grains indivisibles d'énergie, les photons, il obtenait un certain nombre de prédictions précises qui allaient être vérifiées expérimentalement par Millikan. L'importance de ce travail est soulignée par le prix Nobel de 1922 qu'Einstein reçut explicitement pour l'effet photoélectrique, c'est-à-dire pour la quantification du rayonnement lumineux.

Mais on sait aussi que dès la fin des années 1920 Einstein s'est opposé à Niels Bohr sur l'interprétation à donner à la théorie quantique. Le débat a duré jusqu'à la fin de la vie de ces géants, se concentrant à partir de 1935 sur la question soulevée dans le fameux article d'Einstein, Podolsky et Rosen, qui présente pour la première fois une situation où la mécanique quantique prédit des corrélations particulièrement fortes entre deux particules éloignées, préparées dans un état quantique autorisé par la théorie mais aux propriétés surprenantes, appelé aujourd'hui « état intriqué » ou « état EPR ».

De façon inattendue, tant était grand chez la plupart des physiciens le sentiment que Bohr et ses élèves (l'École de Copenhague) avaient totalement clarifié le sujet en répondant aux objections d'Einstein, c'est au niveau du cœur conceptuel de la théorie, et à partir des objections d'Einstein, qu'une nouvelle phase de progrès majeurs a débuté dans les années 1960 avec les travaux de John Bell, suivis d'expériences de plus en plus fines. La violation des inégalités de Bell a apporté l'évidence irréfutable de l'inséparabilité quantique, propriété extraordinaire d'une paire d'objets intriqués qui se comportent comme un système quantique unique, même si les deux objets sont très éloignés l'un de l'autre.

C'est à la même époque que l'on a appris à manipuler un par un des électrons, des ions, des atomes ou des photons, et on a pu observer directement les « sauts quantiques », évolutions brutales du système à des instants aléatoires. On avait longtemps cru cette notion réservée aux discussions conceptuelles sur la possibilité d'utiliser la mécanique quantique pour décrire l'évolution des objets individuels. Ici encore Einstein avait soulevé des objections profondes auxquelles Bohr s'était fait un devoir d'essayer de répondre de façon convaincante. On sait aujourd'hui que la mécanique quantique s'applique avec succès aux objets uniques, et pas seulement de façon statistique à de grands ensembles d'objets identiques comme les atomes d'une gaz. Et on a vu apparaître de nouvelles méthodes de calcul, dites « Monte-Carlo quantique », bien adaptées à la description du comportement des objets quantiques individuels.

Il n'est probablement pas exagéré de dire que ces deux avancées conceptuelles de la fin du XX<sup>e</sup> siècle – la compréhension de l'importance des états intriqués, la compréhension de la dynamique des objets quantiques individuels – signent le début d'une nouvelle révolution quantique. Et il n'est pas interdit d'imaginer qu'au-delà des progrès en physique fondamentale dont nous n'avons sans doute observé que les prémices, et des questions épistémologiques d'interprétation qui restent plus ouvertes que jamais, cette nouvelle révolution quantique pourrait à son tour bouleverser notre société en débouchant sur une nouvelle révolution technologique, la révolution de l'information quantique.

N'est-il pas remarquable que les deux ingrédients de cette nouvelle révolution quantique soient précisément les points qu'Einstein ne cessa de mettre en avant – certes pour les contester, mais surtout parce que, plus que tout autre, il en avait compris le caractère stupéfiant ?

## **2.2 L'émergence de l'intrication quantique : de l'article EPR à la violation des inégalités de Bell**

### **2.2.1 Le débat Bohr-Einstein et l'article EPR**

La mécanique quantique n'a pu se construire qu'au prix de révisions radicales et douloureuses des concepts qui s'étaient imposés en physique classique.



Ainsi, pour prendre en compte la dualité onde-particule, la mécanique quantique a dû renoncer au concept classique de trajectoire. On traduit ce renoncement par les célèbres relations de dispersion de Heisenberg (appelées aussi « relations d'incertitude ») qui décrivent quantitativement l'impossibilité de définir simultanément la position et la vitesse d'une particule avec une précision trop grande. Ce renoncement était si radical que plusieurs physiciens, au premier rang desquels Einstein et de Broglie, n'admettaient pas son caractère inéluctable, à la différence de Bohr qui en avait fait la clef de voûte de l'interprétation qu'il donnait de la nouvelle théorie, avec ses élèves de « l'École de Copenhague ». Aux congrès Solvay de 1927 et 1930, Einstein lança contre cette interprétation une série d'attaques mémorables, basées sur des « expériences de pensée » c'est-à-dire des schémas compatibles avec les lois fondamentales de la physique même s'ils semblaient totalement irréalisables en pratique au vu des moyens expérimentaux de l'époque. Ainsi, raisonnant sur des fentes d'Young mobiles dont on pourrait mesurer le recul sous l'effet de la déflexion d'un seul photon, ou encore sur une boîte contenant des photons dont le nombre serait mesurable à une unité près grâce à une pesée (en utilisant l'équivalence masse énergie), Einstein tenta de démontrer que les relations de Heisenberg ne s'appliquaient pas dans ces situations. On sait qu'en fait Bohr pût répondre de façon convaincante à ces attaques, et il semble qu'à partir de 1930 Einstein ait été convaincu de la validité et de la cohérence interne du formalisme quantique [1].

En fait, s'il ne remettait plus en cause le formalisme mathématique de la mécanique quantique, ni ses prévisions, Einstein restait profondément insatisfait par la théorie quantique telle qu'elle était présentée par Bohr et ses disciples. Il pensait, semble-t-il, que les renoncements mis en avant dans l'interprétation ne traduisaient que l'état d'inachèvement de la théorie. Cette position allait donner lieu à un nouveau débat homérique avec Niels Bohr, qui devait se poursuivre tout le reste de la vie des deux protagonistes. Ce débat débute avec la publication, en 1935, de l'article d'Einstein Podolsky et Rosen (EPR) [2], dont le titre pose la question : « La description quantique de la réalité physique peut-elle être considérée comme complète ? » Dans cet article, Einstein et ses coauteurs montrent que le formalisme quantique prédit l'existence d'états particuliers de deux particules, par exemple deux électrons, caractérisés par de très fortes corrélations à la fois des vitesses et des positions (Fig. 2.1).

Plus précisément, pour l'état EPR découvert par Einstein et ses collègues, le formalisme quantique prédit que des mesures de position sur chacun des deux électrons donneront des valeurs exactement symétriques par rapport à l'origine, et que des mesures de vitesses donneront des résultats toujours identiques. Pour une paire d'électrons dans un tel état, il suffit donc de mesurer la vitesse du premier électron pour connaître avec certitude celle de l'autre électron. Les deux électrons étant éloignés l'un de l'autre, la mesure effectuée sur le premier ne saurait modifier l'état du second, et Einstein et ses coauteurs en déduisent que le second électron possédait, avant la mesure, une vitesse

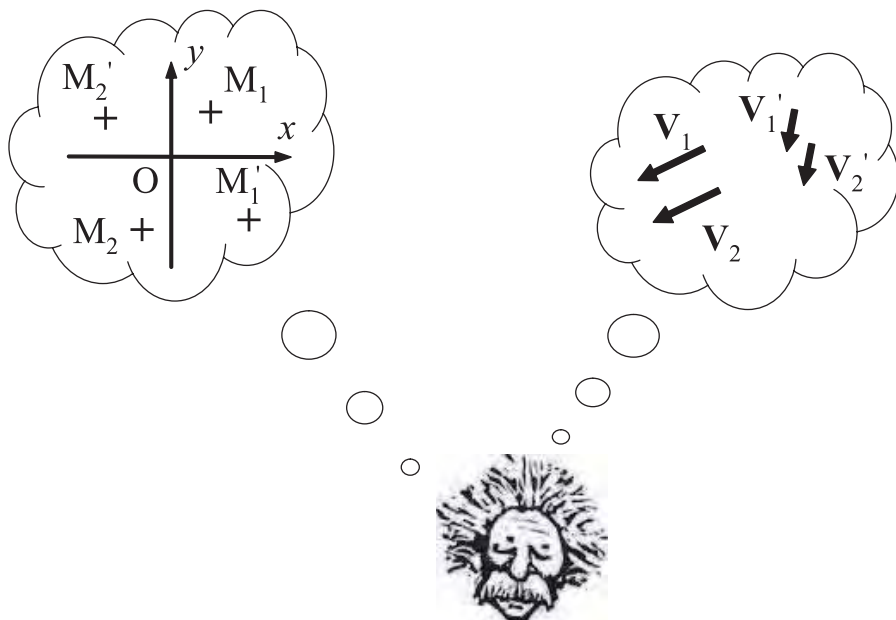


FIG. 2.1 – L'expérience de pensée d'Einstein, Podolsky, et Rosen (1935). Dans l'état quantique EPR, les deux particules  $\nu_1$  et  $\nu_2$  sont totalement corrélées, en position et en vitesse. La mesure de position de la première particule peut donner un résultat quelconque ( $M_1, M_1'$ ) ; mais si on a trouvé  $M_1$  on trouvera avec certitude la deuxième particule en  $M_2$ , symétrique de  $M_1$  par rapport à l'origine, et de même si on a trouvé  $M_1'$  pour la première on trouvera  $M_2'$  symétrique pour la deuxième. De la même façon, les mesures sur les vitesses sont totalement corrélées (on trouve toujours des vitesses identiques :  $\mathbf{V}_2 = \mathbf{V}_1$  ou  $\mathbf{V}_2' = \mathbf{V}_1'$ ). Comme une mesure sur la particule  $\nu_1$  ne saurait affecter instantanément la particule  $\nu_2$  éloignée, EPR en déduisent que la particule  $\nu_2$  possédait, avant les mesures, une valeur parfaitement déterminée de position et de vitesse, alors que la description quantique de l'état EPR ne spécifie aucune valeur précise pour ces grandeurs : EPR en concluent que cette description est incomplète.

parfaitement déterminée. Mais l'état quantique ne spécifie aucune valeur particulière de vitesse pour cet électron, et pour Einstein, Podolsky et Rosen il ne reflète donc pas la totalité des propriétés – de la « réalité physique » – de cet électron. Les auteurs en concluent que le formalisme quantique est incomplet. L'argument d'incomplétude va encore plus loin lorsqu'on remarque qu'au lieu de mesurer sa vitesse, on aurait pu choisir de mesurer la position du premier électron : on en aurait déduit la position du second, qui apparaît, elle aussi, comme parfaitement déterminée et préexistante à toute mesure,

contrairement à la description quantique. Suivant le raisonnement EPR ce sont en définitive aussi bien la vitesse que la position du second électron qui avaient des valeurs parfaitement déterminées avant toute mesure, en contradiction avec la description quantique de cet état qui ne spécifie aucune valeur particulière a priori pour ces grandeurs. De plus, l'affirmation que la vitesse et la position d'une même particule pourraient être a priori simultanément déterminées, entre en contradiction avec les relations de Heisenberg considérées comme centrales par Bohr et ses disciples. Einstein et ses collègues répondent donc négativement à la question posée dans le titre de leur article : le formalisme quantique n'est pas complet, il ne rend pas compte de la totalité de la réalité physique, et il faut donc s'attacher à essayer de le compléter.

Niels Bohr fut, semble-t-il, bouleversé par l'argument EPR qui s'appuie sur le formalisme quantique lui-même pour en démontrer le caractère incomplet, provisoire. Ses écrits révèlent la conviction profonde que si le raisonnement EPR était correct, compléter le formalisme quantique ne serait pas suffisant, c'est toute la physique quantique qui s'effondrerait. Bohr contesta donc immédiatement le raisonnement EPR, en affirmant que dans un état quantique de ce type, « non factorisable », on ne peut parler des propriétés individuelles de chaque électron, et cela même s'ils sont très éloignés l'un de l'autre [3]. Avec Schrödinger, qui découvrait au même moment ces états étonnants [4], on allait désormais parler d'état « intriqué »<sup>1</sup>, pour indiquer que les deux électrons sont indissolublement enchevêtrés, qu'ils forment un objet unique quelle que soit leur distance de séparation.

On pourrait penser que ce débat entre deux géants de la physique du XX<sup>e</sup> siècle eut un immense écho chez les physiciens. En fait, lors de la publication en 1935 de l'article EPR, la mécanique quantique allait de succès en succès et la plupart des physiciens ignorèrent ce débat qui leur paraissait académique : il semblait que l'adhésion à l'une ou l'autre des positions fût une affaire de goût personnel (ou de position épistémologique) sans aucune conséquence pratique sur la mise en œuvre du formalisme quantique, ce qu'Einstein lui-même ne semblait pas contester. Il fallut attendre trente ans pour voir un démenti à cette position relativement consensuelle, avec la parution, en 1964, d'un court article de John Bell qui allait modifier radicalement la situation. Avant de présenter le théorème de Bell proprement dit, nous allons d'abord reformuler le raisonnement EPR en nous intéressant non plus à des observables continues comme la position ou la vitesse, pouvant prendre n'importe quelles valeurs, mais à des observables discrètes ne pouvant prendre que deux valeurs, comme la polarisation d'un photon<sup>2</sup>.

---

<sup>1</sup>On dit aussi « enchevêtré », « entangled » en anglais, « Verschränkung » étant le nom original allemand donné par Schrödinger.

<sup>2</sup>C'est à D. Bohm que l'on doit la première version de l'expérience de pensée EPR portant sur des observables discrètes bivaluées, en l'occurrence des mesures suivant des axes différents de la composante d'un spin 1/2 par un appareil de Stern et Gerlach. Les mesures de polarisation de photon sont équivalentes, et elles ont donné lieu aux premières expériences réelles convaincantes.

## 2.2.2 Paires de photons EPR corrélés en polarisation

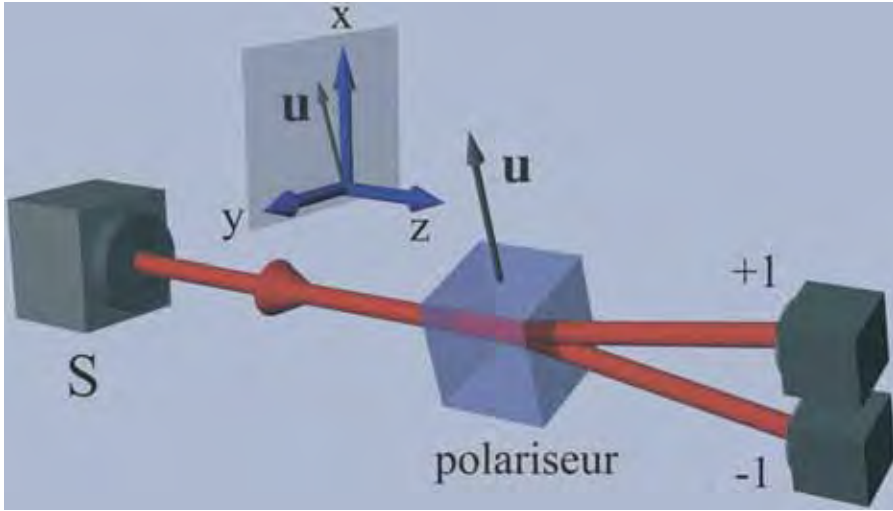
### Mesure de polarisation

On sait que la lumière, qui est un champ électromagnétique vibrant perpendiculairement à sa direction de propagation, peut être polarisée, c'est-à-dire qu'elle vibre suivant une direction particulière. Cela s'observe à l'aide d'un polariseur, que l'on place sur le faisceau lumineux, et qui laisse sortir la lumière soit dans la voie dite « ordinaire » (notée  $+1$  dans la figure 2.2), soit dans la voie « extraordinaire » (notée  $-1$ ). Dans le premier cas, la lumière est polarisée suivant la direction d'analyse du polariseur, repérée par la direction notée  $\mathbf{u}$  dans la figure 2.2. Au contraire, de la lumière polarisée perpendiculairement à  $\mathbf{u}$  sort suivant la voie extraordinaire.

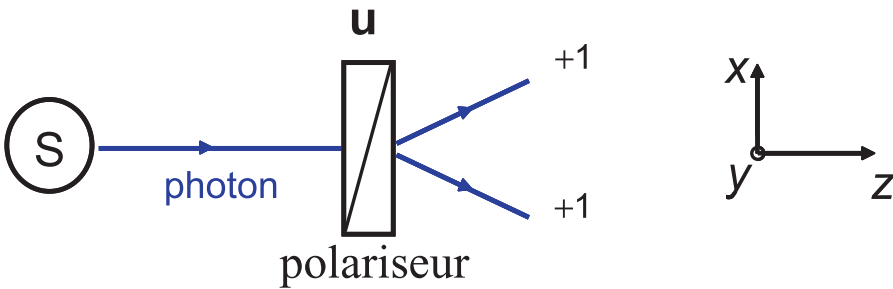
Dans le cas général, le faisceau lumineux n'est pas exactement polarisé parallèlement ou perpendiculairement à  $\mathbf{u}$ , et une fraction de la lumière sort dans la voie  $+1$  (composante polarisée suivant  $\mathbf{u}$ ) tandis que le reste sort dans la voie  $-1$  (composante polarisée perpendiculairement à  $\mathbf{u}$ ). Mais que se passe-t-il pour un photon unique, qui ne saurait être fractionné? Le photon sort soit dans la voie ordinaire soit dans la voie extraordinaire, et la valeur obtenue ( $+1$  ou  $-1$ ) constitue le résultat de la mesure de polarisation suivant la direction  $\mathbf{u}$ . La polarisation suivant  $\mathbf{u}$  est une observable discrète à deux valeurs possibles. Si on répète l'expérience en préparant à chaque fois le photon dans le même état, on a une certaine probabilité d'obtenir le résultat  $+1$ , et la probabilité complémentaire d'obtenir  $-1$ . Si le photon a été préparé dans un état de polarisation bien déterminé, par exemple l'état noté  $|x\rangle$  correspondant à une polarisation suivant  $\mathbf{Ox}$ , et si on effectue une mesure suivant le même axe  $\mathbf{Ox}$ , on trouve le résultat  $+1$  dans 100 % des cas. Mais si on choisit une orientation  $\mathbf{u}$  du polariseur différente de  $\mathbf{Ox}$ , par exemple la direction  $\mathbf{Ox}'$  à 45 degrés de  $\mathbf{Ox}$ , le photon dans l'état  $|x\rangle$  donnera le résultat  $+1$  dans 50 % des cas, et  $-1$  dans 50 % des cas : la valeur de la polarisation n'est pas définie suivant l'axe  $\mathbf{Ox}'$ . En fait, les deux mesures suivant  $\mathbf{Ox}$  et  $\mathbf{Ox}'$  correspondent à deux observables quantiques incompatibles, de même que position et vitesse d'une particule sont deux observables incompatibles.

Considérons maintenant (Fig. 2.3) une paire de photons  $\nu_1$  et  $\nu_2$  émis simultanément vers la gauche et vers la droite, et effectuons sur ces photons, à l'aide des polariseurs I et II orientés suivant  $\mathbf{a}$  et  $\mathbf{b}$ , des mesures conjointes de polarisation. Si la paire de photons est préparée dans l'état de polarisation  $|x, x\rangle$ , chaque photon est polarisé suivant  $\mathbf{Ox}$ , et une mesure conjointe de polarisation suivant  $\mathbf{Ox}$  sur chaque photon donnera toujours le résultat  $(+1, +1)$  : les résultats de mesure suivant  $\mathbf{Ox}$  sont totalement corrélés, puisqu'ils sont identiques. De façon analogue des paires de photon préparées dans l'état  $|y, y\rangle$  et mesurées suivant  $\mathbf{Ox}$  donneront à chaque fois le résultat  $(-1, -1)$ , ce qui correspond encore à une corrélation totale entre les résultats.

Si maintenant on considère une situation où les paires successives sont aléatoirement  $|x, x\rangle$  ou  $|y, y\rangle$ , en proportions égales (50 %), des mesures suivant  $\mathbf{Ox}$  sur chaque photon sont totalement corrélées puisqu'on obtient  $(+1, +1)$



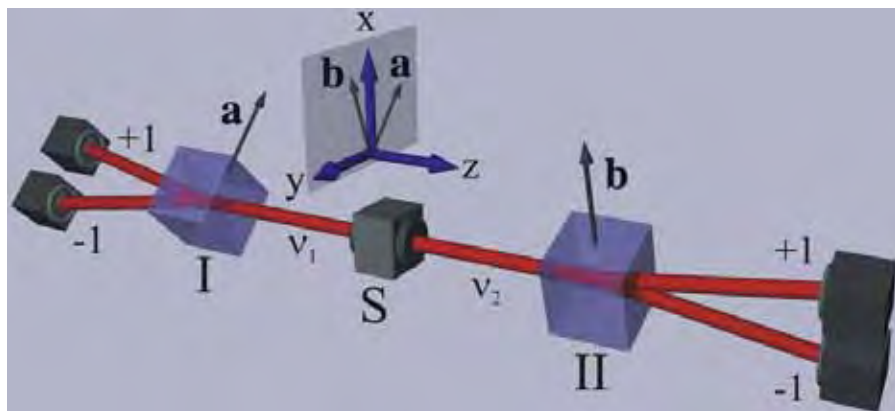
(a)



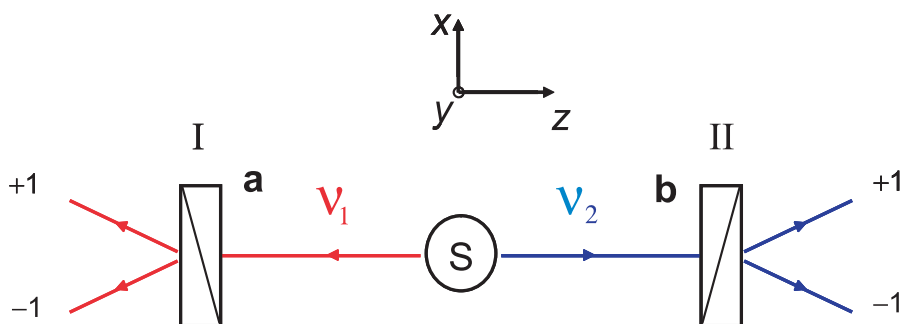
(b)

FIG. 2.2 – *Mesure de polarisation de la lumière. Le polariseur, qui peut tourner autour de l'axe  $Oz$  de propagation de la lumière, effectue une mesure de polarisation suivant  $u$ , perpendiculaire à  $Oz$ . De la lumière polarisée suivant  $u$  sort dans la voie +1, tandis que de la lumière polarisée perpendiculairement à  $u$  sort dans la voie -1. Dans le cas général, une fraction sort dans chaque voie. Un photon unique, qui ne peut se partager, sort soit dans la voie +1 soit dans la voie -1, où il est détecté avec des probabilités qui dépendent de l'état quantique. On dit qu'on effectue sur lui une mesure de polarisation suivant  $u$ , dont les résultats peuvent être +1 ou -1. (a) Vue d'artiste, incluant les détecteurs placés sur les voies +1 et -1. (b) Représentation schématique.*

ou  $(-1, -1)$  mais jamais  $(+1, -1)$  ni  $(-1, +1)$ . Notons en revanche que si l'on effectue des mesures conjointes suivant  $Ox'$  sur chaque photon, les mesures conjointes donneront avec une égale probabilité (25 %) chacun des résultats possibles :  $(+1, +1)$ ,  $(+1, -1)$ ,  $(-1, +1)$ , ou  $(-1, -1)$ . Dans ce cas, on n'observe aucune corrélation.



(a)



(b)

FIG. 2.3 – *Expérience de pensée EPR avec des paires de photons corrélés en polarisation. Les photons  $\nu_1$  et  $\nu_2$  d'une même paire sont analysés en polarisation suivant les directions  $\mathbf{a}$  et  $\mathbf{b}$  par les polariseurs I et II (les vecteurs  $\mathbf{a}$  et  $\mathbf{b}$  qui caractérisent les orientations des polariseurs sont perpendiculaires à  $\mathbf{Oz}$ ). Les résultats des mesures permettent de mettre en évidence les corrélations de polarisation. (a) Vue d'artiste, incluant les détecteurs placés sur les voies +1 et -1. (b) Représentation schématique.*

### État EPR pour les photons

Considérons maintenant des paires de photons dans l'« état EPR » :

$$|\Psi_{\text{EPR}}(\nu_1, \nu_2)\rangle = \frac{1}{\sqrt{2}}(|x, x\rangle + |y, y\rangle) \quad (2.1)$$

Notons d'emblée le caractère particulier de cet état, pour lequel on ne peut attribuer de polarisation précise ni au photon  $\nu_1$  ni au photon  $\nu_2$ . Il est en effet possible de montrer qu'il n'existe aucune transformation mathématique permettant d'écrire cet état sous forme d'un produit de deux termes

respectivement associés à  $\nu_1$  et  $\nu_2$  : cet état est non factorisable, non décomposable en deux termes caractérisant séparément les photons  $\nu_1$  et  $\nu_2$ . Cette propriété mathématique est au cœur de l'intrication.

Comme pour les mesures de vitesse ou de position dans la situation EPR originale, le formalisme de la mécanique quantique prédit que pour des photons dans l'état EPR (Éq. (2.1)) les mesures de polarisation suivant une même orientation sont totalement corrélées, quelle que soit cette orientation. Plus précisément, si on s'intéresse aux résultats des mesures de polarisation conjointes lorsque les deux polariseurs sont orientés suivant la même direction ( $\mathbf{a} = \mathbf{b}$ ), les résultats sont parfaitement corrélés, alors que pris séparément ils apparaissent aléatoires : si on obtient  $+1$  pour  $\nu_1$  (ce qui arrive dans 50 % des cas) alors on obtient avec certitude  $+1$  pour  $\nu_2$ , mais si on obtient  $-1$  pour  $\nu_1$  alors on obtient avec certitude  $-1$  pour  $\nu_2$ . Une confirmation de cette corrélation est la prédiction qu'on ne trouve jamais les résultats opposés  $(+1, -1)$  ou  $(-1, +1)$  lorsque les deux polariseurs sont orientés suivant la même direction d'analyse.

On notera que la situation décrite par l'état EPR (2.1) est très différente de celle du mélange d'états  $|x, x\rangle$  et  $|y, y\rangle$  en proportions égales, vue plus haut (paragraphe 2.2.2). Pour le mélange, on a certes une corrélation totale lorsque les deux polariseurs effectuent les mesures suivant  $\mathbf{Ox}$ , mais il n'y a aucune corrélation pour des mesures suivant  $\mathbf{Ox}'$  à 45 degrés de  $\mathbf{Ox}$ . Au contraire pour l'état EPR la corrélation de polarisation est totale quelle soit la direction (identique) choisie pour les deux polariseurs.

### Coefficient de corrélation de polarisation

Il est possible de donner une mesure de la corrélation entre les résultats de mesure de polarisation suivant des orientations différentes  $\mathbf{a}$  et  $\mathbf{b}$  des polariseurs I et II : on introduit pour cela le coefficient de corrélation  $E(\mathbf{a}, \mathbf{b})$

$$E(\mathbf{a}, \mathbf{b}) = P_{++}(\mathbf{a}, \mathbf{b}) + P_{--}(\mathbf{a}, \mathbf{b}) - P_{-+}(\mathbf{a}, \mathbf{b}) - P_{+-}(\mathbf{a}, \mathbf{b}) \quad (2.2)$$

où  $P_{++}(\mathbf{a}, \mathbf{b})$  désigne la probabilité conjointe d'obtenir  $+1$  pour  $\nu_1$  et  $+1$  pour  $\nu_2$ , les polariseurs I et II étant respectivement orientés suivant  $\mathbf{a}$  et  $\mathbf{b}$ . On note qu'une corrélation totale ( $P_{++} = P_{--} = 1/2$  et  $P_{+-} = P_{-+} = 0$ ) est signalée par un coefficient de corrélation  $E = 1$ , tandis qu'une corrélation nulle ( $P_{++} = P_{--} = P_{+-} = P_{-+} = 1/4$ ) est signalée par une valeur nulle de  $E$ .

Pour une paire de photons dans l'état EPR (2.1), le calcul quantique prédit un coefficient de corrélation de polarisation

$$E_{MQ}(\mathbf{a}, \mathbf{b}) = \cos 2(\mathbf{a}, \mathbf{b}) \quad (2.3)$$

où  $(\mathbf{a}, \mathbf{b})$  est l'angle entre les orientations des polariseurs. La corrélation est donc totale lorsque les directions d'analyse sont identiques ( $E(\mathbf{a}, \mathbf{a}) = 1$ ).

## Raisonnement EPR et variables supplémentaires

Il est alors facile de reprendre mutatis mutandis le raisonnement EPR, les mesures de polarisation suivant un axe (par exemple  $\mathbf{Ox}$ ) correspondant aux mesures de position, tandis que les mesures de polarisation suivant un axe distinct (par exemple  $\mathbf{Ox}'$  à 45 degrés de  $\mathbf{Ox}$ ) correspondent aux mesures de vitesse. La ligne de raisonnement est la suivante. Intéressons nous à une mesure conjointe de polarisation suivant une même direction – par exemple  $\mathbf{Ox}$ , et considérons un cas où on obtient  $(+1, +1)$  : en fait, selon le raisonnement EPR, nul besoin d'effectuer la mesure sur  $\nu_2$  selon  $\mathbf{Ox}$  pour savoir qu'elle donne  $+1$  ; dans l'état EPR, il suffit d'observer  $+1$  sur  $\nu_1$  selon  $\mathbf{Ox}$  pour être certain qu'une mesure sur  $\nu_2$  selon la même direction donnerait  $+1$ . Or  $\nu_2$  est éloigné de  $\nu_1$  et la mesure sur  $\nu_1$  ne saurait l'avoir affecté. On en conclut que la valeur de la polarisation de  $\nu_2$  suivant  $\mathbf{Ox}$  était parfaitement déterminée avant toute mesure, elle valait  $+1$  dans cet exemple. Mais le formalisme quantique – en l'occurrence l'expression (2.1) du vecteur d'état décrivant la paire, n'attribue pas de valeur particulière à la polarisation du photon  $\nu_2$ , d'où l'on conclut que ce formalisme est incomplet<sup>3</sup>.

Ainsi, c'est en s'appuyant sur le formalisme quantique lui même que l'on peut conclure, en suivant Einstein, que ce formalisme doit être complété. Il ne s'agit pas de contester la validité de ce formalisme, mais bien plutôt d'introduire un niveau supplémentaire de description, plus détaillé, dans lequel chaque photon particulier aurait une polarisation bien définie. L'espoir est alors de retrouver au niveau statistique les prévisions de la mécanique quantique habituelle. La démarche serait analogue à celle qui prévaut en théorie cinétique des gaz, où l'on peut en principe décrire chaque molécule d'un gaz avec sa trajectoire, bien qu'il suffise en pratique d'utiliser une description statistique (la distribution de Maxwell-Boltzmann) pour rendre compte des propriétés du gaz qui intéressent le physicien. C'est cette possibilité de compléter le formalisme quantique standard, a priori sans le remettre en cause, que John Bell allait examiner dans toutes ses conséquences.

### 2.2.3 Le théorème de Bell

En 1964 paraît un court article de John Bell, physicien théoricien irlandais travaillant au CERN à Genève. En quelques lignes de calcul, cet article [6] montre que si l'on prend au sérieux l'argument EPR, et qu'on introduit explicitement des variables supplémentaires spécifiant la polarisation de chaque

---

<sup>3</sup>Pour aller jusqu'au bout des arguments de l'article EPR, on peut encore noter qu'il suffit d'envisager une mesure conjointe suivant  $\mathbf{Ox}'$ , à 45 degrés de  $\mathbf{Ox}$ , pour conclure que le photon  $\nu_2$  possédait avant la mesure une polarisation déterminée à la fois suivant  $\mathbf{Ox}$  et suivant  $\mathbf{Ox}'$ . Or dans le formalisme quantique il s'agit d'observables complémentaires incompatibles, analogues à la vitesse et la position d'une particule, et il existe des relations de Heisenberg interdisant à ces deux quantités d'être simultanément déterminées avec précision.



photon, alors on aboutit à une contradiction avec les prédictions quantiques. Plus précisément, même si un tel formalisme à paramètres supplémentaires (on parle aussi de « variables cachées ») permet de rendre compte des corrélations totales prévues par le formalisme quantique lorsque les deux polariseurs sont orientés suivant la même direction d'analyse, cette possibilité ne s'étend pas à la totalité des situations possibles, et aucun modèle à paramètres supplémentaires ne peut donner la valeur prédite par la mécanique quantique pour l'ensemble de toutes les orientations envisageables pour les deux polariseurs. *Il n'est donc pas possible de « comprendre » dans leur totalité les corrélations EPR en complétant le formalisme quantique dans l'esprit suggéré par Einstein.*

Aujourd'hui encore, ce résultat ne laisse pas de nous étonner. S'il s'appliquait au cas des jumeaux humains (vrais), pour lesquels on observe des corrélations fortes dans les caractères physiques (couleur des yeux, des cheveux...) ou dans le profil biologique et médical (groupe d'histocompatibilité, développement de telle ou telle pathologie comme l'hypertension artérielle, ...), on devrait renoncer à expliquer ces corrélations en invoquant des jeux de chromosomes identiques pour les deux jumeaux.

## Le formalisme de Bell : exemple

Pour établir ce résultat, Bell commence par traduire en termes mathématiques la conclusion du raisonnement EPR, et il complète le formalisme quantique en introduisant explicitement un paramètre supplémentaire, noté  $\lambda$ , porté par chacun des deux photons d'une même paire. Ce paramètre commun varie d'une paire à l'autre. Il suffit d'admettre que les résultats des mesures de polarisation sur les photons  $\nu_1$  et  $\nu_2$  dépendent du paramètre commun  $\lambda$  pour obtenir des résultats de mesure fortement corrélés. Par exemple, on peut imaginer que chaque photon possède dès le départ une polarisation dont l'orientation  $\mathbf{p}$  est spécifiée par l'azimut  $\lambda$  qui joue le rôle de paramètre supplémentaire ; on suppose évidemment que le résultat d'une mesure de polarisation suivant la direction  $\mathbf{u}$  (spécifiée par l'azimut  $\theta$ ) dépend de l'angle  $(\mathbf{p}, \mathbf{u}) = \theta - \lambda$  entre  $\mathbf{p}$  et  $\mathbf{u}$ . Il est facile de développer un modèle – c'est-à-dire de spécifier des formes mathématiques particulières pour décrire la distribution des paramètres  $\lambda$  parmi les paires émises, et la façon dont le résultat de mesure dépend de l'angle  $(\theta - \lambda)$  – prédisant une corrélation totale lorsque les polariseurs ont la même orientation [8]. La figure 2.4 montre la prédiction de ce modèle, comparée au résultat (2.3) du calcul quantique. On constate que la prédiction de notre modèle est un peu différente de la prédiction quantique, mais la tendance générale est bien reproduite, et on peut se demander s'il est possible de raffiner le modèle jusqu'à obtenir exactement de la prédiction quantique.

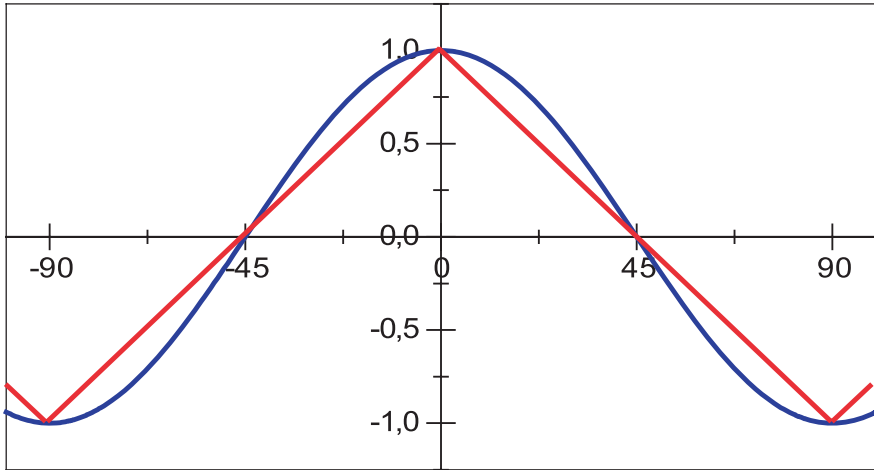


FIG. 2.4 – Coefficient de corrélation de polarisation en fonction de l'angle entre polariseurs. La figure permet de comparer les résultats du calcul quantique (sinusoïde) à ceux d'un modèle à paramètres supplémentaires bien choisi (ligne brisée). On constate que les prédictions des deux modèles sont identiques pour des polariseurs parallèles, perpendiculaires, ou à 45 degrés l'un de l'autre, et qu'elles restent voisines aux angles intermédiaires. Mais le théorème de Bell montre qu'il est impossible de développer un modèle à paramètres supplémentaires locaux dont les prédictions coïncideraient avec les prédictions quantiques pour toutes les orientations.

## Inégalités de Bell

La force du raisonnement de Bell est de répondre négativement à la question de façon générale, sans se référer à un modèle particulier. Il s'applique à tout modèle où l'on rend compte des corrélations en introduisant un paramètre  $\lambda$  identique pour les deux photons d'une même paire, et où le résultat d'une mesure de polarisation est décrit par une fonction ne pouvant prendre que la valeur  $+1$  ou  $-1$ , et ne dépendant que du paramètre supplémentaire  $\lambda$  du photon analysé et de l'orientation  $\theta$  du polariseur effectuant la mesure. Il suffit d'explicitier mathématiquement cette hypothèse pour démontrer les inégalités<sup>4</sup>

$$-2 \leq S(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}') \leq +2 \quad (2.4)$$

avec

$$S(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}') = E(\mathbf{a}, \mathbf{b}) - E(\mathbf{a}, \mathbf{b}') + E(\mathbf{a}', \mathbf{b}) + E(\mathbf{a}', \mathbf{b}') \quad (2.5)$$

La quantité  $S(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}')$  met en jeu quatre mesures de corrélation de polarisation, avec le polariseur I orienté suivant  $\mathbf{a}$  puis  $\mathbf{a}'$ , et le polariseur II orienté

<sup>4</sup>Cette forme des inégalités de Bell, bien adaptée aux tests expérimentaux, a été écrite pour la première fois dans [9].

suivant  $\mathbf{b}$  puis  $\mathbf{b}'$ . Les inégalités de Bell imposent à la combinaison (2.5) des coefficients de corrélation de polarisation  $E$  – définis en (2.2) – d'être compris entre  $+2$  et  $-2$ .

La démonstration de ce résultat crucial est si simple que nous la donnons ici. Appelons  $A(\lambda, \mathbf{a})$  la fonction qui, dans un modèle à paramètres supplémentaires<sup>5</sup>, donne le résultat de la mesure par le polariseur I orienté suivant  $\mathbf{a}$ , lorsque le photon  $\nu_1$  « porte » le paramètre supplémentaire  $\lambda$ . Cette fonction ne peut prendre que deux valeurs,  $+1$  ou  $-1$ . Une fonction analogue  $B(\lambda, \mathbf{b})$  caractérise le résultat de la mesure par le polariseur II orienté suivant  $\mathbf{b}$ , lorsque le photon  $\nu_2$  porte le paramètre supplémentaire  $\lambda$ . Considérons alors la quantité

$$s(\lambda, \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}') = A(\lambda, \mathbf{a}) \cdot B(\lambda, \mathbf{b}) - A(\lambda, \mathbf{a}) \cdot B(\lambda, \mathbf{b}') \\ + A(\lambda, \mathbf{a}') \cdot B(\lambda, \mathbf{b}) + A(\lambda, \mathbf{a}') \cdot B(\lambda, \mathbf{b}') \quad (2.6)$$

qui met en jeu deux orientations ( $\mathbf{a}$  et  $\mathbf{a}'$ ) pour le polariseur I, et deux orientations ( $\mathbf{b}$  et  $\mathbf{b}'$ ) pour le polariseur II. Comme chaque fonction  $A$  ou  $B$  ne peut valoir que  $+1$  ou  $-1$ , il est facile de vérifier (par exemple en mettant  $A(\lambda, \mathbf{a})$  et  $A(\lambda, \mathbf{a}')$  en facteur) que  $s(\lambda, \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}')$  ne peut valoir que  $+2$  ou  $-2$ . Lorsqu'on moyenne cette quantité sur toutes les valeurs possibles de  $\lambda$  (le jeu d'orientations  $\mathbf{a}$ ,  $\mathbf{a}'$ ,  $\mathbf{b}$ ,  $\mathbf{b}'$  étant fixé), on obtient un résultat nécessairement compris entre  $+2$  et  $-2$ . Or cette moyenne n'est autre que la quantité  $S(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}')$  définie en (2.5), car la moyenne de  $A(\lambda, \mathbf{a}) \cdot B(\lambda, \mathbf{b})$  est précisément le coefficient de corrélation de polarisation pour les orientations  $\mathbf{a}$  et  $\mathbf{b}$  des polariseurs.

### Conflit avec la mécanique quantique

Les inégalités de Bell (2.4) sont très générales : elles s'appliquent à tout modèle dans lequel on rend compte des corrélations de polarisation en introduisant des paramètres supplémentaires, comme Einstein l'avait suggéré. Mais il se trouve que les corrélations de polarisation (2.3) prévues par la mécanique quantique pour des photons dans l'état EPR (2.1) violent ces inégalités pour certaines valeurs des orientations. Cette violation peut être considérable puisque la quantité  $S$  calculée par la mécanique quantique atteint, pour les orientations de la figure 2.5, la valeur

$$S_{\text{MQ}}(\mathbf{a}_0, \mathbf{a}'_0, \mathbf{b}_0, \mathbf{b}'_0) = 2\sqrt{2} = 2,828... \quad (2.7)$$

comme le lecteur le vérifiera aisément à partir de (2.3).

Ce résultat a une portée conceptuelle que l'on ne saurait surestimer. La violation des inégalités de Bell par les prédictions quantiques démontre explicitement que les corrélations quantiques EPR sont d'une nature irréductible

---

<sup>5</sup>La notation  $\lambda$  peut désigner un ensemble de paramètres, et la démonstration n'est pas limitée au cas d'un paramètre unique.

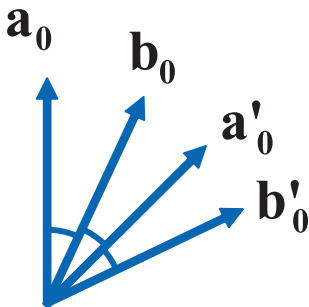


FIG. 2.5 – Violation maximale des inégalités de Bell. Pour le jeu d'orientations  $\{\mathbf{a}_0, \mathbf{a}'_0, \mathbf{b}_0, \mathbf{b}'_0\}$  des polariseurs I et II telles que  $(\mathbf{a}_0, \mathbf{b}_0) = (\mathbf{b}_0, \mathbf{a}'_0) = (\mathbf{a}'_0, \mathbf{b}'_0) = \pi/8$ , la quantité  $S_{\text{MQ}}$ , prédite par la mécanique quantique dans la situation EPR, prend la valeur  $2\sqrt{2}$ , bien supérieure à la limite supérieure des inégalités (2.4).

aux concepts classiques, qui conduisent aux inégalités. Le formalisme de Bell est en effet la forme générale traduisant le raisonnement effectué habituellement en présence de corrélations, par un scientifique aussi bien que par un juge d'instruction. C'est par exemple au terme d'un raisonnement de ce type que les biologistes, observant que les vrais jumeaux possèdent des caractéristiques identiques (couleur des cheveux, couleur des yeux, groupe sanguin, type d'histocompatibilité), ont conclu que ces caractéristiques sont déterminées par des chromosomes, identiques pour les deux jumeaux d'une même paire, et cela bien avant que l'on ait observé les chromosomes au microscope électronique. À la lumière du théorème de Bell, on est forcé de constater que ce type de raisonnement ne s'applique pas aux corrélations quantiques EPR.

L'autre conséquence, tout aussi importante, du théorème de Bell, est la possibilité qu'il ouvre de *trancher par l'expérience le débat entre Einstein et Bohr*. Il suffit en principe de mesurer les corrélations dans une situation où la mécanique quantique prédit une violation des inégalités de Bell pour savoir s'il faut renoncer à une interprétation « à la Einstein », où si au contraire on a identifié une situation où la mécanique quantique est prise en défaut (c'est le destin normal des théories physiques de rencontrer un jour une limite de validité).

## 2.2.4 Condition de localité et expérience avec polariseurs variables

Ayant identifié une contradiction entre les prédictions quantiques et tout modèle à paramètres supplémentaires tentant de compléter la mécanique quantique conformément au programme EPR, on doit s'interroger sur les raisons profondes de ce conflit. Quelles sont les hypothèses, explicites ou implicites, nécessaires pour obtenir les inégalités de Bell ? Avec quatre décennies

de recul, et après des centaines d'articles consacrés à la question, il semble possible de conclure que deux hypothèses suffisent :

- (1) l'existence d'une « réalité physique autonome » (suivant le vocabulaire d'Einstein) pour un système physique isolé, ce qui conduit à l'introduction de paramètres supplémentaires ;
- (2) l'hypothèse de localité, elle aussi clairement invoquée par Einstein, et soulignée par Bell dès son premier article.

Dans le contexte de l'expérience de pensée de la figure 2.3, l'hypothèse de localité stipule que le résultat de la mesure par un polariseur (par exemple le polariseur I) ne peut pas dépendre de l'orientation ( $\mathbf{b}$ ) de l'autre polariseur (II) qui est éloigné – et réciproquement la mesure en II ne peut être influencée par l'orientation  $\mathbf{a}$  de I. De même, l'état des photons lors de leur émission ne saurait dépendre des orientations  $\mathbf{a}$  et  $\mathbf{b}$  des polariseurs qui effectueront ultérieurement les mesures sur ces mêmes photons.

Nous avons implicitement admis cette hypothèse de localité en écrivant le formalisme permettant de démontrer les inégalités de Bell (voir l'expression (2.6)). Ainsi, la fonction  $A(\lambda, \mathbf{a})$  décrivant la mesure par le polariseur I, ne dépend manifestement pas de l'orientation  $\mathbf{b}$  du polariseur II éloigné. De même, lorsqu'on effectue la moyenne de  $s(\lambda, \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}')$  sur l'ensemble des paires, on admet implicitement que la distribution de probabilité des paramètres  $\lambda$  sur les paires ne dépend pas des orientations  $\mathbf{a}$  ou  $\mathbf{a}'$ ,  $\mathbf{b}$  ou  $\mathbf{b}'$ , des polariseurs qui effectueront la mesure. Ces conditions de localité sont nécessaires pour obtenir les inégalités de Bell. Il est en effet facile de vérifier, en reprenant la démonstration, que si par exemple on autorisait la réponse du polariseur I à dépendre de  $\mathbf{b}$ , en l'écrivant  $A(\lambda, \mathbf{a}, \mathbf{b})$ , alors on ne pourrait pas démontrer que la quantité  $s(\lambda, \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}')$  vaut  $\pm 2$ .

Pour naturelle qu'elle soit, la condition de localité ne découle d'aucune loi fondamentale : comme Bell le faisait remarquer [6], rien n'interdit qu'une interaction inconnue permette à l'orientation du polariseur II d'influencer le polariseur I. Mais Bell ajoutait aussitôt que si on pouvait réaliser *une expérience dans laquelle les orientations des polariseurs sont modifiées rapidement* pendant la propagation des photons entre la source et les polariseurs, alors c'est la *causalité relativiste* d'Einstein qui interdirait l'existence d'une telle interaction (voir Fig. 2.6). En effet, aucune interaction ne pouvant se propager plus vite que la lumière, la valeur la plus récente de l'orientation du polariseur II n'arrivera pas à temps en I pour influencer la mesure. De façon encore plus évidente, dans une telle expérience de pensée, l'état initial des photons au moment de leur préparation dans la source ne peut pas dépendre des orientations que les polariseurs n'ont pas encore mais qu'ils auront au moment de la mesure. Avec un tel schéma, *la condition de localité* (condition (2) ci-dessus) *n'est plus une hypothèse, elle devient une conséquence de la causalité relativiste d'Einstein.*

Dans l'expérience de pensée de la figure 2.6, on teste l'ensemble des idées qu'Einstein ne cessa de mettre en avant dans le contexte des corrélations EPR :

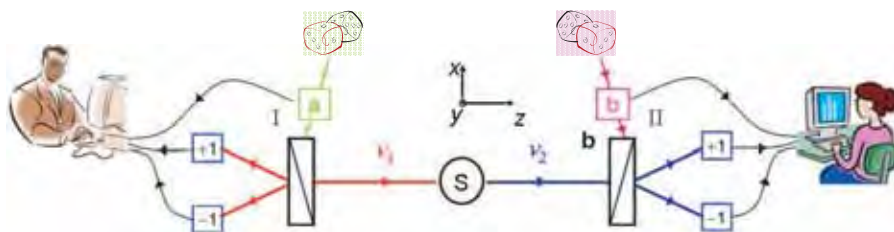


FIG. 2.6 – *Expérience de pensée avec polariseurs variables. S’il est possible de modifier aléatoirement les orientations  $\mathbf{a}$  et  $\mathbf{b}$  des polariseurs I et II pendant le temps de propagation des photons entre la source et les polariseurs, alors la condition de localité de Bell devient une conséquence de la causalité relativiste d’Einstein. Pour souligner la séparation entre les opérations de mesure, on a indiqué qu’il est possible d’enregistrer séparément à chaque poste de mesure les données suivantes : le résultat de la mesure de polarisation, l’orientation du polariseur au moment où il a donné ce résultat, et la date de ce résultat. La corrélation sera déterminée a posteriori par comparaison entre les résultats obtenus au même instant aux deux postes de mesure.*

d’une part la possibilité (ou même pour Einstein la nécessité) de compléter le formalisme quantique ; d’autre part l’impossibilité d’une interaction directe entre événements séparés par un « intervalle relativiste du genre espace », c’est-à-dire tels qu’aucun signal ne peut se propager de l’un à l’autre à une vitesse inférieure ou égale à celle de la lumière. C’est donc bien le conflit entre une certaine vision du monde défendue par Einstein et la mécanique quantique que l’on va mettre à l’épreuve dans les expériences de tests des inégalités de Bell.

## 2.2.5 Le recours à l’expérience

De façon surprenante, malgré des décennies de résultats expérimentaux en accord avec la théorie quantique, on ne disposait en 1964 d’aucun résultat permettant un test des inégalités de Bell. On s’aperçut que les situations dans lesquelles la théorie prévoyait une violation des inégalités de Bell (situations « sensibles ») étaient extrêmement rares, et il fallut donc imaginer des expériences sensibles réalisables avec les moyens de l’époque. En 1969, l’article de J. Clauser, M. Horne, A. Shimony et R. Holt [9] montra que la paire de photons visibles produits dans la désexcitation d’un atome sautant de niveau en niveau (cascade radiative) pouvait constituer un bon candidat, à condition de bien choisir les niveaux (voir par exemple l’insert de la figure 2.8). Sous l’impulsion de cet article, des expériences pionnières virent le jour dès le début des années 1970, montrant la possibilité de créer des paires de photons intriqués, même si la très faible efficacité des sources lumineuses rendait ces expériences extrêmement difficiles. On ne s’étonnera donc pas que les premiers tests des inégalités de Bell aient donné des résultats quelque peu contradictoires [8],

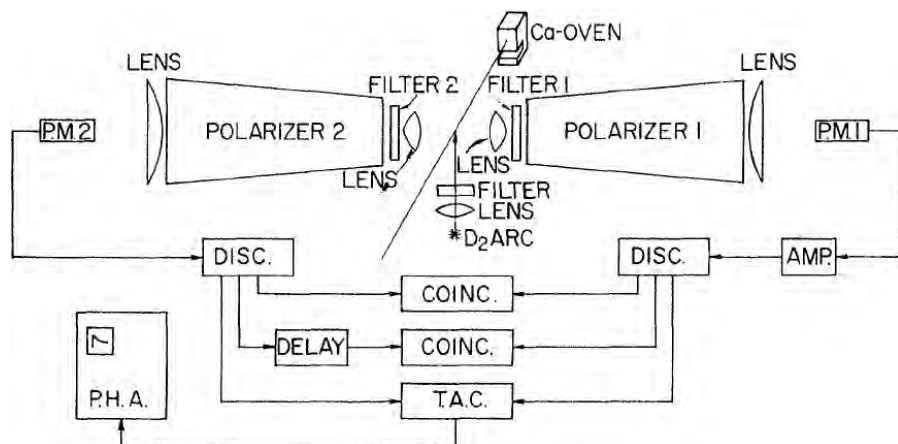


FIG. 2.7 – Expériences avec polariseurs à une seule voie. L'expérience ci-dessus (Clauser et Freedman, 1972), premier test des inégalités de Bell réalisé avec des photons visibles, utilisait des polariseurs à « piles de glaces » dont un seul résultat de mesure (+1) est accessible, le résultat complémentaire (−1) donnant lieu à une absorption du photon, et donc à aucun signal détectable. En utilisant des calibrations auxiliaires pendant lesquelles un ou les deux polariseurs sont retirés du faisceau, il est possible, au prix d'hypothèses supplémentaires, d'estimer les résultats −1, et donc de réaliser un test des inégalités de Bell. L'expérience de Clauser et Freedman a conclu à un accord avec les prédictions quantiques, et à une violation des inégalités de Bell par 5 écarts types. À la même époque une expérience similaire (Holt et Pipkin) aboutissait à un résultat opposé (accord avec les inégalités de Bell, contradiction avec les prédictions quantiques). Quatre ans plus tard, une expérience améliorée grâce à l'utilisation d'un laser (Fry et Thompson) tranchait clairement en faveur de la mécanique quantique (voir les références de ces expériences dans [8]).

avec pourtant un avantage à la mécanique quantique, mais cela dans des schémas expérimentaux encore éloignés de l'expérience de pensée idéale. En effet, dans toutes ces expériences de première génération, les polariseurs étaient de type « à une seule voie » (voir Fig. 2.7), c'est-à-dire que seul le résultat +1 (voir Fig. 2.3) était accessible à la mesure, le résultat −1 donnant lieu à une absorption du photon, c'est-à-dire à aucun signal explicite. Comme l'a montré l'article [9], il est possible dans une certaine mesure de « reconstituer » les résultats −1 manquant, à condition de recourir à des calibrations auxiliaires où les polariseurs sont retirés du faisceau : on comprend que l'on puisse identifier la différence entre les signaux sans et avec polariseurs au résultat −1, et aboutir ainsi à un test significatif des inégalités de Bell. Néanmoins, avec cette procédure, deux problèmes se posent, l'un conceptuel, l'autre expérimental. D'une part la stratégie esquissée ci-dessus ne permet de passer au test des inégalités de Bell qu'au prix d'une hypothèse supplémentaire excluant

une « conspiration des polariseurs et des détecteurs » ; même si elle est raisonnable, cette nouvelle hypothèse rend le test un peu moins direct. D'autre part, sur le plan expérimental, la validité de la comparaison entre des mesures avec et sans polariseurs exige une stabilité extrême de l'expérience et en particulier de la source de paires de photons intriqués, difficile à assurer pour des durées de mesure de plusieurs dizaines d'heures.

On mentionnera ici pour mémoire des expériences réalisées non pas avec des photons optiques, mais avec des photons gamma, ou encore des protons. Ces expériences étaient encore plus indirectes, car à la différence de l'optique il n'existe pas pour les photons gamma de polariseur susceptible de donner directement les résultats  $+1$  ou  $-1$ . La méthode consiste alors à utiliser un phénomène, la diffusion Compton, sensible à la polarisation des photons gamma mais donnant un résultat continu et non bivalué. Pour passer de ce résultat continu à ce que donnerait un hypothétique polariseur réalisant la mesure quantique, on doit recourir au formalisme quantique. On comprend alors que les tests des inégalités de Bell basés sur ce type d'expériences mettent en jeu des hypothèses supplémentaires nettement plus fortes que celles relatives aux expériences avec photons visibles, et même que la démarche ne soit pas exempte d'une certaine circularité. Ce que l'on peut affirmer c'est que ces expériences ont donné des résultats en accord avec les prédictions quantiques, dans une situation qui serait « sensible » s'il existait des polariseurs à résultats bivalués.

Après les expériences pionnières du début des années 1970, on allait voir apparaître une nouvelle génération d'expériences lançant la course vers le Graal de « l'expérience idéale », avec des schémas toujours plus proches de l'expérience de pensée.

## 2.2.6 Les expériences de deuxième génération

Avec le développement des lasers, il devint possible au début des années 1980 de mettre au point une nouvelle source de paires de photons intriqués beaucoup plus intense (Fig. 2.8), qui donna des résultats beaucoup plus précis, et qui surtout permit de mettre en œuvre des schémas nettement plus proches de l'expérience de pensée. Ainsi, après une première expérience utilisant encore des polariseurs à une seule voie, mais ayant montré que l'intrication survit sans affaiblissement même si on éloigne les polariseurs à plus de six mètres de la source, A. Aspect, P. Grangier et G. Roger allaient développer des polariseurs à deux voies pour réaliser l'expérience de la figure 2.9, manifestement très proche du schéma idéal de la figure 2.3. En un seul enregistrement de durée 100 secondes, il était possible de mesurer, à un niveau de précision de 1 %, le coefficient de corrélation de polarisation  $E$  pour une orientation donnée  $(\mathbf{a}, \mathbf{b})$ . Pour tester les inégalités de Bell, il suffisait de répéter cette mesure en orientant les polariseurs suivant les directions « optimales » de la figure 2.5. En quelques minutes seulement, cette expérience a obtenu des résultats violant les





FIG. 2.8 – Expériences d’Orsay (1982) : la source de photons intriqués. On distingue les lasers permettant d’exciter les atomes de Calcium dans le niveau e (par absorption non linéaire des deux photons  $\nu_{Kr}$  et  $\nu_{col}$ ). Par désexcitation en cascade vers r puis f, l’atome émet la paire de photons intriqués  $\nu_1$  et  $\nu_2$ . Cette source d’une efficacité sans précédent a permis d’effectuer des mesures de corrélation de polarisation avec une précision de 1 % en moins de 2 minutes, autorisant la mise en œuvre de schémas expérimentaux nouveaux de plus en plus proches de l’expérience de pensée. Les photons étaient émis dans des faisceaux opposés, peu divergents, de quelques centimètres de diamètre seulement, ce qui a permis d’éloigner les polariseurs à 6 mètres de la source de photons située dans l’enceinte à vide qu’on distingue au centre de la photo. Avec cette géométrie, un signal se propageant à la vitesse de la lumière mettrait 40 nanosecondes pour aller d’un polariseur à l’autre.

inégalités de Bell par plus de 40 « barres d’incertitude », puisqu’on a obtenu

$$S_{exp} = 2,697 \pm 0,015 \tag{2.8}$$

ce qui dépasse de façon non ambiguë la limite supérieure des inégalités de Bell (2.4). Ce résultat est en excellent accord avec la prédiction quantique ( $S_{MQ} = 2,70 \pm 0,005$ ) prenant en compte les imperfections résiduelles des polariseurs. On notera que cette valeur est proche de la valeur maximale  $S_{MQ}^{max} = 2\sqrt{2} = 2,828...$  qui serait obtenue dans une expérience idéale (Éq. (2.7)).

Après ce premier pas significatif vers une expérience idéale, l’équipe d’Orsay allait s’attaquer à l’hypothèse de localité. Dans l’expérience de la figure 2.9, les polariseurs étaient éloignés de 12 mètres l’un de l’autre (Fig. 2.8), ce qui correspond à un temps de propagation de 40 nanosecondes pour la lumière, tandis que l’intervalle de temps entre les détections des deux photons d’une même paire n’excédait pas 10 nanosecondes : les détections étaient donc indépendantes au sens relativiste – on dit qu’elles étaient séparées par un « intervalle du genre espace » – puisque la vitesse de propagation d’un

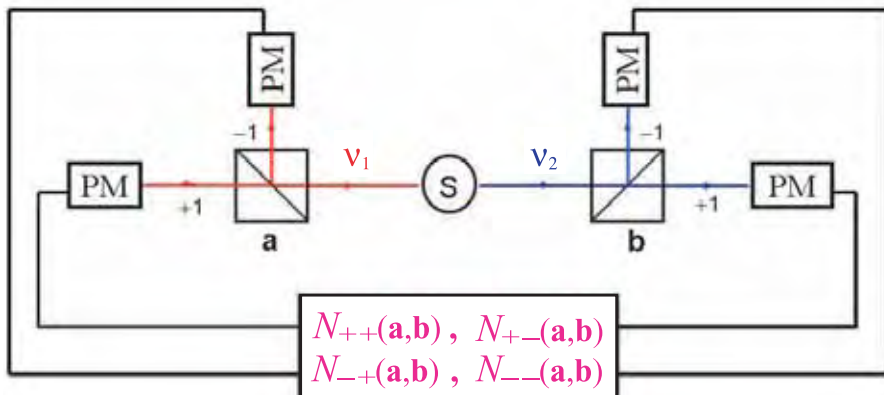


FIG. 2.9 – *Expérience avec polariseurs à deux voies. En mettant en œuvre des polariseurs à deux voies (issus de la technique des multicouches diélectriques) et un système de détection de coïncidences multiples (issu des techniques électroniques utilisées auprès des accélérateurs de particule), il a été possible de mesurer en un seul enregistrement les quatre nombres de coïncidences  $N_{++}(\mathbf{a}, \mathbf{b})$ ,  $N_{--}(\mathbf{a}, \mathbf{b})$ ,  $N_{+-}(\mathbf{a}, \mathbf{b})$ ,  $N_{-+}(\mathbf{a}, \mathbf{b})$  pour une orientation  $(\mathbf{a}, \mathbf{b})$  donnée, et d'en déduire le coefficient de corrélation de polarisation correspondant  $E(\mathbf{a}, \mathbf{b})$ . Grâce à la nouvelle source de paires de photons intriqués (Fig. 2.8) une mesure durant 100 secondes permettait d'obtenir  $E(\mathbf{a}, \mathbf{b})$  avec une précision de 1 %. Cette expérience a permis d'observer, en quelques minutes de mesure, une violation des inégalités de Bell par plus de 40 écarts-type (A. Aspect, P. Grangier et G. Roger, 1982).*

hypothétique signal connectant les détections aurait été supérieure à celle de la lumière, ce qui est interdit par la causalité d'Einstein. Il ne s'agissait pourtant pas encore de la mise en œuvre du schéma ultime suggéré par Bell, dans lequel les orientations des polariseurs devraient être modifiées aléatoirement à une cadence assez élevée pour que la relativité interdise toute dépendance directe entre le choix de l'orientation d'un polariseur et la mesure par l'autre (Fig. 2.6).

Un premier pas vers la réalisation de ce schéma idéal a été accompli en 1982. Comme il n'est pas question de modifier physiquement l'orientation de polariseurs massifs en quelques nanosecondes (aucun matériau n'y résisterait), A. Aspect, J. Dalibard et G. Roger ont contourné la difficulté en développant des « commutateurs optiques » rapides capable soit de laisser passer la lumière, soit de l'aiguiller vers une seconde voie de sortie (Fig. 2.10). Ainsi, tantôt le commutateur  $C_1$  laisse passer le photon  $\nu_1$  vers un polariseur dans l'orientation  $\mathbf{a}$ , tantôt il le défléchit vers un polariseur dans l'orientation  $\mathbf{a}'$ . L'ensemble constitué du commutateur et des deux polariseurs est équivalent à un seul polariseur basculant entre les orientations  $\mathbf{a}$  et  $\mathbf{a}'$ , à une cadence limitée seulement par le commutateur. Un second commutateur suivi de deux

polariseurs dans les orientations  $\mathbf{b}$  et  $\mathbf{b}'$  opère de même sur le photon  $\nu_2$ . Dans cette expérience, la distance entre les commutateurs était de douze mètres, ce qui correspond à un temps de propagation de quarante nanosecondes pour la lumière, nettement plus long que l'intervalle de temps entre deux basculements successifs des commutateurs (dix nanosecondes). Bien que moins précise que l'expérience de la figure 2.9, du fait de l'ouverture limitée des commutateurs qui réduisait les taux de comptage par un ordre de grandeur, cette expérience aboutit à une violation convaincante des inégalités de Bell (par 6 écarts type), dans une situation où la condition de localité n'était plus posée a priori comme une hypothèse raisonnable, mais pouvait être déduite du principe de causalité d'Einstein.

### 2.2.7 Vers une expérience idéale : la chasse aux échappatoires

En fait, l'expérience de la figure 2.10 ne satisfaisait pas totalement les conditions de l'expérience de pensée idéale de la figure 2.6. En effet, les changements d'orientation des polariseurs n'avaient pas vraiment lieu à des instants aléatoires. Pour des raisons liées aux limites technologiques de l'époque, on avait choisi de réaliser les commutateurs optiques à partir de modulateurs acousto-optiques nécessairement pilotés par des générateurs périodiques. Afin de se rapprocher le plus possible du schéma idéal, les deux générateurs pilotant les deux commutateurs éloignés opéraient à des fréquences différentes,

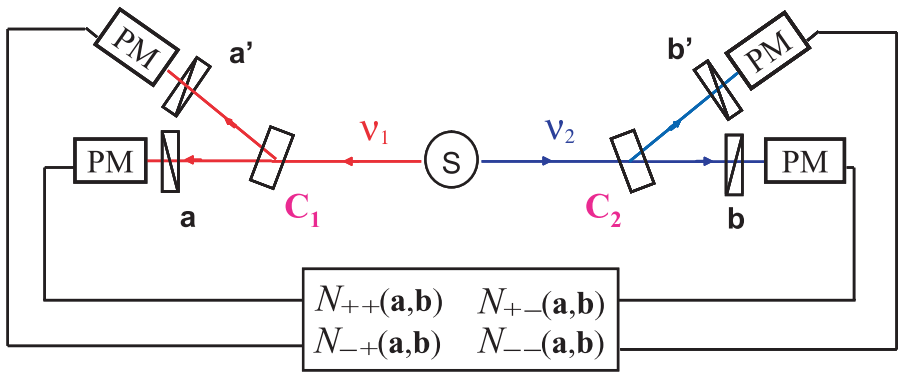


FIG. 2.10 – *Expérience avec polariseurs variables. Dans cette expérience, le commutateur optique rapide  $C_1$  permet d'aiguiller le photon  $\nu_1$  soit vers un polariseur orienté suivant  $\mathbf{a}$ , soit vers un polariseur orienté suivant  $\mathbf{a}'$  : l'ensemble est équivalent à un polariseur unique basculant rapidement (toutes les 10 nanosecondes environ) entre  $\mathbf{a}$  et  $\mathbf{a}'$ . Un système analogue réalise un polariseur variable basculant entre  $\mathbf{b}$  et  $\mathbf{b}'$ , pour analyser  $\nu_2$ . Les deux commutateurs étant distants de 12 mètres, les changements d'orientation sont séparés au sens relativiste (A. Aspect, J. Dalibard et G. Roger, 1982).*



FIG. 2.11 – *Expérience de Genève (1998)*. Dans cette expérience (W. Tittel, J. Brendel, H. Zbinden, et N. Gisin) les photons intriqués se propagent dans les fibres optiques du réseau commercial de télécommunication. La source (située à Cornavin) est à plus de dix kilomètres des détecteurs (Bellevue et Bernex).

et de plus leurs phases dérivait de façon indépendante : les changements d'orientations aux deux extrémités apparaissaient donc statistiquement indépendants. Une expérience avec des changements à des instants véritablement aléatoires restait pourtant souhaitable. C'est un nouveau progrès majeur sur les sources de paires de photons intriqués qui allait permettre un nouveau pas vers l'expérience idéale.

Vers la fin des années 1980, on commença en effet à développer une troisième génération de sources de paires de photons intriqués, basées sur des effets d'optique non linéaire non plus dans des atomes comme la source d'Orsay, mais dans des cristaux anisotropes. Le progrès majeur qui en est résulté est la maîtrise de la direction d'émission des photons intriqués, ce qui permet d'injecter les deux membres de chaque paire dans deux fibres optiques dirigées vers des directions opposées. On a ainsi pu réaliser des expériences avec des distances de source à détecteur de plusieurs centaines de mètres, voire de dizaines de kilomètres comme dans l'expérience de Genève qui utilise le réseau commercial de fibres optiques de la compagnie suisse de télécommunications (Fig. 2.11).

Avec de telles distances entre source et stations de mesure, il devient possible de choisir de façon totalement aléatoire l'orientation de chaque polariseur

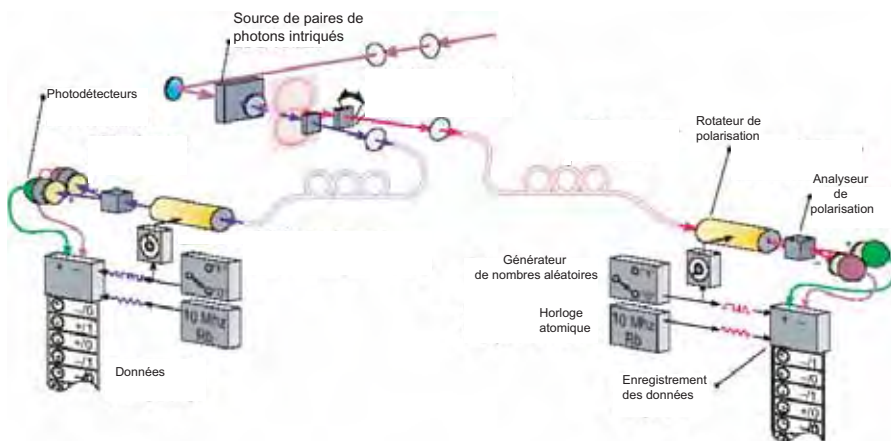


FIG. 2.12 – *Expérience d'Innsbruck (1998)*. La distance entre source et polariseurs (plus de 300 mètres de fibre optique) est suffisante pour permettre un tirage aléatoire de l'orientation de chaque polariseur, pendant le temps de propagation des photons (de l'ordre de une microseconde) (G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, et A. Zeilinger)

pendant le temps de propagation des photons depuis la source. Une telle expérience, suivant exactement le schéma de la figure 2.6 a été réalisée à Innsbruck en 1999 (Fig. 2.12), confirmant sans ambiguïté la violation des inégalités de Bell dans une situation où la condition de localité est une conséquence de la causalité relativiste.

Toutes ces expériences ont confirmé la violation des inégalités de Bell, à un niveau de précision étonnant (on a ainsi rapporté une violation des inégalités de Bell par plus de 100 écarts type). À ce stade de raffinement, on peut se demander si on a complètement atteint l'idéal de l'expérience de pensée. En fait, pour l'avocat déterminé des théories à variables cachées locales, il reste encore une échappatoire, liée à la détectivité imparfaite des détecteurs de photons placés dans les voies  $+1$  et  $-1$  des polariseurs. Il se trouve en effet qu'en régime de comptage de photons aucun détecteur n'a une sensibilité lui permettant de détecter un photon avec une probabilité de 100 %, loin de là. Cette probabilité est plutôt de l'ordre de 10 à 20 %. On peut alors, sans insulter la logique, imaginer que le détecteur « conspire » avec le polariseur pour sélectionner un sous ensemble de photons détectés non représentatif de la totalité des photons, mais choisis pour donner des résultats reproduisant les résultats quantiques. Un certain nombre de modèles, évidemment ad hoc, ont été développés pour montrer que cette position n'est pas totalement intenable. Pour fermer cette échappatoire, il faudrait disposer de détecteurs de sensibilité suffisante pour ne manquer que très peu de photons, ne laissant plus de place à la conspiration. On montre qu'à partir de 70 % de probabilité

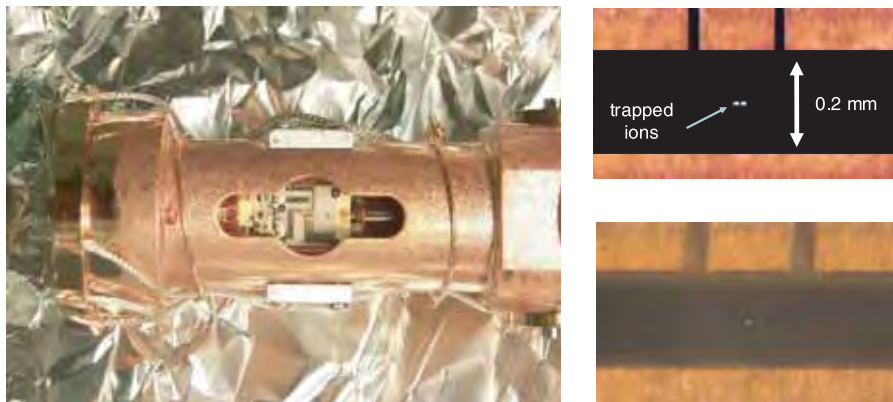


FIG. 2.13 – *Expérience avec détection de grande sensibilité (Boulder, 2001). Dans cette expérience (D. Wineland et col., 2001), les observables intriquées sont non pas deux polarisations de deux photons, mais deux états internes de deux ions. Grâce à une méthode de détection par fluorescence induite par laser, il est possible de distinguer sans ambiguïté entre un état qui ne fluoresce pas (état sombre) et l'autre état qui fluoresce (état brillant). La photographie de gauche montre le piège qui permet de maintenir les deux ions à quelques micromètres l'un de l'autre dans un état intriqué. Les deux vues de droite, considérablement agrandies, montrent des situations différentes où un seul ion, ou les deux, sont dans l'état brillant. Cette expérience a permis dans une large mesure de fermer l'échappatoire de la détectivité insuffisante.*

de détection, la plupart des conspirations imaginables sont éliminées, et avec une détectivité de 80 %, une violation des inégalités de Bell serait irréfutable. Or, comme John Bell l'a fait remarquer il y a longtemps [7], nous ne connaissons aucune raison physique fondamentale empêchant les détecteurs de photons d'avoir un rendement de 100 %, et c'est un fait d'expérience que les progrès technologiques ont permis une amélioration jamais démentie de ces rendements au cours des décennies. On peut donc, avec John Bell, être sceptique sur la possibilité qu'un phénomène physique aussi fondamental que l'intrication quantique change de nature à la faveur d'une simple amélioration d'un paramètre des détecteurs. Mais il n'en restait pas moins souhaitable de fermer l'échappatoire de la détectivité.

En fait, c'est en passant des photons aux ions que l'échappatoire de la détectivité a pu être fermée en 2001. Il a en effet été possible d'intriquer non pas deux polarisations de deux photons mais deux états quantiques internes de deux ions. Or, par une méthode de fluorescence induite par laser (Fig. 2.13), il est possible de distinguer entre ces deux états avec certitude, la probabilité de « manquer » le résultat étant très faible : la situation est donc équivalente à celle où un polariseur à deux voies serait suivi de deux détecteurs de détectivité

parfaite. Même si cette expérience souffre du fait que l'intrication de ces paires d'ions est loin d'être aussi forte que celle obtenue avec les photons, l'expérience a permis de conclure à une violation non ambiguë des inégalités de Bell, sans échappatoire liée à une détectivité insuffisante.

Peut-on alors considérer que l'expérience idéale a été réalisée, reproduisant strictement toutes les caractéristiques de l'expérience idéale, la « *Gedanken-Experiment* » sur laquelle s'appuient les raisonnements théoriques ? À ce stade, on ne peut pas complètement l'affirmer, pour la raison suivante : dans l'expérience de Boulder, les deux ions intriqués sont à des distances microscopiques, de l'ordre du micromètre, et les détections ne sont manifestement pas séparées au sens relativiste du terme : un signal se propageant à la vitesse de la lumière pourrait connecter les deux détections. À plus forte raison, les choix de l'observable mesurée sur chaque ion – l'équivalent de l'orientation du polariseur – ne sont pas faits (et ne pourraient l'être) de façons séparées au sens relativiste. Ainsi, l'expérience reste à faire dans laquelle on aurait d'une part des détections parfaitement efficaces, et d'autre part des mesures suffisamment éloignées pour assurer la séparation relativiste des mesures et des choix des observables mesurées.

Il n'en est pas moins vrai que l'on dispose aujourd'hui d'un ensemble impressionnant de données expérimentales qui violent sans ambiguïté les inégalités de Bell, dans des schémas extrêmement variés, permettant ainsi de clore séparément les diverses échappatoires qui ont été envisagées dans des discussions d'un très grande rigueur. De plus, il faut noter que ces expériences ont non seulement montré une violation non ambiguë des inégalités de Bell, mais aussi permis de confronter avec une très grande précision les résultats de mesure aux prédictions de la mécanique quantique, parfaitement précises dès lors que les situations expérimentales sont bien contrôlées. L'accord observé est impressionnant, et il est difficile d'imaginer une « conspiration » d'un monde non quantique capable de simuler les prédictions de la mécanique quantique à ce niveau de précision et de raffinement. Même si une expérience ultime serait la bienvenue<sup>6</sup>, nous pouvons raisonnablement considérer que les inégalités de Bell ont été violées, que les prédictions quantiques stupéfiantes relatives aux systèmes intriqués ont été confirmées, et essayer d'en tirer les conséquences.

### 2.2.8 Conclusion : la non localité quantique

Quelle conclusions tirer de la violation des inégalités de Bell ? Tout d'abord nous devons accepter l'idée que le monde ne peut pas toujours se concevoir comme formé de sous-systèmes séparés, dont les propriétés physiques seraient définies localement et ne sauraient s'influencer mutuellement lorsque les systèmes sont séparés au sens relativiste. Cette notion de « séparabilité » semblait pourtant tellement fondamentale à Einstein qu'il en avait fait la pierre angulaire de sa démonstration de la nécessité de compléter la mécanique quantique :

---

<sup>6</sup>Pour connaître l'état de l'art, voir par exemple la proposition et les références dans [10].

« On peut échapper à cette conclusion [la mécanique quantique est incomplète, note de l'auteur] seulement en acceptant que la mesure de S1 change la situation réelle de S2 (par télépathie) ou en niant l'indépendance des situations réelles de tels objets, spatialement séparés l'un de l'autre. Les deux termes de l'alternative me semblent également inacceptables. » [11]. Aujourd'hui, avec la violation des inégalités de Bell, nous ne pouvons plus rejeter cette alternative, et il nous faut renoncer à la vision « réaliste locale » du monde que défendait Einstein.

Avant de discuter ce renoncement, il convient de replacer la position d'Einstein dans une perspective historique, afin de ne pas commettre d'anachronisme. Lorsqu'il écrit cette phrase, en 1949, Einstein ne sait pas qu'il y a une incompatibilité irréductible entre sa vision du monde et les prédictions quantitatives de la mécanique quantique, puisque cette incompatibilité ne sera établie qu'en 1964 par Bell. Pour lui, la divergence ne porte que sur l'interprétation de la théorie, pas sur son application aux phénomènes physiques, comme il l'écrit en 1948 dans une lettre à Max Born [12] : « *Il ne semble pas faire de doute que les physiciens qui tiennent la description donnée par la mécanique quantique pour définitive en son principe réagiront à ces considérations [le raisonnement EPR, note de l'auteur] de la façon suivante : ils laisseront tomber l'exigence d'une existence autonome de la réalité physique présente en diverses portions de l'espace ; ils peuvent invoquer à bon droit le fait que la théorie quantique ne fait nulle part explicitement usage de cette exigence. Dont acte, mais je note ceci : lorsque je considère les phénomènes physiques que je connais, y compris ceux que la mécanique quantique a étudiés avec tant de succès, je ne trouve nulle part de fait qui me fasse paraître vraisemblable que l'on veuille abandonner cette exigence<sup>7</sup>. Aussi suis-je enclin à croire que l'on doit considérer la description donnée par la mécanique quantique comme une description incomplète et indirecte de la réalité, destinée à être remplacée plus tard par une description exhaustive et directe* ». Même si nul n'est autorisé à dire quelle aurait été la réaction d'Einstein face au théorème de Bell établissant l'incompatibilité entre le réalisme local et la mécanique quantique, et face aux expériences tranchant en faveur de la mécanique quantique, il nous semble légitime de nous laisser guider par ces citations d'Einstein, et de pousser son raisonnement à son terme logique : *nous avons aujourd'hui des faits qui nous amènent à renoncer à une description réaliste locale des systèmes intriqués*.

On a parfois pu se demander quel était le concept à abandonner, entre la localité d'une part, et le réalisme de l'autre. Il nous semble pour notre part difficile d'envisager ces deux notions comme indépendantes : comment concevoir des réalités physiques indépendantes pour deux systèmes séparés spatialement mais qui resteraient en contact via une interaction instantanée, supraluminale ? *Pour nous, la non localité de la mécanique quantique – souvent présentée comme la conclusion à tirer de la violation des inégalités de Bell –*

---

<sup>7</sup>Souligné par nous.



correspond à une négation en bloc de la vision réaliste locale du monde : les photons intriqués jumeaux ne sont pas deux systèmes distincts portant deux copies identiques d'un même ensemble de paramètres. Une paire de photons intriqués doit être considérée comme un système unique, inséparable, décrit par un état quantique global, impossible à décomposer en deux états relatifs à chacun des deux photons : les propriétés de la paire ne se résument pas à la réunion des propriétés des deux photons. Pour souligner cette caractéristique du formalisme quantique, qui contraste avec une conception réaliste locale, on peut parler de « holisme quantique » [13, 14].

Soulignons ici qu'une utilisation insuffisamment précautionneuse de l'expression « non localité » pourrait amener, à tort, à conclure à la possibilité de l'envoi d'un signal utilisable à une vitesse supérieure à celle de la lumière. Reprenant le schéma de la figure 2.6, on pourrait en effet imaginer qu'une mesure par le polariseur II de la polarisation du photon  $\nu_2$  pourrait permettre de connaître instantanément la direction du polariseur I qui vient juste d'effectuer une mesure sur le photon  $\nu_1$ . Imaginons en effet qu'une mesure sur  $\nu_1$ , effectuée suivant la direction  $\mathbf{a}$ , donne le résultat  $+1$ . On peut alors, dans une certaine mesure, dire que le photon  $\nu_2$  prend instantanément, au moment de la mesure de  $\nu_1$ , une polarisation linéaire parallèle à  $\mathbf{a}$ . Ne peut-on prendre ainsi connaissance en II, instantanément, de la valeur de l'orientation  $\mathbf{a}$  qui pourrait être choisie au dernier moment en I ? On aurait ainsi transmis une information (l'orientation  $\mathbf{a}$ ), plus vite que la lumière. En fait, on montre – en se référant aux principes de base de la mécanique quantique – qu'il est impossible de déterminer complètement l'état d'un système unique, ici la direction suivant laquelle le photon  $\nu_2$  est polarisé. Pour déterminer complètement l'état d'un système, il faudrait être capable d'en préparer un grand nombre de copies, toutes exactement dans le même état, et de multiplier les mesures en changeant d'observable (ici avec une série de polariseurs orientés différemment). Mais en fait il est impossible de dupliquer un système quantique, c'est-à-dire d'obtenir deux systèmes strictement dans le même état que le système initial : c'est le « théorème de non clonage quantique » (voir appendice 1). On ne peut donc connaître l'orientation de l'état de polarisation de  $\nu_2$ , et il faut renoncer au rêve, cher aux auteurs de science fiction, d'un télégraphe supra luminal.

À l'issue d'un voyage de sept décennies, qui nous a mené de l'article EPR (1935) aux inégalités de Bell (1964) et aux expériences qu'elles ont suscitées, on pourrait avoir le sentiment frustrant d'une série de conclusions négatives : les propriétés avérées de l'intrication quantique nous forcent à renoncer à une vision réaliste locale ; le théorème de non clonage quantique interdit le télégraphe supra luminal... En fait, ces renoncements sont porteurs de formidables progrès potentiels : on commence à savoir tirer parti des propriétés quantiques que nous venons de voir, dans des concepts nouveaux de traitement et de transmission de l'information, où l'intrication quantique et le théorème de non clonage jouent un rôle central : c'est le domaine de l'information quantique, que nous présentons maintenant.

## 2.3 L'information quantique

### 2.3.1 Introduction

Bien que les implications de la révolution conceptuelle que nous venons de présenter n'aient sans doute pas été encore toutes explorées, il est légitime de se demander quelles peuvent être ses conséquences pratiques. En fait, l'examen approfondi des concepts quantiques a ouvert vers le milieu des années 1980 un nouveau domaine de recherche, devenu extrêmement actif, appelé « information quantique ». L'idée directrice de l'information quantique est que l'on peut, en utilisant la physique quantique, concevoir de nouvelles façon de calculer et de communiquer, dont les « règles du jeu » ne sont plus celles que l'on connaissait classiquement.

On peut alors concevoir de nouvelles méthodes de cryptographie dont la sécurité s'appuie sur les bases même de la physique, ou de nouvelles méthodes de calculs qui peuvent être exponentiellement plus efficaces que les méthodes classiques. L'information quantique ne concerne donc pas seulement les physiciens, mais aussi les théoriciens de l'information, les algorithmiciens, et les mathématiciens travaillant sur la théorie de la complexité. Ces recherches ont ainsi permis de proposer de nouveaux algorithmes, et aussi de nouvelles architectures de calcul, basées sur des « portes logiques quantiques » sans équivalent classique. Toujours sur le plan fondamental, la rencontre entre la théorie de l'information et la mécanique quantique – qui constitue le cœur de l'information quantique – conduit à un renouvellement très stimulant des outils théoriques utilisés de part et d'autre. On peut ainsi envisager de nouvelles approches des principes fondamentaux de la théorie quantique, associées à de nouvelles façons de définir et de traiter l'information. Selon la célèbre formule de Rolf Landauer : « l'information est de nature physique », et il n'est finalement pas surprenant que la mécanique quantique, qui sous-tend toute la physique actuelle, se découvre des liens étroits avec la théorie de l'information.

### 2.3.2 Paires de photons et cryptographie quantique

Un premier exemple d'utilisation de ces concepts est la cryptographie quantique. De façon générale, le but de la cryptographie est la transmission d'un message secret entre un émetteur (Alice) et un récepteur (Bob), en minimisant les risques qu'un espion puisse intercepter et décoder ce message. La cryptographie joue depuis très longtemps un rôle important pour protéger des secrets commerciaux ou militaires, mais elle est aussi devenue l'affaire de tous avec la généralisation des échanges électroniques « grand public », qui vont des cartes bancaires aux achats sur internet. La cryptographie classique utilise le plus souvent des algorithmes de codage sophistiqués, qui ne peuvent être « cassés » en un temps raisonnable compte-tenu des moyens de calcul actuellement disponibles. La sécurité ainsi obtenue est acceptable, mais elle

n'est pas absolue, car elle dépend des moyens mis en jeu par l'adversaire. De plus, elle n'est en général pas démontrable mathématiquement.

Il existe cependant une méthode de cryptographie simple et « inconditionnellement sûre » d'un point de vue mathématique, qui repose sur le fait qu'Alice et Bob ont échangé par avance une « clé secrète », c'est-à-dire une longue suite de caractères aléatoires connus d'eux seuls. Si cette clé est aussi longue que le message, et si elle est utilisée une seule fois, alors la sécurité absolue du cryptage résulte d'un théorème mathématique, démontré par Claude Shannon en 1948. Ce théorème reporte donc la sécurité de la communication sur la sécurité du partage de la clé. C'est à ce stade qu'intervient la cryptographie quantique : elle permet à Alice et Bob d'échanger une clé secrète, avec une sécurité garantie par les principes mêmes de la physique quantique !

Il existe actuellement de nombreux « protocoles » de cryptographie quantique, et nous allons présenter ici une méthode qui découle directement de l'expérience EPR présentée dans la première partie de cet article. Ceci mettra clairement en évidence à nouveau que « ce qui est transmis » dans une expérience EPR n'est pas un message, mais une suite de nombres aléatoires corrélés, c'est-à-dire précisément une clé secrète ! Suivant le principe exposé ci-dessus, cette clé pourra ensuite être utilisée pour crypter le « vrai » message, avec une sécurité démontrée mathématiquement.

Supposons donc qu'Alice et Bob se partagent des paires de photons intriqués, suivant le protocole déjà présenté. Alice et Bob peuvent choisir arbitrairement les mesures qu'ils vont effectuer sur leur photon, mais l'espion (Ève) ne peut pas connaître ces mesures lorsque le photon est transmis dans la ligne. Par ailleurs, Ève ne peut pas « cloner » le photon qui lui parvient, c'est-à-dire en faire une copie à l'identique (voir appendice 1). En fait, nous allons voir que toute tentative d'Ève pour intercepter le photon va perturber l'état de ce photon, et créer des erreurs de transmission qui pourront être détectées par Alice et Bob. En revanche, en l'absence de toute erreur de transmission, Alice et Bob sauront qu'il n'y a pas d'espion sur la ligne.

Plus précisément, Alice et Bob conviennent d'effectuer leurs mesures suivant 4 états de polarisation linéaire, orientés suivant un axe horizontal (noté  $h = |x\rangle$ ), vertical (noté  $v = |y\rangle$ ), incliné à 45 degrés à droite (noté  $d = |x'\rangle$ ), ou à 45 degrés à gauche (noté  $g = |y'\rangle$ ). Les états orthogonaux  $|x\rangle$  et  $|y\rangle$  sont faciles à distinguer l'un de l'autre, car ils donnent les résultats « + » et « - » pour une même orientation du polariseur, appelée « base  $hv$  ». De même, les états  $|x'\rangle$  et  $|y'\rangle$  donnent les résultats « + » et « - » pour une orientation du polariseur à 45 degrés de  $hv$ , appelée « base  $dg$  ». En revanche, les bases  $hv$  et  $dg$  sont dites « incompatibles » : si l'état de polarisation est connu dans l'une des bases, il est complètement aléatoire dans l'autre.

Après avoir reçu les photons, Bob va diffuser publiquement l'ensemble de ses choix d'axes de mesures,  $hv$  ou  $dg$ , ainsi qu'une fraction des résultats obtenus, + ou -. À l'examen de ces résultats, Alice peut détecter la présence éventuelle d'un espion, en faisant le raisonnement suivant. L'espion Ève ne connaît pas plus que Bob l'orientation  $hv$  ou  $dg$  qu'elle a choisi pour mesurer

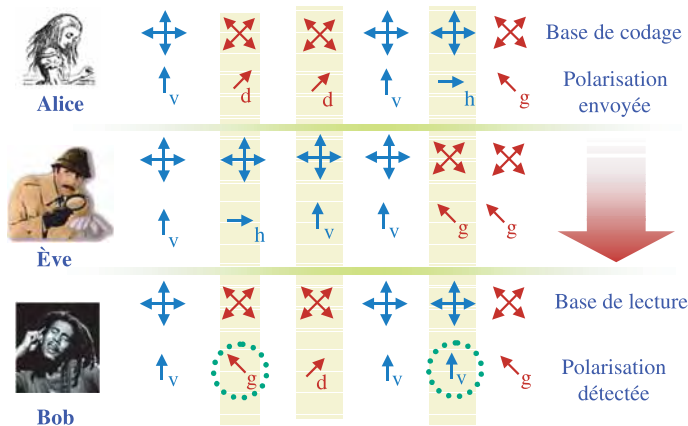


FIG. 2.14 – Protocole de cryptographie quantique. On n'a illustré que des situations où Alice et Bob ont choisi la même base, ce qui se produit en moyenne une fois sur deux. Dans la deuxième colonne, Alice a détecté un photon  $d$ , mais Bob a détecté un photon  $g$  : cette erreur a été provoquée par l'intervention d'Ève, qui a détecté et réémis un photon  $h$ .

chaque photon qu'elle a reçu. Supposons donc qu'Ève oriente elle aussi son polariseur de manière arbitraire selon  $hv$  ou  $dg$ , et ré-émette à chaque détection un photon dont l'état de polarisation est identique à celui qu'elle vient de mesurer. Ainsi, si elle choisit  $hv$  et mesure  $+$ , elle ré-émet vers Bob un photon polarisé suivant  $h$ . Mais cette intervention d'Ève est en fait détectable, car elle introduit des erreurs au niveau des détections de Bob.

Considérons par exemple le cas illustré sur la figure 2.14 (deuxième colonne), où Alice a détecté un photon  $d$ , et où Bob a également orienté son polariseur dans la base  $dg$ , mais où Ève a orienté le sien selon  $hv$ . Ève va alors mesurer  $+$  avec une probabilité  $1/2$ , et  $-$  avec une probabilité  $1/2$ . Selon son résultat, elle ré-émet ensuite vers Bob un photon dans l'état  $h$  ou  $v$ . Dans les deux cas, avec son polariseur orienté selon  $dg$ , Bob peut mesurer  $+$  ( $d$ ) avec une probabilité  $1/2$ , et  $-$  ( $g$ ) avec une probabilité  $1/2$ . Au contraire, si Ève n'avait pas été présente, Bob aurait dû mesurer  $d$  avec une probabilité de 1. L'intervention de l'espion crée donc des erreurs dans 25 % des cas, et dans ce cas Alice et Bob le détectent et interrompent la transmission.

De façon plus générale, il y aura toujours des erreurs sur la ligne de transmission, dues le plus souvent à des imperfections techniques. Par prudence, Alice et Bob doivent attribuer toutes ces erreurs à un espion potentiel. On pourrait en conclure que la ligne sera alors inutilisable, mais ce n'est pas le cas : en fait, Alice et Bob vont tout d'abord évaluer le taux d'erreur de la ligne, en utilisant des données « tests » que Bob révèle publiquement. La connaissance de ce taux d'erreur va leur permettre de borner quantitativement la quantité d'information dont a pu s'emparer Ève : plus le taux d'erreur est

petit, plus l'information sur les photons transmis disponible pour Ève sera petite également. La suite de bits échangés et non divulgués par Bob va alors constituer la base de la clé secrète. Pour que cette clé soit utilisable, Alice et Bob doivent éliminer les erreurs qu'elle peut contenir, en utilisant des protocoles classiques de correction d'erreurs. Ensuite, en fonction du taux d'erreur qu'ils ont mesuré, ils vont réduire encore le nombre de bits utiles pour obtenir une clé finale plus petite, mais totalement inconnue d'Ève. Un taux d'erreur de plus en plus élevé sur la ligne ne compromettra pas la sécurité de cette clé finale, mais sa taille : on montre que si le taux d'erreur sur les photons mesurés est supérieur à 11 %, le protocole de correction ne produira finalement aucun bit secret. En revanche, pour un taux d'erreur inférieur à 11 %, Alice et Bob disposeront d'une clé sans erreurs et parfaitement sûre. La présence d'erreurs techniques, produisant typiquement des taux d'erreurs de quelques pourcents, n'empêche donc pas de transmettre un clé secrète.

Nous avons supposé qu'Ève pouvait choisir de manière arbitraire l'orientation de son polariseur pour chaque photon, puis réémettre vers Bob un photon polarisé correspondant à son résultat de mesure. On peut se demander s'il s'agit de la meilleure stratégie pour lui éviter d'être détectée. En particulier, si elle pouvait « cloner » (ou dupliquer) le photon sans modifier son état de polarisation, il lui serait possible de renvoyer vers Bob un des deux clones, tout en gardant l'autre pour effectuer sa propre mesure, et l'espionnage serait alors indétectable. Ce clonage d'un état inconnu est (heureusement pour Alice et Bob) impossible en mécanique quantique, comme l'on démontré W.K. Wootters et W.H. Zurek [15]. En effet, on ne peut pas produire de manière fiable une ou plusieurs copies d'un état quantique, à moins que cet état ne soit partiellement connu auparavant (voir appendice 1).

Le raisonnement que nous venons de faire s'appuie sur le partage de paires de photons intriqués entre Alice et Bob. En fait, la source EPR peut être placée chez Alice, si bien qu'Ève et Bob n'ont accès qu'au deuxième photon de la paire. On montre alors que rien ne change ni pour Ève ni pour Bob si Alice n'utilise pas la source EPR, mais envoie simplement à Bob un photon polarisé suivant une des 4 directions  $h$ ,  $v$ ,  $d$  ou  $g$ , qu'elle a choisie aléatoirement. Plus généralement, on peut montrer que la sécurité du protocole cryptographique ne dépend pas de la transmission effective d'intrication, mais plutôt de la capacité du canal à transmettre de l'intrication. Le seuil d'erreur qui empêche la production d'une clé secrète est ainsi aussi celui où l'intrication de la paire EPR serait détruite par l'attaque d'Ève. Dans le cas où Alice envoie simplement un photon polarisé à Bob, l'équivalent du protocole que nous venons de présenter a été proposé par Charles Bennett et Gilles Brassard en 1984, et est connu sous l'acronyme de « protocole BB84 ».

### 2.3.3 Expériences de cryptographie quantique

Dans les montages réels, diverse méthodes sont utilisées pour coder de l'information sur les photons. Par exemple, pour le codage en polarisation

introduit dans le paragraphe précédent, Alice utilise quatre états définissant deux bases non orthogonales, qui permettent chacune de coder les bits 0 et 1, en choisissant par exemple  $h = 1$ ,  $v = 0$ ,  $d = 1$ ,  $g = 0$ .

Un enjeu important des montages réels de cryptographie quantique est d'obtenir une distance de transmission suffisante. Ces distances sont actuellement de l'ordre de quelques dizaine de kilomètres, en utilisant des techniques empruntées aux télécommunications optiques, et en particulier en transmettant les photons dans des fibres optiques.

Un autre point important à considérer est la source lumineuse utilisée. Le théorème de non clonage quantique, essentiel pour garantir la sécurité du système, s'applique à des photons considérés individuellement. Au contraire, les impulsions lumineuses généralement utilisées dans les systèmes de télécommunications contiennent un très grand nombre de photons, typiquement plus d'un million. Si l'on utilise un codage en polarisation pour de telles impulsions, le théorème de non-clonage quantique ne s'applique pas car l'espion peut adopter la tactique suivante : il prélève pour chaque impulsion une petite partie de la lumière envoyée par Alice, et il peut ainsi identifier la polarisation utilisée, en introduisant des erreurs négligeables. Pour garantir la sécurité du message, il faut donc en principe que chaque impulsion contienne un photon et un seul. Des expériences récentes [16,17] ont mis en œuvre des sources lumineuses émettant des photons individuels « à la demande » dans des dispositifs complets de cryptographie quantique (voir Fig. 2.15).

La réalisation d'une source à un photon demeurant relativement complexe, les systèmes commercialisés actuellement se contentent souvent de la méthode suivante : Alice atténue fortement les impulsions pour que la probabilité  $p$  d'avoir un photon dans chaque impulsion soit petite devant un. La probabilité d'avoir deux photons dans la même impulsion sera alors  $p^2/2 \ll p$ , ce qui signifie que les impulsions à deux photons (ou davantage) seront très peu nombreuses. Évidemment, puisque  $p \ll 1$ , la plupart des impulsions ne contiendront aucun photon, ce qui est un défaut de la méthode et diminue le taux de transmission de manière significative. On considère généralement qu'une valeur de  $p$  comprise entre 0,01 et 0,1 constitue un compromis acceptable [18]. La source est alors un laser impulsif fortement atténué, et le codage de l'information s'effectue directement dans la fibre optique grâce à des modulateurs intégrés. Les impulsions atténuées sont détectées grâce à des « photodiodes à avalanche », qui transforment un photon unique en impulsion électrique macroscopique grâce à un processus de multiplication d'électrons. Afin d'identifier sans ambiguïté les photons émis par Alice et détectés par Bob, des impulsions électriques synchrones du laser émetteur sont transmises par voie conventionnelle, et jouent le rôle d'horloge. Finalement, une gestion informatique est indispensable pour réaliser toutes les procédures décrites dans le paragraphe précédent, et en particulier pour corriger les erreurs et « distiller » la clé finale. À l'heure actuelle, les systèmes qui ont été réalisés ont permis de tester divers paramètres pertinents, comme la distance et le taux de transmission, les taux d'erreurs... Le développement de ces systèmes a pour l'instant

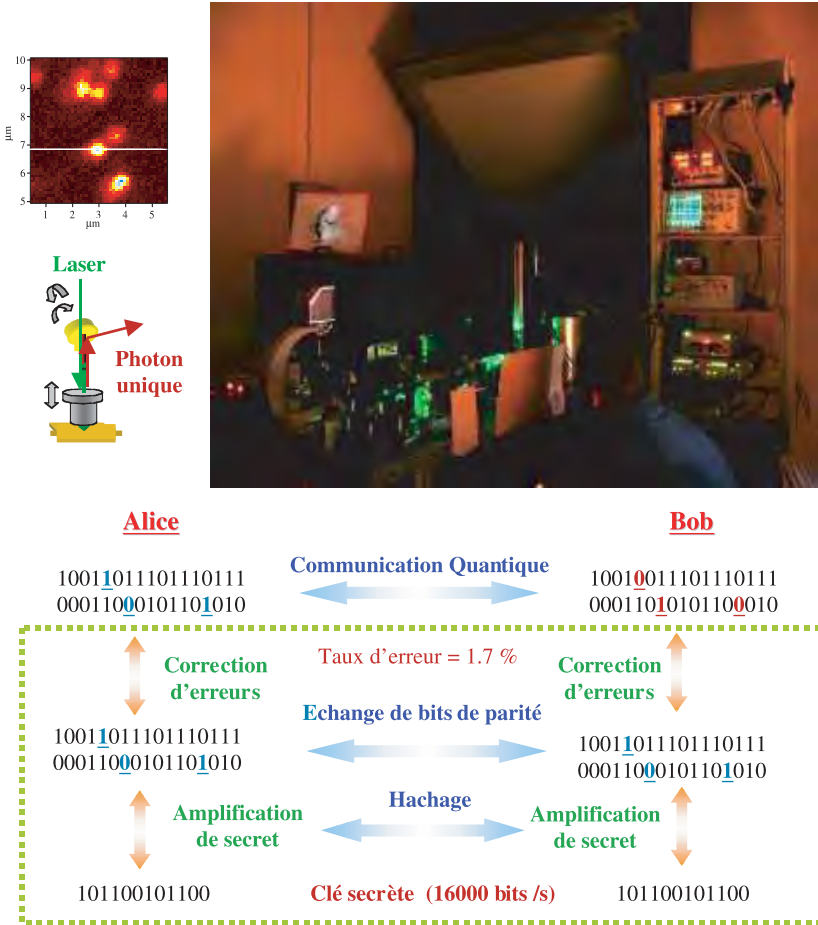


FIG. 2.15 – Échange quantique de clé utilisant une source de photons uniques, obtenue en excitant des nanocristaux de diamant par un laser impulsif. La photo en haut à gauche montre l'émission de lumière par les nanocristaux de diamant (points lumineux sur la photo). La photo de droite représente le montage expérimental d'Alice, et les photons sont envoyés (par la fenêtre) vers le montage de Bob situé dans un autre bâtiment. La partie inférieure de la figure montre les différentes étapes du protocole aboutissant à la constitution de la clé secrète.

un caractère prospectif, car il entre en concurrence avec les systèmes cryptographiques conventionnels (non quantiques), qui sont considérés comme très sûrs par leurs utilisateurs, civils ou militaires. Pourtant, cette confiance a été un peu ébranlée en 1994 ; ce point sera détaillé dans le paragraphe suivant, consacré à l'ordinateur quantique.

### 2.3.4 L'ordinateur quantique

#### Les bits quantiques ou « qubits »

Nous avons vu dans le paragraphe précédent qu'il est possible de coder un bit d'information (0 ou 1) sur deux états orthogonaux d'un photon polarisé. Que se passe-t-il alors, du point de vue du contenu d'information, si le photon est placée dans une combinaison linéaire quantique de ces deux états ? En termes imagés, le bit ne vaudra plus 0 ou 1, mais sera dans une superposition linéaire de ces deux valeurs. Pour prendre en compte cette possibilité, on introduit la notion de « qubit », ou bit quantique, qui contrairement au bit classique, admet l'existence de telles superpositions linéaires. Nous allons voir que cette notion a des implications très intéressantes si l'on considère un ordinateur quantique, basé sur la manipulation d'un grand nombre de qubits.

Nous prendrons une définition très simplifiée d'un ordinateur, en le considérant comme un appareil capable d'effectuer des opérations sur des ensembles de  $N$  bits appelés « registres ». Le contenu d'un registre est un mot binaire, qui représente un nombre mémorisé par l'ordinateur. Pour  $N = 3$ , on a ainsi  $2^3 = 8$  mots possibles, qui sont les triplets 111, 110, 101, 011, 100, 010, 001, 000. Considérons maintenant un q-registre, formé d'un ensemble de  $N$  qubits. Les  $2^N$  états possibles du registre classique vont alors définir une base de l'espace des états du q-registre, qui pourra quant à lui être placé dans une superposition linéaire arbitraire de tous les états de base :

$$|\Psi\rangle = c_0 |000\rangle + c_1 |001\rangle + c_2 |010\rangle + c_3 |011\rangle \\ + c_4 |100\rangle + c_5 |101\rangle + c_6 |110\rangle + c_7 |111\rangle \quad (2.9)$$

Supposons maintenant que l'ordinateur calcule, c'est-à-dire effectue une opération sur l'état du q-registre. Puisque cette opération est réalisée sur une superposition linéaire d'états, on peut considérer qu'elle est effectuée « en parallèle » sur les  $2^N$  nombres classiques. Au contraire, un ordinateur classique devrait effectuer une à une les  $2^N$  opérations. Cette notion de parallélisme quantique est à la base d'un gain d'efficacité de l'ordinateur, qui peut en principe être exponentiel si les  $2^N$  calculs correspondant à  $N$  qubits sont effectivement réalisés simultanément.

Plusieurs questions se posent immédiatement : sur le plan fondamental, quel type de calculs et quel type d'algorithmes peut-on effectuer avec un tel dispositif, sur le plan pratique, comment peut-on envisager de le réaliser ?

#### L'algorithme de factorisation de Peter Shor

Dans le paragraphe précédent, nous avons fait allusion aux systèmes cryptographiques non quantiques, qui sont souvent appelés protocoles algorithmiques. Un de ces protocoles est fondé sur le fait que certaines opérations mathématiques sont très faciles à réaliser dans un sens, mais beaucoup plus difficiles à réaliser dans l'autre. Par exemple, il est simple et rapide pour un



ordinateur de calculer le produit de deux nombres; en revanche, il est en général beaucoup plus difficile de décomposer un produit en ses facteurs premiers. Ainsi, si l'on considère le produit  $P$  de deux grands nombres premiers, il faut effectuer approximativement  $\sqrt{P}$  divisions pour identifier les facteurs. Le temps de calcul est alors de la forme  $e^{a \log P}$ , c'est-à-dire qu'il augmente exponentiellement avec le nombre de chiffres (ou de bits) de  $P$ , et devient rapidement rédhibitoire<sup>8</sup>. La méthode cryptographique « RSA », fondée sur cette remarque et initialement proposé par Rivest, Shamir et Adelman, est actuellement très répandue (cartes bancaires, transactions télématiques...), et elle est considérée comme étant extrêmement sûre.

Aussi peut-on imaginer sans peine l'impact qu'a eu un article publié en 1994 par un chercheur américain, Peter Shor, qui a affirmé qu'un ordinateur quantique pourrait factoriser le produit  $P$  de deux nombres premiers en un temps qui n'est plus que de l'ordre de  $(\log P)^3$ , c'est-à-dire qui est réduit d'un facteur exponentiel par rapport aux ordinateurs classiques! L'effervescence étant maintenant retombée, la situation semble être la suivante : l'algorithme proposé par Peter Shor est correct dans son principe, et apporte bien le gain d'efficacité escompté. En revanche, dans l'état actuel de nos connaissances, la réalisation d'un ordinateur quantique compétitif semble encore hors de portée de la technologie actuelle, bien qu'elle ne soit pas exclue par les lois de la physique. La mise en œuvre d'un ordinateur quantique apparaît donc comme un enjeu scientifique à long terme (quelques dizaines d'années), et non comme une menace immédiate sur les systèmes de cryptographie algorithmique.

### Principe du fonctionnement d'un ordinateur quantique

Nous ne tenterons pas ici d'explicitier l'algorithme de Shor (pour plus de détails voir appendice 2), mais de donner quelques idées intuitives sur la façon dont un ordinateur quantique peut effectuer un calcul. Le principe de base est que le calcul doit pouvoir se ramener à une évolution quantique de l'état initial, suivie d'une « mesure » qui détermine l'état du q-registre, mais interrompt aussi son évolution. Conformément aux principes de la mécanique quantique, la valeur trouvée sera associée à un des états propres de l'observable mesurée, qui correspond ici à un état classique du registre, c'est-à-dire à un mot binaire. En revanche, l'évolution de l'ordinateur lors du calcul lui-même aura mis en jeu en parallèle les  $2^N$  états correspondant à tous les nombres que le registre peut contenir.

Pour pouvoir effectuer des opérations successives, on devra faire évoluer le système de qubits de manière contrôlée, sous l'action d'une horloge qui détermine le rythme du calcul. À première vue, la détermination de cette évolution semble être un problème inextricable si l'on souhaite effectuer un calcul non trivial. En fait, on montre que cette construction peut être menée à bien assez simplement, car un calcul réel peut être décomposé en une

---

<sup>8</sup>Plus précisément, le meilleur algorithme de factorisation connu actuellement a un comportement en  $e^{1,9(\log P)^{1/3}(\log \log P)^{2/3}}$ .

succession d'opérations simples n'affectant qu'un ou deux qubits. Comme en informatique classique, ces opérations simples sont effectuées par des « portes logiques », dont des exemples bien connus classiquement sont les portes *NON*, *ET*, *OU*, ... En revanche, les portes quantiques requises par l'algorithme de Shor présentent certaines particularités :

(1) elles doivent être « réversibles », pour être compatibles avec l'évolution quantique du q-registre ;

(2) elles doivent manipuler des qubits, sur lesquels on peut effectuer certaines opérations logiques inconcevables classiquement.

Des exemples simples de portes quantiques sont les portes  $\sqrt{I}$  et  $\sqrt{NON}$ , qu'il faut appliquer deux fois pour obtenir l'identité (porte  $\sqrt{I}$ ), ou pour inverser 0 et 1 (porte *NON*). Ces deux portes préparent en fait un qubit dans une superposition linéaire à poids égaux des valeurs 0 et 1. Par exemple, deux applications de la porte  $\sqrt{I}$ , appelée aussi porte de Hadamard, effectuent les transformations successives :

$$\begin{aligned} |0\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \longrightarrow \frac{1}{2}((|0\rangle + |1\rangle) + (|0\rangle - |1\rangle)) = |0\rangle \\ |1\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \longrightarrow \frac{1}{2}((|0\rangle + |1\rangle) - (|0\rangle - |1\rangle)) = |1\rangle \end{aligned} \quad (2.10)$$

Une autre porte très importante est la porte « *C - NOT* » (controlled-not), qui est une porte à deux qubits effectuant l'opération suivante :

$$\begin{aligned} |0\rangle |0\rangle &\longrightarrow |0\rangle |0\rangle \\ |0\rangle |1\rangle &\longrightarrow |0\rangle |1\rangle \\ |1\rangle |0\rangle &\longrightarrow |1\rangle |1\rangle \\ |1\rangle |1\rangle &\longrightarrow |1\rangle |0\rangle \end{aligned}$$

On voit que cette porte conserve le premier qubit inchangé, et écrit un « *OU* exclusif » sur le second qubit. On peut aussi dire que le second qubit est basculé (porte *NON*) si le premier est dans l'état  $|1\rangle$ , on réalise donc bien un « *NON* contrôlé » (controlled-not). Examinons alors l'effet de l'application d'une porte de Hadamard sur le premier qubit, puis d'une porte *C - NOT* :

$$\begin{aligned} |0\rangle |0\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |0\rangle + |1\rangle |0\rangle) \longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |0\rangle + |1\rangle |1\rangle) \\ |0\rangle |1\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |1\rangle + |1\rangle |1\rangle) \longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |1\rangle + |1\rangle |0\rangle) \\ |1\rangle |0\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |0\rangle - |1\rangle |0\rangle) \longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |0\rangle - |1\rangle |1\rangle) \\ |1\rangle |1\rangle &\longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |1\rangle - |1\rangle |1\rangle) \longrightarrow \frac{1}{\sqrt{2}}(|0\rangle |1\rangle - |1\rangle |0\rangle) \end{aligned}$$

Les deux qubits, qui étaient initialement dans 4 états possibles tous factorisés, se retrouvent dans 4 états tous intriqués ! Ces états sont d'ailleurs appelés

« états de Bell », et forment une base « maximalement intriquée » de l'espace des états de deux qubits. Il est aussi intéressant de remarquer que l'opération inverse de la précédente, c'est-à-dire l'application de la porte CNOT puis de la porte de Hadamard, « désintrique » les deux qubits et permet ainsi d'identifier les 4 états de Bell. Cette opération inverse s'appelle donc une « mesure de Bell ». Un exemple simple et spectaculaire de l'utilisation de ces outils est la « téléportation quantique » d'un qubit, qui est détaillée dans l'appendice 3.

Comme nous l'avons déjà vu, il est possible de décomposer un calcul arbitraire en une suite d'applications de portes à un qubit et à 2 qubits. On pourrait penser qu'il suffit ensuite de faire évoluer l'ordinateur vers un état à une seule composante, qui serait le résultat du calcul. En fait, très peu d'algorithmes se prêtent à une manipulation aussi simple. L'état final de l'ordinateur est en général encore une superposition linéaire, et le résultat obtenu est donc aléatoire. Par exemple, si l'on considère l'algorithme de Shor, le résultat obtenu doit plutôt être considéré comme un « indice » permettant la factorisation ; il est facile de vérifier par un moyen conventionnel si la réponse est la bonne, et sinon de relancer le calcul. Peter Shor a montré que cette procédure d'essai-erreur fournit la bonne réponse avec une probabilité arbitrairement proche de un, en effectuant un nombre d'essais qui croît linéairement – et non plus exponentiellement – avec le nombre de chiffres du nombre à factoriser (voir détails dans l'appendice 2).

## La décohérence

Le principe d'un ordinateur quantique est donc compatible avec les lois de la physique, et un tel ordinateur semble réalisable, du moins tant que l'on considère des calculs simples, ne faisant intervenir qu'un petit nombre de portes. Lorsqu'on atteint des tailles de calcul importantes, l'état global de l'ordinateur se présente comme une superposition linéaire d'un très grand nombre d'états, dont l'évolution doit être contrôlée tout en préservant toutes les propriétés de la superposition linéaire. Il n'est pas clair à l'heure actuelle que ce type de système soit réalisable en pratique, et les recherches portent essentiellement dans deux directions :

(1) d'une part, il faut que le q-registre en évolution soit extraordinairement bien isolé de l'environnement extérieur. Tout couplage avec cet environnement va en effet induire un effet de « décohérence », susceptible de brouiller les superpositions quantiques. Cette exigence n'est compatible avec le contrôle de l'évolution du calcul que pour des systèmes très bien maîtrisés ;

(2) d'autre part, il faut prévoir, pour lutter contre les perturbations résiduelles, des « codes de correction » capables de remettre l'ordinateur dans l'état qui était le sien avant l'action de la perturbation extérieure. Il a été montré théoriquement que de tels codes correcteurs quantiques sont réalisables en principe, et pourraient assurer un fonctionnement de l'ordinateur exempt de toute erreur, à condition que le taux d'erreur par opération soit assez faible, et que le taux de redondance du calcul soit assez élevé.

Ces deux voies – choix d’un système présentant une décohérence très faible et codes de correction quantiques – sont très activement étudiées, et les questions soulevées ont stimulé aussi bien l’algorithmique que la physique quantique expérimentale. Il est actuellement très difficile de prévoir l’issue de ces recherches, mais il est tout à fait envisageable que des algorithmes quantiques simples (voir l’appendice 3) trouvent à moyen terme des applications dans les systèmes de cryptographie quantique.

### Exemples de mises en œuvre et de manipulations de qubits

Malgré les conditions restrictives dues à la nécessité de réduire les effets de décohérence, le nombre de candidats au titre de qubit est actuellement très grand. Les photons polarisés (dans le domaine optique) sont de « bons » qubits, qui se propagent et se manipulent facilement – c’est en fait la base du succès de la cryptographie quantique. En revanche, ils se prêtent mal à la réalisation de portes logiques – pour y parvenir, il faut coupler entre deux photons, ce qui requiert d’utiliser les concepts et les méthodes de l’électrodynamique quantique en cavité. Un groupe du Laboratoire Kastler-Brossel à l’École Normale Supérieure a ainsi réalisé des portes logiques faisant intervenir des photons micro-onde (de fréquence 51 GHz), couplés à des atomes de Rydberg dans des cavités supraconductrices.

D’autres qubits très étudiés sont les atomes et les ions piégés : l’information est alors codée sur des états internes, par exemple sur le spin des noyaux atomiques. Les ions individuels confinés et refroidis dans des pièges radio-fréquence sont en fait les dispositifs plus avancés actuellement, et ont donné lieu en 2003–2004 à la réalisation de portes quantiques à deux qubits très bien maîtrisées, ainsi qu’à des expériences de téléportation mettant en jeu trois qubits (NIST à Boulder et Université d’Innsbruck (Fig. 2.16), voir l’appendice 3).

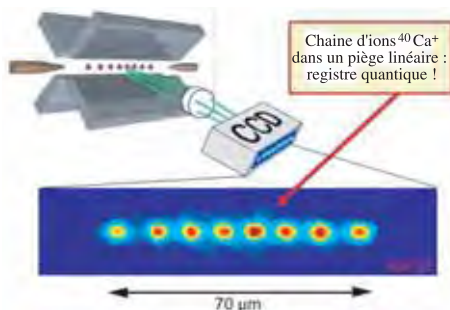


FIG. 2.16 – Exemple de mise en œuvre de qubits codés sur les états quantiques d’ions piégés (Université d’Innsbruck, Autriche).

D’autres systèmes, étudiés en particulier en France, sont les jonctions supraconductrices (groupe « Qnantronique », CEA Saclay), les atomes froids

(Institut d'Optique à Orsay), les « boîtes quantiques » (Institut d'Électronique Fondamentale à Orsay, Laboratoire de Photonique et Nanostructures à Marcoussis). Cette liste n'est pas exhaustive, certaines études en cours font intervenir des fullerènes ou des nanotubes de carbone, des électrons flottant sur l'hélium liquide, etc.

Une tendance forte de ces recherches est que l'on souhaite créer une « ingénierie des objets quantiques », qui n'existe pas spontanément dans la nature, bien qu'elle soit permise par les lois physiques – ceci est sans doute une bonne définition d'une technologie. En pratique, les idées qui semblent les plus prometteuses pour réaliser cette « ingénierie quantique » sont clairement reliées à la manipulation d'objets quantiques individuels (photons, atomes, ions, spins, quantum dots...), ou à des nanocircuits quantiques (jonctions Josephson). À terme il devrait devenir possible d'assembler des tels objets à grande échelle, mais cette étape n'a pas encore de réalité pratique, bien que plusieurs approches aient été proposées. Par exemple, un effort très important de nanofabrication orientée vers l'information quantique a été lancé par des équipes australiennes et américaines, avec l'objectif de détecter et de contrôler des ions implantés individuellement dans une matrice de silicium.

### 2.3.5 Conclusion

Il est très difficile à l'heure actuelle de prédire si l'ordinateur quantique calculera un jour, et si oui s'il ressemblera à ce que l'on peut imaginer aujourd'hui, ou fera appel à des technologies auxquelles personne n'a encore songé. Il semble néanmoins établi que l'intrication est un concept extraordinaire, donnant accès à des idées physiques et à des perspectives d'applications radicalement nouvelles. Il est clair aujourd'hui qu'Einstein a eu l'intuition de cette révolution conceptuelle des dizaines d'années avant que la communauté scientifique en prenne même conscience. Cette capacité à reconnaître l'incompréhensible et à identifier les frontières de la connaissance est certainement la marque des plus grandes figures de l'histoire des sciences.

## 2.4 Appendices

### 2.4.1 Le théorème de non-clonage

Le clonage d'un état inconnu est (heureusement pour Alice et Bob) impossible en mécanique quantique. Pour démontrer ce résultat, notons  $|\alpha_1\rangle$  un état quantique original à dupliquer. Le système sur lequel la copie doit « s'imprimer » est initialement dans un état connu que nous notons  $|\phi\rangle$  (l'équivalent d'une feuille blanche dans une photocopieuse). L'évolution du système total *original + copie* durant l'opération de clonage doit donc être :

$$\text{Clonage : } |\text{original} : \alpha_1\rangle |\text{copie} : \phi\rangle \longrightarrow |\text{original} : \alpha_1\rangle |\text{copie} : \alpha_1\rangle \quad (2.11)$$

Cette évolution est régie par un hamiltonien que nous ne chercherons pas à préciser, mais qui ne dépend pas de  $|\alpha_1\rangle$  puisque celui-ci est par hypothèse inconnu. Pour un autre état de l'original  $|\alpha_2\rangle$  (orthogonal à  $|\alpha_1\rangle$ ), on doit également avoir :

$$\text{Clonage : } |\text{original : } \alpha_2\rangle|\text{copie : } \phi\rangle \longrightarrow |\text{original : } \alpha_2\rangle|\text{copie : } \alpha_2\rangle \quad (2.12)$$

L'impossibilité du clonage apparaît alors pour l'état initial

$$|\alpha_3\rangle = \frac{1}{\sqrt{2}} (|\alpha_1\rangle + |\alpha_2\rangle) \quad (2.13)$$

Si l'opération de copie fonctionne pour cet état, on doit trouver :

$$\text{Clonage : } |\text{original : } \alpha_3\rangle|\text{copie : } \phi\rangle \longrightarrow |\text{original : } \alpha_3\rangle|\text{copie : } \alpha_3\rangle \quad (2.14)$$

Mais la linéarité de l'équation de Schrödinger impose par combinaison linéaire de (2.11) et (2.12) :

$$\begin{aligned} |\text{original : } \alpha_3\rangle|\text{copie : } \phi\rangle \longrightarrow \\ (1/\sqrt{2}) (|\text{original : } \alpha_1\rangle|\text{copie : } \alpha_1\rangle + |\text{original : } \alpha_2\rangle|\text{copie : } \alpha_2\rangle) \end{aligned}$$

Cet état final intriqué est différent de l'état souhaité (2.14).

L'examen de cette démonstration permet de comprendre l'apport de la mécanique quantique à la cryptographie. Si on se limite à une transmission à deux états orthogonaux  $|\alpha_1\rangle = |x\rangle$  et  $|\alpha_2\rangle = |y\rangle$ , alors l'espion peut rester indétectable comme nous l'avons vu dans les paragraphes précédents. Les deux opérations (2.11) et (2.12) sont possibles, ne serait-ce qu'en mesurant la polarisation du photon dans la base hv, puis en le réémettant dans le même état. C'est le fait de pouvoir utiliser simultanément les états  $|\alpha_1\rangle$ ,  $|\alpha_2\rangle$  et des combinaisons linéaires de ces états  $|\alpha_3\rangle = |x'\rangle$  ou  $|\alpha_3\rangle = |y'\rangle$  qui fait l'originalité de la cryptographie quantique, et qui interdit toute duplication fiable d'un message intercepté par un espion.

## 2.4.2 L'algorithme de Shor

Les calculs s'effectuent au moyen d'algorithmes, que les mathématiciens savent classer en terme de difficulté : c'est l'objet de la théorie de la complexité. Ce classement conduit à penser que la difficulté d'un problème est une propriété du problème, et non de la machine qui effectue le calcul. C'est effectivement vrai tant que la physique qui régit le calculateur reste inchangée : la factorisation est par exemple un problème « difficile » pour tous les ordinateurs classiques. Cela signifie que le temps nécessaire pour faire le calcul augmente très vite (exponentiellement) lorsqu'on cherche à factoriser des nombres de plus en plus grands. Mais Peter Shor a montré en 1994 que ce problème devient « facile » sur un ordinateur quantique : le fait de changer la physique du calculateur permet donc de concevoir un nouvel algorithme, qui dans ce cas précis est exponentiellement plus rapide.

La méthode utilisée dans l'algorithme de Shor est subtile. L'algorithme naïf qui consiste à effectuer des divisions est dans tous les cas très peu efficace, et il existe des algorithmes classiques bien meilleurs, qui permettent par exemple de factoriser un nombre de 155 chiffres en quelques mois de calcul distribué. L'algorithme de Shor utilise encore une autre méthode, qui s'appuie sur la théorie des nombres. Un théorème dit en effet que pour factoriser un nombre  $P$ , on peut construire une fonction simple  $F$  de  $P$  et de la variable entière  $n$ , qui est périodique en  $n$ . Plus précisément, on a  $F(P, n) = a^n$  modulo  $P$ , où  $a$  est un entier premier avec  $P$ . Si on connaît la période de  $F$ , notée  $R$ , on obtient facilement les facteurs de  $P$ , qui sont donnés par  $\text{pgcd}(P, a^{R/2} + 1)$  et  $\text{pgcd}(P, a^{R/2} - 1)$ . Ce raisonnement est vrai classiquement, mais pour trouver  $R$  il faut évaluer la fonction  $F$  un très grand nombre de fois, ce qui rend le calcul peu efficace et donc « difficile ».

Au contraire, l'ordinateur quantique évalue en parallèle toutes les valeurs prises par  $F$ , pour toutes les valeurs de  $n$  que peuvent contenir ses registres. On effectue ensuite plusieurs opérations (mesure projective, transformée de Fourier...), puis on obtient un nombre aléatoire, « résultat » du calcul. On répète ensuite ce calcul, et on obtient une série de nombres, qui en fait ne sont pas complètement aléatoires, mais répartis régulièrement. En analysant ces régularités on peut remonter à la période  $R$ , et donc factoriser. On voit donc que le calcul quantique fonctionne ici sur un mode très particulier : il ne donne pas vraiment un « résultat » au sens habituel, mais plutôt un indice qui permet d'obtenir (exponentiellement) plus vite le résultat cherché. Ceci illustre le fait que l'ordinateur quantique, s'il est un jour mis en œuvre, n'aura pas pour but de remplacer les calculettes de poche, mais bien plutôt de traiter l'information suivant des concepts entièrement nouveaux.

### 2.4.3 La téléportation quantique

Comme nous l'avons vu plusieurs fois, il est impossible de cloner ou de déterminer parfaitement l'état d'un qubit unique si on n'a aucune information sur cet état. Mais est-il possible de « transférer à distance » cet état inconnu ? Curieusement, la réponse à cette question est positive [19], et l'opération correspondante est connue sous le vocable très médiatique de « téléportation quantique ». Ces termes ne sont pas totalement usurpés, puisque la téléportation requiert bien la destruction du système original, et permet de reproduire parfaitement à distance un état qu'on ne connaît pas – et qu'on ne peut d'ailleurs pas connaître.

La téléportation de l'état d'un qubit A vers un qubit cible C peut être considérée comme un petit algorithme quantique, dont les étapes sont les suivantes :

- (1) Alice dispose initialement de 3 qubits : le qubit A, placé dans un état arbitraire  $|\Psi_A\rangle = \alpha|0_A\rangle + \beta|1_A\rangle$  qu'Alice ne connaît pas, le qubit cible C, et un qubit auxiliaire B. Les qubits B et C sont placés initialement dans l'état  $|0\rangle$  ;

- (2) Alice intrique les qubits B et C par la méthode exposée plus haut (porte de Hadamard et porte  $C - NOT$ ), puis elle conserve le qubit B, et envoie le qubit C à Bob ;
- (3) Alice effectue ensuite une mesure de Bell entre les qubits A et B (voir plus haut). Cette mesure « projette » la paire (A, B) dans l'un des 4 états de Bell, qui est le résultat de la mesure. L'état du qubit A est détruit dans cette opération ;
- (4) Alice transmet à Bob le résultat de sa mesure, c'est-à-dire lequel des 4 états de Bell elle a obtenu. Bob dispose donc du qubit C, et de deux bits d'information classique ( $m = 0, 1, 2$  ou  $3$ ). Bob applique alors à C une transformation qui dépend de la valeur de  $m$ , et reconstitue ainsi l'état du qubit A (voir calcul ci-dessous).

La suite d'opérations effectuées sur l'état  $|\Psi_{ABC}\rangle$  des trois qubits est donc (le calcul effectué à l'étape 2 est essentiel) :

$$\text{Étape 1 : } |\Psi_{ABC}\rangle_{initial} = (\alpha |0_A\rangle + \beta |1_A\rangle) |0_B\rangle |0_C\rangle$$

$$\begin{aligned} \text{Étape 2 : } & (\alpha |0_A\rangle + \beta |1_A\rangle) \frac{1}{\sqrt{2}}(|0_B0_C\rangle + |1_B1_C\rangle) = \\ & \frac{1}{\sqrt{2}}(\alpha|0_A0_B0_C\rangle + \alpha|0_A1_B1_C\rangle + \beta|1_A0_B0_C\rangle + \beta|1_A1_B1_C\rangle) = \\ & \frac{1}{2^{3/2}} (|0_A0_B\rangle + |1_A1_B\rangle) (\alpha |0_C\rangle + \beta |1_C\rangle) + \\ & \frac{1}{2^{3/2}} (|0_A0_B\rangle - |1_A1_B\rangle) (\alpha |0_C\rangle - \beta |1_C\rangle) + \\ & \frac{1}{2^{3/2}} (|0_A1_B\rangle + |1_A0_B\rangle) (\alpha |1_C\rangle + \beta |0_C\rangle) + \\ & \frac{1}{2^{3/2}} (|0_A1_B\rangle - |1_A0_B\rangle) (\alpha |1_C\rangle - \beta |0_C\rangle) \end{aligned}$$

Étape 3 : État mesuré par Alice	Action de Bob sur le qubit C
$\frac{1}{\sqrt{2}}( 0_A0_B\rangle +  1_A1_B\rangle)$	Rien
$\frac{1}{\sqrt{2}}( 0_A0_B\rangle -  1_A1_B\rangle)$	Porte $PI$
$\frac{1}{\sqrt{2}}( 0_A1_B\rangle +  1_A0_B\rangle)$	Porte $NON$
$\frac{1}{\sqrt{2}}( 0_A1_B\rangle -  1_A0_B\rangle)$	Portes $NON$ et $PI$

où la porte  $NON$  (basculement du bit) a déjà été définie, et la porte  $PI$  (basculement de phase) réalise l'opération  $|0\rangle \rightarrow |0\rangle, |1\rangle \rightarrow -|1\rangle$ .

En examinant l'expression de l'état donné à la fin de l'étape 2, on voit que dans tous les cas le qubit C se retrouve dans l'état initial  $|\Psi_A\rangle$  du qubit A. Il est très important de remarquer que ni Bob ni Alice n'ont accès à  $|\Psi_A\rangle$  : toutes les opérations sont effectuées « en aveugle » sur un qubit dont ils ne



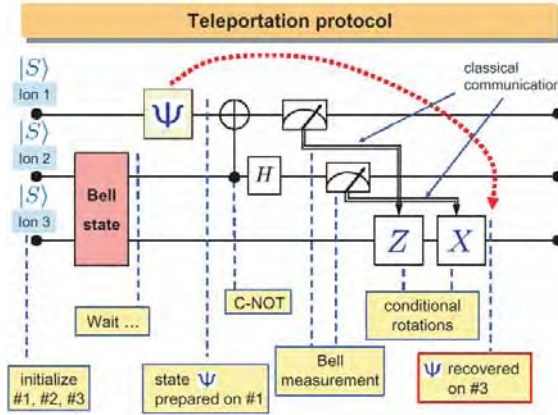


FIG. 2.17 – Protocole de téléportation quantique réalisé à l'Université d'Innsbruck en 2004. On reconnaît les différentes étapes de l'algorithme, qui sont mises en œuvre une à une en pilotant l'évolution de 3 qubits codés sur l'état quantique d'ions individuels, séparés de quelques micromètres dans un piège électromagnétique (Fig. 2.16).

connaissent pas l'état. Cette possibilité d'agir sur un qubit sans connaître son état joue un rôle fondamental dans le calcul quantique, et elle est en particulier à la base des codes quantiques de correction d'erreur.

La téléportation quantique a d'abord été réalisée avec des photons intriqués, puis avec des ions piégés [20, 21], simultanément par une équipe américaine et une équipe autrichienne. Ces expériences avec des ions piégés sont très spectaculaires, dans le sens où elles suivent pas à pas toutes les étapes de l'algorithme décrit ci-dessus (Fig. 2.17). On peut considérer que la mise en œuvre de ces « mini-algorithmes » représente l'état de l'art du calcul quantique en 2004.

## 2.5 Références

- [1] M. Jammer, *The Philosophy of Quantum Mechanics*, Wiley (1974), p. 303. Bien qu'il date d'une époque où l'on sous-estimait généralement la contribution de John Bell – et donc la portée des objections d'Einstein, cet ouvrage constitue une source intéressante d'informations et de références sur l'histoire de la mécanique quantique et sur les débats conceptuels associés.
- [2] A. Einstein, B. Podolsky, N. Rosen, *Can Quantum-Mechanical description of physical reality be considered complete?*, Phys. Rev. **47**, 777 (1935).
- [3] N. Bohr, *Can Quantum-Mechanical description of physical reality be considered complete?*, Phys. Rev. **48**, 696 (1935).
- [4] E. Schrödinger, Proc. Camb. Phil. Soc. **31**, 555–563 (1935).
- [5] D. Bohm, *Quantum Theory*, Prentice Hall (1951) ; Dover (1989).

- [6] J.S. Bell, *On the Einstein-Einstein-Podolsky-Rosen-Paradox*, *Physics* **1**, 195–200 (1964). Article reproduit dans *Speakable and unspeakable in quantum mechanics*, J.S. Bell, Cambridge University Press, 2<sup>e</sup> édition (2004).
- [7] J.S. Bell, *Atomic-cascade photons and quantum-mechanical nonlocality*, *Comments on Atomic and Molecular Physics* **9** (1980). Article reproduit dans *Speakable and unspeakable in quantum mechanics*, op. cit. dans [6].
- [8] A. Aspect, *Bell's theorem: the naive view of an experimentalist*, dans *Quantum [un]speakables, from Bell to Quantum information*, R.A. Bertlmann and A. Zeilinger (eds.), Springer (2002). Texte disponible sur la toile (<http://hal.ccsd.cnrs.fr/ccsd-00001079>). Cet article contient de nombreuses références, en particulier celles des expériences citées dans ce chapitre.
- [9] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, *Proposed experiment to test local hidden-variable theories*, *Phys. Rev. Lett.* **23**, 880 (1969).
- [10] R. García-Patron, J. Fiuracek, N.J. Cerf, J. Wenger, R. Tualle-Brouiri, Ph. Grangier, *Proposal for a Loophole-Free Bell Test Using Homodyne Detection*, *Phys. Rev. Lett.* **93**, 130409 (2004).
- [11] P.A. Schilp (ed.), *A. Einstein philosopher scientist*, Open court and Cambridge university press (1949).
- [12] *Einstein-Born, Correspondance*, Traduction française, Seuil (1972), p. 188.
- [13] P. Grangier, Contextual objectivity and quantum holism, e-print quant-ph/0301001, <http://fr.arxiv.org/abs/quant-ph/0301001>
- [14] R. Healey, *Holism and Nonseparability in Physics*, Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/physics-holism/>
- [15] W.K. Wothers, W.H. Zurek, *A Single Quantum Cannot be Cloned*, *Nature* **299**, 802–803 (1982).
- [16] A. Beveratos et al., *Single Photon Quantum Cryptography*, *Phys. Rev. Lett.* **89**, 187901 (2002).
- [17] E. Waks et al., *Secure communication: Quantum cryptography with a photon turnstile*, *Nature* **420**, 762 (2002).
- [18] N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, *Quantum cryptography*, *Rev. Mod. Phys.* **74**, 145 (2002).
- [19] C.H. Bennett et al., *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- [20] M. Riebe et al., *Deterministic quantum teleportation with atoms*, *Nature* **429**, 734–737 (2004).
- [21] M.D. Barrett et al., *Deterministic quantum teleportation of atomic qubits*, *Nature* **429**, 737–739 (2004).

# Chapitre 3

## La condensation de Bose-Einstein dans les gaz

Par *Claude Cohen-Tannoudji,*  
*Jean Dalibard et Franck Laloë*

### Sommaire

<b>3.1</b>	<b>Le « gaz parfait saturé »</b> . . . . .	<b>92</b>
3.1.1	Dénombrement des états dans un gaz . . . . .	92
3.1.2	La distribution de Bose-Einstein et le phénomène de condensation . . . . .	95
3.1.3	Interprétation ondulatoire du seuil de condensation . . . . .	98
<b>3.2</b>	<b>Évolution des idées après 1925</b> . . . . .	<b>99</b>
3.2.1	Des particules qui ne sont plus indépendantes ? . . . . .	99
3.2.2	Une vraie condensation ? La limite thermodynamique . . . . .	100
3.2.3	Un lien avec la superfluidité ? . . . . .	102
3.2.4	Le renouveau des gaz quantiques . . . . .	104
3.2.5	Un régime intermédiaire semi-quantique . . . . .	106
<b>3.3</b>	<b>La percée des alcalins</b> . . . . .	<b>107</b>
3.3.1	La manipulation d'atomes par laser . . . . .	108
3.3.2	Vers la condensation . . . . .	110
3.3.3	Interférences de condensats . . . . .	113
<b>3.4</b>	<b>Au-delà du gaz parfait</b> . . . . .	<b>114</b>
3.4.1	Description, manipulation et effets des interactions à l'équilibre . . . . .	114
3.4.2	La superfluidité . . . . .	117
3.4.3	Condensat en rotation et tourbillons quantiques . . . . .	119
3.4.4	Du superfluide vers l'isolant de Mott . . . . .	120
<b>3.5</b>	<b>Conclusion et perspectives</b> . . . . .	<b>122</b>
<b>3.6</b>	<b>Références générales</b> . . . . .	<b>125</b>



### **Claude Cohen-Tannoudji**

Claude Cohen Tannoudji est professeur honoraire au Collège de France, membre de l'Académie des sciences. Toute sa carrière de chercheur s'est déroulée à l'École Normale Supérieure. Il a écrit quatre ouvrages sur la mécanique quantique, l'électrodynamique quantique, l'optique quantique et les statistiques de Lévy, ainsi que de très nombreux articles théoriques et expérimentaux sur divers problèmes de physique atomique et d'optique quantique. Il est en particulier l'auteur du formalisme universellement utilisé de l'atome « habillé » par le champ électromagnétique. Il a partagé en 1997 le Prix Nobel de physique avec Steve Chu et William Phillips pour ses travaux sur le refroidissement et le piégeage d'atomes par des faisceaux laser.



### **Jean Dalibard**

Jean Dalibard est directeur de recherches au CNRS, chercheur au département de Physique de l'École Normale Supérieure (Laboratoire Kastler Brossel) et professeur à l'École Polytechnique. Il est membre de l'Académie des sciences. Ses travaux de recherche ont d'abord porté sur la manipulation et le refroidissement des atomes par laser. Il étudie désormais les gaz quantiques, en particulier les condensats de Bose-Einstein, en s'intéressant tout particulièrement à leurs propriétés de cohérence et de superfluidité.



### **Franck Laloë**

Franck Laloë est directeur de recherche au CNRS, chercheur au département de physique de l'École Normale Supérieure, au sein du Laboratoire Kastler Brossel dont il est le directeur actuel. Ses travaux ont porté sur le pompage optique et l'orientation nucléaire dans l'hélium trois, puis sur le transport dans les fluides quantiques polarisés, en particulier les ondes de spin. Il a aussi contribué à la compréhension des fondements de la mécanique quantique et s'est intéressé à l'acoustique musicale. Il travaille actuellement sur les propriétés statistiques des gaz à très basse température.



# La condensation de Bose-Einstein dans les gaz

L'INTERPRÉTATION DE LA DISTRIBUTION SPECTRALE DU RAYONNEMENT en équilibre thermodynamique avec les parois d'une enceinte dans laquelle il est enfermé<sup>1</sup> a stimulé de nombreux travaux à la fin du XIX<sup>e</sup> siècle. C'est pour en rendre compte que Max Planck introduisit en 1900, et pour la première fois en physique, l'idée de la quantification des échanges d'énergie entre oscillateurs matériels et rayonnement, ainsi que la constante  $h$  qui porte son nom. La formule de Planck, donnant la répartition spectrale du rayonnement du corps noir, peut ainsi être considérée comme marquant la naissance de la physique quantique. Plusieurs démonstrations de la formule de Planck ont été données. L'une d'entre elles fut trouvée en 1924 par le jeune physicien indien Satyendra Nath Bose qui traitait le rayonnement comme un gaz de particules identiques, les quanta lumineux. En cherchant la distribution la plus probable en énergie des particules d'un tel gaz, Bose retrouvait la loi de Planck. Albert Einstein, à qui Bose avait adressé son article, fut très intéressé par ce travail. Il traduisit l'article en allemand et aida Bose à le publier [1]. Il généralisa peu après cette approche à l'étude d'un gaz parfait de particules matérielles identiques, et dans deux articles parus en 1924 et 1925 [2], il établit l'expression de la distribution en énergie d'un gaz parfait monoatomique, connue maintenant sous le nom de distribution de Bose-Einstein<sup>2</sup>.

Les formules établies par Einstein dans son second article le conduisaient à prévoir un effet curieux. Si, à volume  $V$  et à nombre total  $N$  d'atomes fixés, on diminue la température  $T$  du gaz, on trouve qu'en dessous d'une certaine température critique  $T_C$  les atomes doivent s'accumuler dans l'état fondamental de la boîte contenant les atomes. Cet état acquiert ainsi une population macroscopique, qui peut devenir de l'ordre de  $N$  quand on continue à diminuer la température. Les valeurs prévues pour  $T_C$  étaient cependant très basses, et à ces températures, tout corps ne semblait pouvoir exister qu'à l'état liquide ou solide. C'est pourquoi la condensation prévue par Einstein pour un gaz

---

<sup>1</sup>Ce rayonnement est traditionnellement appelé rayonnement du corps noir.

<sup>2</sup>Pour plus de détails sur les échanges entre Einstein et Bose, voir [3].

parfait fut plutôt considérée comme un phénomène purement académique par les physiciens de l'époque, y compris Einstein lui-même<sup>3</sup>. Certes, des rapprochements furent suggérés plus tard par Fritz London entre la condensation prévue par Einstein et la superfluidité de l'hélium liquide ; nous y reviendrons plus en détail au paragraphe 3.2.3. Toutefois il fallut attendre jusqu'en 1995, soit 70 ans après l'article d'Einstein, pour que les progrès réalisés sur la manipulation d'échantillons de gaz alcalins ultrafroids et polarisés permettent d'observer la condensation de Bose-Einstein dans un gaz. Depuis cette date, on assiste à une véritable explosion de l'activité théorique et expérimentale dans ce domaine. Le but de ce chapitre est de décrire brièvement la genèse et le développement de ce domaine de recherche qui est devenu l'un des plus actifs de la physique.

Nous commençons (premier paragraphe) par rappeler les principaux résultats établis par Einstein : importance de l'indiscernabilité des particules dans le dénombrement des états microscopiques du gaz et dans l'établissement de la statistique de Bose-Einstein ; phénomène de condensation conduisant Einstein à considérer la partie non condensée du gaz comme un « gaz parfait saturé » ; interprétation ondulatoire du seuil de condensation. Nous décrivons ensuite (paragraphe 2) les débats qui ont été stimulés par l'article d'Einstein, les nouvelles idées qui ont été suggérées, le renouveau d'intérêt qui s'est manifesté pour l'étude des effets quantiques dans les gaz. Les premières observations de la condensation de Bose-Einstein sur les gaz ultrafroids d'atomes alcalins obtenus par refroidissement laser et refroidissement par évaporation sont présentées (paragraphe 3) ainsi que quelques expériences démontrant la cohérence des ondes de matière ainsi obtenues. L'importance des interactions entre atomes est abordée au paragraphe 4. On montre comment on peut les décrire et les manipuler. Le rôle qu'elles jouent dans divers effets physiques, comme la superfluidité, les tourbillons quantiques et la transition superfluide-isolant de Mott est également analysé. Une conclusion (paragraphe 5) mentionne enfin quelques développements récents sur d'autres gaz quantiques dégénérés formés d'atomes fermioniques.

## 3.1 Le « gaz parfait saturé »

### 3.1.1 Dénombrement des états dans un gaz

L'élément essentiel dans l'établissement de la distribution en énergie d'un gaz parfait monoatomique mit un certain temps à être formulé de manière explicite<sup>4</sup>. Il concerne la manière de dénombrer le nombre d'états microscopiques distincts associés à un état macroscopique du gaz, défini par Einstein comme la donnée des nombres  $N_i$  d'atomes dans chaque domaine  $\Delta_i$  d'énergie comprise entre  $E_i$  et  $E_i + \Delta E_i$ . Chacun de ces domaines peut être divisé

<sup>3</sup>Il écrivait à ce propos : « l'idée est belle, mais contient-elle une part de vérité ? » [4].

<sup>4</sup>Voir par exemple l'article d'Alfred Kastler sur ce sujet [5].



en  $g_i$  cellules différant par d'autres grandeurs physiques que l'énergie. Une cellule correspond tout simplement en mécanique quantique à un état individuel d'une particule ; elle occupe dans l'espace des phases un volume  $h^3$  (où  $h$  est la constante de Planck). Chaque état macroscopique est alors défini par une certaine répartition des  $N_i$  particules parmi les  $g_i$  cellules, pour tous les domaines  $\Delta_i$ . L'état d'équilibre est l'état macroscopique qui est réalisé par le plus grand nombre d'états microscopiques, compte tenu des contraintes imposées au système (on fixe le nombre de particules et l'énergie totale).

En mécanique statistique classique, des particules identiques sont considérées comme discernables. On peut numéroter à chacune des particules à l'instant initial et suivre individuellement au cours de l'évolution temporelle chacune de ces particules qui conserve ainsi le numéro qui lui a été attribué. Deux états microscopiques se déduisant l'un de l'autre par une permutation de 2 particules identiques se trouvant dans 2 cellules différentes doivent donc être considérés comme distincts dans le dénombrement des états microscopiques associés à un état macroscopique. On peut montrer que l'état macroscopique le plus probable, c'est-à-dire l'état d'équilibre, suit alors la distribution de Maxwell-Boltzmann.

Les calculs faits par Bose pour le rayonnement, puis par Einstein pour un gaz parfait monoatomique, correspondent à un autre type de dénombrement où des particules identiques sont considérées comme indiscernables : deux états microscopiques se déduisant l'un de l'autre par une permutation quelconque de particules identiques doivent être considérés comme un même état. Le dénombrement des états microscopiques correspondant à un état macroscopique est alors différent du précédent et c'est ce qui explique pourquoi la distribution d'équilibre ainsi obtenue, la distribution de Bose-Einstein, est différente de la statistique de Maxwell-Boltzmann. Dans le cas étudié par Bose, les « particules » du gaz sont des quanta d'excitation d'un mode du rayonnement, et il est assez naturel de les considérer comme indiscernables. L'extension d'un tel mode de dénombrement à un gaz d'atomes était moins évidente ; l'idée d'Einstein peut être considérée comme prémonitoire des statistiques quantiques résultant du postulat de symétrisation des fonctions d'onde.

Pour illustrer les considérations précédentes sur le dénombrement des états microscopiques de particules identiques discernables ou indiscernables, considérons l'exemple simple de trois particules réparties dans deux cellules (Fig. 3.1). Si les particules sont discernables, et donc numérotables (partie gauche de la figure), il y a un seul état où les 3 particules sont dans la cellule de gauche, un seul état où ces 3 particules sont dans la cellule de droite. Le résultat est le même quand les particules sont indiscernables (partie de droite de la figure). En revanche, si l'on met une particule dans la cellule de gauche et 2 dans la cellule de droite, 3 états distincts apparaissent quand les 3 particules sont discernables alors qu'un seul état existe lorsqu'elles sont indiscernables. Un résultat analogue est obtenu quand on met 2 particules dans la cellule de gauche et 1 particule dans la cellule de droite. Le nombre total

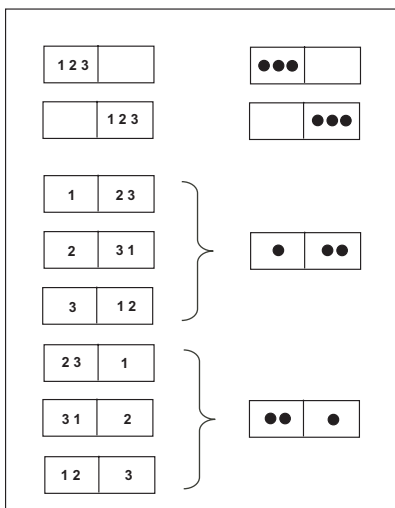


FIG. 3.1 – États microscopiques de 3 particules identiques réparties dans 2 cellules. La partie de gauche de la figure correspond au cas où les 3 particules sont discernables et repérées par un numéro ; la partie de droite au cas où les particules sont indiscernables.

d'états microscopiques est donc clairement différent dans les deux cas, 8 pour des particules discernables, 4 pour des particules indiscernables.

Une autre idée importante est également suggérée par cette figure. Les états microscopiques où toutes les particules sont dans la même cellule sont en même nombre dans les deux cas. En revanche, dès que les particules sont réparties dans des cellules différentes, les permutations de particules discernables donnent naissance à un nombre d'états microscopiques plus élevé que pour des particules indiscernables. Le poids relatif des états où les particules sont toutes dans la même cellule par rapport aux autres états où elles sont réparties dans des cellules différentes est donc plus important quand les particules sont indiscernables. Pour le cas de la figure 3.1, ce poids relatif vaut  $2/8 = 1/4$  pour des particules discernables et  $2/4 = 1/2$ , soit 2 fois plus grand, pour des particules indiscernables. On peut ainsi comprendre pourquoi la statistique de Bose-Einstein va favoriser l'accumulation des particules dans une même cellule par rapport à ce qui se passe pour la statistique de Maxwell-Boltzmann.

Notons enfin que, dans tout ce qui précède, chaque cellule est supposée pouvoir contenir un nombre arbitraire de particules identiques, qu'elles soient discernables ou indiscernables. Deux années après la publication des articles de Bose et d'Einstein, paraissaient les travaux de Fermi et de Dirac introduisant l'idée qu'il pouvait exister des particules, obéissant au principe de Pauli, pour lesquelles les nombres d'occupation de chaque cellule ne pouvaient être égaux

qu'à 0 ou à 1. Le dénombrement des états microscopiques de telles particules, considérées elles aussi comme indiscernables, conduit à la statistique de Fermi-Dirac.

### 3.1.2 La distribution de Bose-Einstein et le phénomène de condensation

Considérons un ensemble d'atomes identiques, sans interactions mutuelles, en contact avec un thermostat. Il s'agit de trouver l'état macroscopique le plus probable du gaz pour des valeurs fixées du nombre total  $N$  d'atomes et de l'énergie totale moyenne  $E$  de ces atomes

$$\sum_i N_i = N \qquad \sum_i N_i E_i = E \qquad (3.1)$$

Nous ne détaillerons pas ici le calcul, fait la première fois par Einstein, du nombre d'états microscopiques distincts correspondant à un état macroscopique défini par l'ensemble  $\{N_i\}$  des nombres  $N_i$  d'atomes dans les domaines  $\Delta_i$ , calcul fait en supposant les atomes indiscernables. La recherche du maximum de ce nombre, compte tenu des contraintes (3.1) donne pour le nombre moyen d'occupation  $n_i = N_i/g_i$  d'une cellule ou état quantique  $i$  d'énergie  $E_i$  le résultat :

$$n_i = \frac{1}{\exp[\beta(E_i - \mu)] - 1} \qquad (3.2)$$

qui est la distribution de Bose-Einstein. Dans cette expression,  $\beta$  et  $\mu$  sont des constantes (« multiplicateurs de Lagrange ») provenant des contraintes (3.1). La première constante,  $\beta$ , est reliée à la température  $T$  du thermostat par l'équation  $\beta = 1/k_B T$ , où  $k_B$  est la constante de Boltzmann ; la seconde,  $\mu$ , est le potentiel chimique, c'est-à-dire l'énergie à fournir pour ajouter un atome au gaz. Il est commode d'utiliser, au lieu de  $\mu$ , la fugacité  $z$  définie par  $z = \exp(\beta\mu)$ , qui est donc un nombre réel et positif, de sorte que (3.2) peut encore s'écrire :

$$n_i = \frac{1}{z^{-1} \exp(\beta E_i) - 1} = \frac{z}{\exp(\beta E_i) - z} \qquad (3.3)$$

On peut toujours prendre pour zéro d'énergie l'énergie de l'état fondamental  $i = 0 : E_0 = 0$ . On a alors  $\exp(\beta E_0) = 1$ . Comme  $N_0$  ne peut pas être négatif,  $z$  est donc nécessairement plus petit que 1, ce qui donne pour les valeurs possibles de  $z$  le domaine

$$0 \leq z < 1 \qquad (3.4)$$

En reportant (3.3) dans la première équation (3.1), on obtient

$$N = \sum_i N_i = \sum_i g_i n_i = \sum_i g_i \frac{z}{\exp(\beta E_i) - z} \qquad (3.5)$$

équation qu'il faut ensuite inverser pour obtenir  $z$  en fonction de  $N$  et de  $T$ .

Considérons un gaz d'atomes contenu dans une boîte cubique de côté  $L$  et de volume  $V = L^3$ . Supposons qu'à température  $T$  et volume  $V$  fixés, on augmente progressivement le nombre d'atomes  $N$  contenus dans la boîte. Aux faibles valeurs de  $N$ , la résolution de l'équation (3.5) donne une valeur très petite pour  $z$  :  $z \ll 1$ . Dans le dénominateur de l'équation (3.3), on peut alors négliger  $z$  devant l'exponentielle, qui est toujours supérieure à 1, ce qui donne :

$$n_i \simeq z \exp(-\beta E_i) \quad (3.6)$$

La distribution de Bose-Einstein est alors très peu différente de la distribution classique de Maxwell-Boltzmann. Quand le nombre  $N$  d'atomes devient suffisamment grand,  $z$  se rapproche de la limite supérieure  $z = 1$  de l'intervalle (3.4) et un comportement tout à fait nouveau apparaît. Pour le décrire, il est commode de séparer dans (3.5) la contribution  $N_0$  de l'état fondamental supposé non dégénéré ( $g_0 = 1$ ), contribution égale à  $z/(1-z)$  puisque  $E_0 = 0$ , de celle des autres états excités  $i \neq 0$  :

$$N = \frac{z}{1-z} + \sum_{i \neq 0} g_i \frac{z}{\exp(\beta E_i) - z} \quad (3.7)$$

Quand  $z$  se rapproche de 1,  $N_0 = z/(1-z)$  devient très grand. En revanche, la somme  $N_{\text{exc}}$  des populations des états excités  $i \neq 0$  demeure toujours inférieure à une valeur  $N_{\text{max}}$  obtenue en remplaçant chacun des termes de la somme apparaissant au second membre de (3.7) par sa valeur maximale atteinte pour  $z = 1$ .

$$N_{\text{exc}} = \sum_{i \neq 0} g_i \frac{z}{\exp(\beta E_i) - z} \leq N_{\text{max}} = \sum_{i \neq 0} g_i \frac{1}{\exp(\beta E_i) - 1} \quad (3.8)$$

Le point important, qui permet de comprendre le phénomène de condensation, est que la valeur de  $N_{\text{max}}$ , qui peut être calculée, soit numériquement, soit après remplacement de la somme discrète par une intégrale, est finie. À température  $T$  et volume  $V$  fixés, quand le nombre total  $N$  d'atomes devient supérieur à cette valeur  $N_{\text{max}}$ , il est impossible pour les populations des états excités d'accueillir l'ensemble des atomes. L'excédent de populations  $N - N_{\text{max}}$  se condense alors nécessairement dans l'état fondamental, puisque  $N_0 = z/(1-z)$  peut prendre une valeur aussi élevée que l'on veut pour  $z$  suffisamment proche de 1. Une fois le seuil  $N = N_{\text{max}}$  atteint, tout atome ajouté au gaz a une probabilité proche de 1 d'aller se condenser dans l'état fondamental<sup>5</sup>. Pour décrire le gaz d'atomes non condensés<sup>6</sup> incapable d'accueillir une population supplémentaire, Einstein parle de « gaz parfait saturé ».

<sup>5</sup>La croissance de  $N_0$  est régulière et commence avant que  $N$  atteigne la valeur « critique »  $N_{\text{max}}$  ; mais si l'on porte  $N_0/N$  en fonction de  $N/N_{\text{max}}$ , on obtient une courbe dont la pente passe discontinûment de 0 à 1 à la limite où  $V$  est très grand.

<sup>6</sup>Il est d'usage maintenant d'appeler « condensat » l'ensemble des atomes condensés dans l'état fondamental et « nuage thermique » l'ensemble des atomes non condensés.

La valeur de  $N_{\max}$  dépend du problème considéré, par exemple gaz homogène contenu dans une boîte de volume  $V$ , ou gaz piégé dans un potentiel harmonique. Nous ne détaillerons pas ici les calculs correspondants et nous contenterons de donner les résultats. Au lieu de faire varier  $N$  à température et volume fixés comme plus haut, nous faisons varier maintenant  $T$  à  $N$  et  $V$  fixés. Quand on diminue progressivement  $T$  à partir d'une valeur élevée, la valeur de  $N_{\max}$ , initialement très grande, diminue et devient égale à  $N$  pour une certaine température  $T_C$ . Quand  $T < T_C$ ,  $N_{\max}$  devient inférieur à  $N$ , les populations des états excités sont saturées et l'excédent de population  $N - N_{\max}$  se condense dans l'état fondamental. La température  $T_C$  introduite plus haut est donc la température critique en dessous de laquelle se produit la condensation de Bose-Einstein.

Pour un gaz parfait homogène de  $N$  atomes contenu dans une boîte de volume  $V$ , on trouve que la valeur  $T_C$  de la température critique est donnée par :

$$\frac{N}{V} = \zeta(3/2) \left[ \frac{m k_B T_C}{2\pi \hbar^2} \right]^{3/2} \quad (3.9)$$

où  $m$  est la masse des atomes,  $\hbar = h/2\pi$  et  $\zeta(3/2)$  est un nombre sans dimensions  $\zeta(3/2) \simeq 2,612$ . Nous donnerons plus loin une interprétation ondulatoire de cette équation. Notons que la valeur de  $T_C$  donnée par (3.9) demeure finie et inchangée quand on fait tendre  $N$  et  $V$  vers l'infini en gardant la densité  $N/V$  constante (limite thermodynamique), pour le problème à 3 dimensions considéré ici. Une étude analogue peut être faite pour un gaz à 2 ou 1 dimensions. On trouverait alors que  $T_C$  tend vers 0 à la limite thermodynamique.

Il est également possible de calculer la fraction  $N_0/N$  d'atomes condensés quand  $T$  décroît au dessous de  $T_C$ . On trouve<sup>7</sup> :

$$\frac{N_0}{N} = 1 - \left( \frac{T}{T_C} \right)^{3/2} \quad (3.10)$$

Des calculs analogues peuvent être faits pour un gaz parfait piégé dans un potentiel harmonique isotrope où les atomes oscillent avec une fréquence (angulaire)  $\omega$ . On trouve que la température critique  $T_C$  au dessous de laquelle  $N_{\max}$  devient inférieur à  $N$  est donnée par l'équation :

$$k_B T_C = \hbar \omega \left[ \frac{N}{\zeta(3)} \right]^{1/3} \quad (3.11)$$

---

<sup>7</sup>Ici aussi, un calcul plus précis montre que la croissance de  $N_0/N$  au voisinage de  $T/T_C = 1$  n'est pas aussi brutale que l'indique l'équation (3.10). Elle s'effectue sur un intervalle de valeurs de  $T/T_C$  d'autant plus étroit que  $N$  est plus grand. Une discontinuité réelle dans la vitesse de variation de  $N_0/N$  n'apparaît qu'à la limite thermodynamique. Voir plus loin la discussion du paragraphe 3.2.2.

où  $\zeta(3)$  est un nombre sans dimensions  $\zeta(3) \simeq 1,202$ . La fraction d'atomes condensés est donnée par une équation analogue à (3.10)

$$\frac{N_0}{N} = 1 - \left( \frac{T}{T_C} \right)^3 \quad (3.12)$$

Si les atomes obéissaient à la statistique de Maxwell-Boltzmann, l'occupation de l'état fondamental deviendrait importante quand la température  $T$  atteint une valeur telle que  $k_B T$  devient de l'ordre de  $\hbar\omega$ . Pour des particules obéissant à la statistique de Bose-Einstein, cette accumulation d'atomes dans l'état fondamental apparaît d'après (3.11) à des températures  $T_C$  augmentées par un facteur  $N^{1/3}$ , qui peut être grand puisque  $N \gg 1$ . La condensation de Bose-Einstein ne doit pas donc être confondue avec l'effet thermique trivial qui apparaît quand l'énergie thermique  $k_B T$  devient de l'ordre de la séparation  $\hbar\omega$  entre l'état fondamental du système et le premier état excité.

### 3.1.3 Interprétation ondulatoire du seuil de condensation

Revenons au cas d'un gaz parfait homogène contenu dans une boîte de volume  $V$  et à l'expression (3.9) de la température critique  $T_C$ . Il est intéressant de réécrire cette expression en y faisant apparaître la longueur d'onde thermique  $\lambda_T$  des atomes, qui caractérise leur comportement ondulatoire, plus précisément la longueur de cohérence des paquets d'ondes qui leur sont associés :

$$\lambda_T = \sqrt{\frac{2\pi\hbar^2}{mk_B T}} \quad (3.13)$$

Comme  $N/V$  n'est autre que la densité spatiale  $\rho$  uniforme du gaz, l'équation (3.9) donne à  $T = T_C$  :

$$\rho [\lambda_{T_C}]^3 = \zeta(3/2) \simeq 2,612 \quad (3.14)$$

qui, compte tenu de (3.13), permet de calculer la température critique pour toute densité  $\rho$ ; plus cette dernière est faible, plus la température critique est basse.

La densité spatiale  $\rho$  est égale à  $d^{-3}$ , où  $d$  est la distance moyenne entre atomes. L'équation (3.14) exprime donc que, pour  $T = T_C$ , la distance moyenne  $d$  entre atomes est de l'ordre de la longueur d'onde de de Broglie thermique  $\lambda_{T_C}$ . Analysons alors l'évolution des phénomènes quand on fait décroître  $T$  d'une valeur supérieure à  $T_C$  à une valeur inférieure. Il est clair d'après (3.13) que  $\lambda_T$  est une fonction décroissante de  $T$ . Quand  $T \gg T_C$ , on a donc  $\lambda_T \ll d$ . Les paquets d'ondes des divers atomes sont alors séparés par une distance très supérieure à leur longueur de cohérence. L'aspect ondulatoire des atomes est peu important et on peut les traiter comme des objets classiques, ce qui explique pourquoi la statistique de Maxwell-Boltzmann est

une bonne approximation de la statistique de Bose-Einstein aux températures élevées. Quand  $T$  décroît,  $\lambda_T$  croît. Les paquets d'ondes se retrouvent séparés par une distance moyenne qui devient de l'ordre de grandeur de leur longueur de cohérence quand  $T$  devient de l'ordre de  $T_C$ . Ils peuvent alors interférer et cette interférence devient de plus en plus importante quand  $T$  décroît au dessous de  $T_C$ .

Il est intéressant de noter que, au moment où il écrivait son second article introduisant le phénomène de condensation, Einstein venait de recevoir la thèse de Louis de Broglie que Paul Langevin lui avait fait parvenir. Einstein invoque dans cet article l'interférence entre les ondes de de Broglie associées aux divers atomes pour expliquer, non pas le phénomène de condensation, mais un terme apparaissant dans l'expression des fluctuations du gaz parfait, terme analogue à celui trouvé auparavant pour les fluctuations du rayonnement et qui est relié aux interférences entre ondes lumineuses.

## 3.2 Évolution des idées après 1925

La condensation de Bose-Einstein a eu un destin historique plutôt inhabituel dans l'histoire de la physique. Habituellement, c'est assez rapidement que les grandes percées théoriques sont reconnues par les contemporains, et donnent lieu à des vérifications expérimentales. Ici, non seulement l'idée d'Einstein a été accueillie avec des doutes sérieux, mais environ 70 ans se sont écoulés avant que des observations expérimentales vraiment indiscutables permettent de confirmer clairement l'analyse théorique initiale.

Il faut en effet réaliser à quel point l'article original d'Einstein de 1924 [2] est apparu tôt dans l'histoire de la physique quantique, en fait avant même que les lois de la mécanique quantique aient vraiment émergé. Par exemple l'équation de Schrödinger, qui peut être considérée comme le fondement de la physique quantique moderne, n'a été introduite qu'un an plus tard. On peut même souligner un contraste réel entre l'introduction de la condensation de Bose-Einstein et la découverte la plus connue d'Einstein, la relativité restreinte. Cette dernière s'appuyait sur de longs travaux d'approche menés par des physiciens aussi célèbres que Lorentz et Poincaré. Il semble parfaitement possible que, sans Einstein, d'autres auraient assez rapidement introduit des idées très semblables. Rien de tel dans le cas de l'introduction de la condensation de Bose-Einstein, totalement en avance sur son temps ; comme nous l'avons déjà mentionné, son idée a été accueillie avec un certain scepticisme.

### 3.2.1 Des particules qui ne sont plus indépendantes ?

Dès 1925, Schrödinger écrit à Einstein pour lui dire à quel point il est surpris par la distribution (3.3), qui ne correspond pas à la distribution de Boltzmann bien connue, et qui sous-entend que les particules d'un gaz parfait (sans interaction) ne sont plus « indépendantes ». La question est parfaitement

naturelle ; elle vient de ce que Schrödinger n'avait pas perçu la différence introduite implicitement par Einstein entre particules indiscernables, obéissant à la nouvelle statistique, et particules discernables, qui correspondent effectivement à la statistique de Boltzmann. Dans sa réponse, Einstein souligne que, dans la nouvelle statistique, les quanta ne doivent pas être traités séparément les uns des autres ; les « molécules » n'apparaissent donc pas comme localisées indépendamment, chacune ayant une certaine propension à se placer dans la même « cellule » qu'une autre. C'est probablement la première discussion historique des propriétés très importantes, mais peu intuitives, des statistiques quantiques.

En termes modernes, on n'emploierait plus exactement les mêmes termes que Schrödinger ; on dirait que les particules identiques de l'article d'Einstein sont bien indépendantes, puisqu'elles n'interagissent pas, mais qu'elles sont corrélées par leur statistique quantique. Comme mentionné au paragraphe 3.1, toute la physique de la transition de Bose-Einstein tient dans les nouvelles règles de comptage introduites par Bose, qui ne considère pas comme distincts deux états physiques où un état excité est occupé par deux particules numérotées différemment. On dirait donc maintenant que c'est le peu d'entropie associée à la présence d'excitations qui, dans un gaz de Bose-Einstein, permet l'apparition d'un condensat, contrairement à ce qui se produit pour des particules de Boltzmann.

### 3.2.2 Une vraie condensation ? La limite thermodynamique

Un ami proche d'Einstein, Paul Ehrenfest, est probablement celui qui éleva les objections les plus porteuses d'idées nouvelles. Einstein et Ehrenfest avaient l'habitude de se rencontrer lors de séjours à Leyde, aux Pays-Bas, traditionnellement un haut lieu de la physique (Huyghens, Lorentz et d'autres y vécurent). Il faut avouer que l'article d'Einstein laissait en sous-entendu bien des points délicats ! En fait, dans son article ne figure pas de somme discrète sur les valeurs de  $i$  comme celle de (3.5) ou (3.8), mais directement la valeur limite où  $V$  est très grand ; les sommes sur  $i \neq 0$  peuvent alors être remplacées par le produit de  $V$  par une intégrale finie, limitée supérieurement comme mentionné plus haut (paragraphe 3.1.2). De là, Einstein déduit l'existence d'une saturation de la population non condensée, et donc d'une accumulation dans l'état fondamental à partir d'une certaine densité ; en d'autres termes, il apparaît une singularité pour cette valeur de la densité (transition de phase). Or Ehrenfest faisait remarquer que, pour tout système physique donné, donc de taille fixe, il suffit de revenir tout simplement à la somme discrète initiale pour faire disparaître cet effet : tout est alors continu, régulier, aucune singularité ne se produisant quand la densité ou la température varient. Par exemple, la figure 3.2 représente le rapport entre le nombre de particules condensées  $N_0$  et le nombre total de particules  $N$  en fonction de la température  $T$ , pour des systèmes contenant  $10^2$ ,  $10^3$  ou  $10^4$  particules ; aucune de ces courbes ne



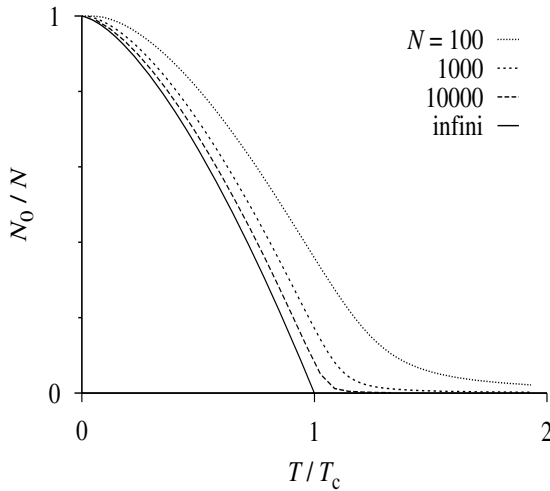


FIG. 3.2 – Variations du rapport entre le nombre de particules condensées  $N_0$  et le nombre total de particules  $N$  en fonction de la température (en unités de la température critique prévue par l'équation (3.14)); la courbe supérieure correspond à  $N = 100$ , celle d'en dessous à  $N = 10^3$  et la suivante à  $N = 10^4$  particules. La courbe la plus basse correspond à la limite thermodynamique où  $N$  est infini, et présente un point anguleux caractéristique d'une transition de phase (la condensation de Bose-Einstein). Cette figure a été réalisée par Werner Krauth (LPS, ENS).

présente de point anguleux ou de discontinuité. Où était alors la transition dont Einstein prédisait l'existence? N'était-ce pas un pur artéfact d'une approximation non contrôlée, le remplacement d'une somme discrète par une intégrale?

Les réactions d'Einstein à cette objection ne sont pas connues, au moins des auteurs de ce texte. Dans l'article original, le passage de la somme à une intégrale est fait sans le moindre commentaire. Un homme aussi épris de logique pure s'est certainement construit une opinion sur le sujet, et l'a probablement exprimée oralement à son ami. Ce qui est clair est que cette période de l'histoire de la physique coïncide avec l'émergence d'une notion très importante, celle de limite thermodynamique, qui permet de répondre à l'objection d'Ehrenfest. C'est précisément dans le grand amphithéâtre de physique de Leyde qu'a eu lieu une discussion célèbre sur la question suivante : les transitions de phase sont-elles contenues dans le formalisme de la mécanique statistique, ou sont-elles intrinsèquement au-delà de ce formalisme, nécessitant l'adjonction de nouveaux postulats? On sait que les transitions de phase impliquent des singularités des fonctions thermodynamiques (le palier bien connu des isothermes de pression en fonction de la densité par exemple). Or l'expression que donne la mécanique statistique de l'énergie libre d'un système est une fonction régulière (analytique) des différents paramètres tels que  $T$  ou

le volume  $V$ . On a alors l'impression que cette expression ne peut jamais prédire l'existence d'une transition de phase. Qu'en est-il réellement ? On dit que la question était si vivement débattue, sans qu'un consensus entre les grands physiciens présents ne puisse être obtenu que, par amusement, il a été procédé à un vote !

Ces discussions signalent la difficile émergence de la notion générale de « limite thermodynamique », limite où le système physique devient de taille infinie à densité constante, c'est-à-dire à rapport  $N/V$  constant. On sait maintenant que ce n'est que dans cette limite qu'apparaissent les singularités caractéristiques des transitions de phase ; le physicien Hendrik Kramers a joué un rôle important dans l'apparition de cette idée. Par exemple, les courbes supérieures de la figure 3.2 sont régulières et analytiques, sans singularité, alors que la courbe inférieure correspond à la limite thermodynamique et présente effectivement un point anguleux pour  $T = T_C$ . On voit clairement comment, au fur et à mesure que  $N$  et  $V$  croissent en gardant un rapport constant, la courbe fait apparaître une région de courbure de plus en plus grande ; pour  $N$  infini, elle devient un point anguleux, ce qui signale la présence d'un changement de phase. Ceci illustre sur un cas particulier comment une transition de phase correspond à un phénomène où certaines propriétés physiques du système (ici par exemple la dérivée de  $N_0$  par rapport à la température) varient d'autant plus rapidement sur un petit intervalle des paramètres que le système est grand.

La découverte d'Einstein a joué un rôle dans cette percée conceptuelle, en prédisant l'existence d'une transition dans un système suffisamment simple pour que des calculs mathématiques soient possibles (un gaz parfait). Auparavant, toutes les transitions de phase connues mettaient en jeu des interactions<sup>8</sup>, ce qui rendait les calculs beaucoup plus difficiles, voire infaisables. Le calcul exact d'Einstein a donc créé le cadre propice à l'apparition d'une notion nouvelle et universelle en physique [4, 6].

### 3.2.3 Un lien avec la superfluidité ?

La superfluidité d'un liquide est sa capacité à entretenir des écoulements sans friction, qui ne tendent donc pas à s'amortir rapidement par viscosité comme ils le font dans tous les liquides normaux. Depuis longtemps (1911), les expérimentateurs du laboratoire de Kamerlingh Onnes à Leyde avaient mis en évidence le phénomène de supraconductivité, phénomène assimilable à une superfluidité des électrons de conduction dans un métal, et se traduisant par une annulation brutale et complète de la résistivité électrique en dessous

---

<sup>8</sup>Comme rappelé plus haut, la transition de Bose-Einstein du gaz parfait est de nature purement entropique, pas énergétique (contrairement à la transition ferromagnétique par exemple, qui est induite par la tendance des spins à s'orienter les uns par rapport aux autres pour réduire leur énergie d'interaction ; elle met donc en jeu une compétition énergie-entropie).

d'une certaine température très basse (quelques kelvins)<sup>9</sup>. Puis, en 1937, est apparue la superfluidité de l'hélium 4 liquide : Kapitza d'une part, Allen et Misener d'autre part, observèrent qu'en dessous d'une température critique traditionnellement notée  $T_\lambda$  (environ 2,2 kelvins) apparaissent toute une série de phénomènes physiques qui seraient impossibles dans un fluide visqueux, comme par exemple l'écoulement rapide dans des films d'hélium liquide très minces. Ni dans un cas ni dans l'autre, les physiciens n'avaient de théorie satisfaisante pour expliquer ces phénomènes pourtant très frappants. En 1938, toutefois, Fritz London fait un premier pas, en émettant la conjecture selon laquelle la transition qui apparaît à  $T_\lambda$  provient d'une condensation de Bose-Einstein des atomes du liquide. Cette idée permit ensuite à Laslo Tisza de construire une théorie phénoménologique de l'hélium 4 liquide, dite du modèle à deux fluides, toujours en usage aujourd'hui.

Pour étayer son idée, London comparait la température  $T_\lambda$  à la température de transition  $T_C$  d'un gaz parfait de bosons de même masse et même densité qu'un liquide d'hélium 4 ; il trouvait pour  $T_C$  une valeur de 3,2 kelvins, du même ordre de grandeur que  $T_\lambda$ . C'était un excellent argument de plausibilité, mais cet accord relatif n'était-il pas le fruit d'un pur hasard ? En effet, dans un liquide, les atomes « roulent » en quelque sorte les uns sur les autres, restant en contact la plupart du temps ; il est donc évident que les interactions, au lieu d'être négligeables, jouent un rôle important, de sorte que le système est très différent d'un gaz parfait. De plus, on sait maintenant qu'un gaz sans interactions, comme celui considéré par Einstein, n'est jamais superfluide, ce qui montre bien les limites de l'intuition de London. La perplexité des physiciens sur la nature exacte du lien entre transition superfluide et condensation de Bose-Einstein était donc réelle. D'ailleurs le très grand physicien russe Lev Landau, souvent considéré comme le père de la théorie moderne des propriétés de l'hélium 4 superfluide, développa son approche sous une forme qui ne fait intervenir à aucun stade la notion de condensation de Bose-Einstein.

Il est amusant de remarquer que, au moment où London faisait sa suggestion, non seulement il était difficile pour les contemporains de dire s'il avait raison ou tort, mais personne ne savait définir précisément ce qu'il fallait entendre par condensation de Bose-Einstein dans un système en interaction. Ce n'est qu'en 1956 que des articles clarificateurs de Penrose et Onsager permirent de donner un sens précis à cette notion, en faisant intervenir à nouveau la notion de limite thermodynamique. Ces deux auteurs donnèrent même une

---

<sup>9</sup>La supraconductivité des métaux ne concerne pas des particules obéissant à la statistique de Bose-Einstein, mais de Fermi-Dirac. Nous savons maintenant grâce à la théorie BCS (du nom des trois physiciens Bardeen, Cooper et Schrieffer, qui l'ont introduite en 1957) que la condensation de paires de fermions dans la théorie BCS peut être vue comme une extension des idées d'Einstein sur la condensation dans un gaz de bosons. Ces paires sont faiblement interagissantes, ce qui ajoute des éléments théoriques essentiels non contenus dans l'article initial d'Einstein.

évaluation de la fraction condensée<sup>10</sup> dans l'hélium liquide très proche de la valeur acceptée actuellement. Curieusement, la qualité de ce résultat numérique impressionnant est maintenant considérée comme un hasard [7], mais ceci n'ôte rien à la grande valeur des idées théoriques générales introduites par London, Penrose et Onsager.

### 3.2.4 Le renouveau des gaz quantiques

On peut se demander pourquoi, sitôt l'idée d'Einstein connue, les expérimentateurs ne se sont pas précipités pour tenter de refroidir des gaz dilués, jusqu'à ce qu'ils atteignent la température prévue pour la transition. La raison en est simple : bien avant qu'on atteigne cette température, toutes les substances gazeuses subissent une condensation tout à fait ordinaire vers une phase liquide, puis vers une phase solide ; on est alors en présence de systèmes denses où les interactions dominent, et où la théorie d'Einstein ne s'applique plus. L'impression générale était donc que cette théorie ne donnait de prédictions intéressantes que dans un domaine de paramètres inaccessible en pratique, ce qui limitait son intérêt en le rendant purement académique.

Toutefois, à partir de 1975 environ, un vif renouveau d'intérêt se fait sentir pour le sujet ; plusieurs raisons l'expliquent. En premier lieu, toute une série d'expériences sur les semi-conducteurs [8] tendent à indiquer l'existence d'une condensation de Bose-Einstein pour des particules de brève durée de vie, les excitons<sup>11</sup>. Ces expériences sont très remarquées, mais leur interprétation reste délicate, du fait des nombreux paramètres qui interviennent dans des systèmes physiques aussi complexes. En second lieu, alors qu'il avait semblé sans espoir de tenter d'observer de « vrais » gaz à très basse température, apparaissent des suggestions pour éviter la condensation ordinaire vers un liquide, et maintenir des systèmes gazeux même à très basse température. Il s'agit de s'intéresser à des états gazeux métastables, mais de durée de vie suffisamment longue pour permettre des observations précises. En particulier, reprenant une suggestion de Hecht [9], un article de Stwalley et Nosanov [10] propose d'utiliser une polarisation des spins électroniques de l'hydrogène atomique pour bloquer la recombinaison moléculaire des atomes, et stabiliser ainsi un gaz d'hydrogène jusqu'à des températures arbitrairement basses. Ceci donnerait en principe accès à une condensation de Bose-Einstein dans un système dilué, donc proche des conditions où la théorie d'Einstein reste valable.

---

<sup>10</sup>Pour cette définition de la fraction condensée, des outils théoriques non disponibles à l'époque de l'article d'Einstein sont nécessaires. Penrose et Onsager donnent une définition générale, qui n'est pas limitée à un système sans interactions ou homogène (où la condensation se produit simplement dans le niveau d'énergie le plus bas accessible à une particule isolée). Il s'agit de considérer l'opérateur densité à  $N$  particules, de prendre la trace partielle sur  $N - 1$  particules pour obtenir l'opérateur à une particule, et enfin de considérer les valeurs propres de ce dernier opérateur. S'il existe une valeur propre qui croît proportionnellement à  $N$  dans la limite thermodynamique, on dit qu'il y a condensation de Bose-Einstein.

<sup>11</sup>Il s'agit de paires liées électron-trou, analogues à des atomes d'hydrogène légers, qui peuvent être formées dans des cristaux semi-conducteurs.

Quelques années plus tard (1980) un groupe expérimental, à Amsterdam, montre qu'il est effectivement possible d'obtenir un tel gaz d'hydrogène atomique polarisé à une température de l'ordre de 0,3 kelvin, avec une densité numérique de quelques  $10^{14}$  atomes par centimètre cube [11]. Cette performance était obtenue grâce un recouvrement des parois de la cellule contenant le gaz à basse température par l'hélium 4 liquide, afin d'éviter la recombinaison des atomes sur la paroi. Hélas, ces conditions sont toujours loin de la condensation de Bose-Einstein, puisque qu'avec ces chiffres le facteur  $\rho\lambda_T^3$  est seulement de l'ordre de  $10^{-6}$ , bien inférieur à la valeur 2,6 de la formule (3.14). C'est le début d'une longue course à la fois vers des températures plus basses et des densités plus élevées, afin de franchir enfin les conditions permettant d'obtenir le but final, la condensation de Bose-Einstein. Cette quête durera presque 30 ans et fournira au passage toute une série de résultats passionnants.

Plusieurs groupes expérimentaux entreprirent de réussir cet exploit [11,12]. Une première nécessité était de réduire encore les problèmes difficiles de recombinaison sur la paroi, subsistant malgré l'utilisation des enduits d'hélium liquide. Le piégeage magnétique, initialement développé par W. Paul dans les années 70 pour confiner des neutrons, fut adapté aux atomes neutres. Dans ces pièges, les atomes sont maintenus en lévitation grâce à un champ magnétique inhomogène, au centre d'une enceinte où règne un vide très poussé ( $10^{-9}$  Pa). Chaque atome porte un moment magnétique  $\boldsymbol{\mu}$ , qui se couple au champ magnétique local pour donner naissance à l'énergie d'interaction  $E = -\boldsymbol{\mu} \cdot \mathbf{B}$ . Cette énergie joue le rôle d'une énergie potentielle pour le mouvement du centre de masse de l'atome, et la force magnétique qui en résulte permet de confiner l'atome. Considérons par exemple un atome au voisinage d'un point  $O$  où l'amplitude  $|\mathbf{B}(\mathbf{r})|$  du champ magnétique est minimale. Si le moment magnétique de l'atome a une direction opposée à celle du champ magnétique local, l'énergie magnétique s'écrit  $E(\mathbf{r}) = -\boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{r}) = |\boldsymbol{\mu}||\mathbf{B}(\mathbf{r})|$ . Cette énergie potentielle est elle-même minimale en  $O$ , et l'atome effectue des oscillations stables autour de ce point. Dans les conditions usuelles de fonctionnement des pièges magnétiques, ces oscillations se font à une fréquence relativement basse, de l'ordre de la centaine de Hertz, et l'approximation qui consiste à supposer que les orientations de  $\boldsymbol{\mu}$  et  $\mathbf{B}$  restent en permanence opposées est très bien vérifiée.

Des techniques de piégeage magnétique furent donc mises en œuvre pour éviter tout contact du gaz avec des parois solides (techniques également appliquées aux atomes alcalins). Une autre idée qui devait s'avérer très importante pour l'observation de la condensation de Bose-Einstein dans les alcalins était celle de mesurer des profils de densité d'un gaz soumis à un gradient de champ magnétique, afin d'obtenir une signature claire du phénomène [13,14]. Surtout, une autre grande percée vint de l'idée proposée par Hess [15] : utiliser la méthode dite du refroidissement par évaporation, où l'on sacrifie volontairement une partie des atomes, ceux dont l'énergie est la plus grande, à qui l'on fait quitter le piège magnétique. Ceci a pour effet de laisser dans le piège un gaz composé d'atomes dont l'énergie moyenne est plus faible, donc un gaz plus

froid. Mais les expériences avec l'hydrogène sont tellement difficiles qu'il fallut attendre presque 20 ans avant que l'observation de la condensation de Bose-Einstein dans l'hydrogène atomique soit réalisée [16] ! Entre-temps, comme nous allons le voir, une grande percée avait été effectuée avec les atomes alcalins.

### 3.2.5 Un régime intermédiaire semi-quantique

En attendant le succès des expériences de condensation de Bose-Einstein, tout un courant de recherches apparut en parallèle, portant également sur les gaz en régime quantique, mais dans une situation moins difficile à réaliser expérimentalement. L'idée est de s'intéresser à un régime intermédiaire, celui où la longueur d'onde thermique (3.13) des atomes est petite devant leur distance moyenne de séparation, mais grande par rapport à la portée du potentiel interatomique. Dans ces conditions, les propriétés d'équilibre sont données avec une bonne approximation par la mécanique statistique classique<sup>12</sup>, mais les effets quantiques jouent fortement sur les propriétés des collisions, et donc sur les propriétés de transport dans le gaz, la conduction calorifique par exemple.

Au début des années 60 régnait un certain mystère concernant les propriétés de transport dans l'hélium gazeux, en particulier l'hélium 3. Il s'agissait pourtant de systèmes dilués où on aurait pu attendre un excellent accord entre théorie et expérience ! Lorsque deux atomes dans le gaz subissent une collision, la section efficace de ce processus est affectée par la statistique quantique. Pour deux atomes d'hélium 4 à très basse température, on montre que l'effet de la statistique de Bose-Einstein est simplement un doublement de la section efficace de collision – nous reviendrons sur ce sujet au paragraphe 3.4.1 ; effectivement, on constate expérimentalement que la viscosité et la conduction calorifique sont deux fois plus faibles que la valeur qu'elles auraient pour un gaz de Boltzmann. Mais, pour l'hélium 3, l'accord théorie-expérience n'était pas bon. Les atomes de cet isotope obéissent à la statistique de Fermi-Dirac, et ont deux états internes de spin. Lorsque ces atomes sont dans le même état interne, cette statistique intervient fortement et modifie la section efficace qui, par exemple, s'annule à la limite des basses énergies. En revanche, lorsque les atomes sont dans des états internes différents, ils sont en principe discernables et la section efficace est la même que pour des particules de Boltzmann (qui ne s'annule pas à cette limite). Comme il y a, dans chaque collision, une chance sur deux pour que les deux atomes soient dans le même état de spin, il est naturel de pondérer également les deux processus de collision pour le calcul du coefficient de transport. Cette façon de faire donnait effectivement des résultats théoriques en bon accord avec les expériences pour la conduction calorifique et la viscosité, mais en net désaccord pour la diffusion de spin. Il y avait donc un réel problème à résoudre !

Pour résoudre ce paradoxe, Emery [17] proposa en 1964 d'utiliser des sections efficaces différentes selon les coefficients de transport étudiés : la

<sup>12</sup>La condensation de Bose-Einstein est donc exclue.

pondération qui vient d'être mentionnée s'applique pour la conduction calorifique et la viscosité ; pour le calcul du coefficient de diffusion de spin, c'est uniquement la section efficace de particules discernables qu'il faut prendre en compte. L'approche d'Emery était basée sur des arguments physiques simples et convaincants, mais ne disait pas de façon systématique quelle section efficace devait être utilisée dans chaque cas ; de plus, elle supposait implicitement que la polarisation moyenne de spin du gaz était nulle.

Pour bien comprendre le problème, il fallait donc une théorie quantique générale du transport dans un gaz classique, mais en régime de collisions quantiques. Elle fut proposée en 1982 [18] sous une forme qui fournit, non seulement une confirmation et une généralisation des idées de Emery, mais l'apparition de plusieurs idées nouvelles. La première était de contrôler la valeur de la viscosité ou de la conduction calorifique d'un gaz d'hélium 3 en variant sa polarisation nucléaire [19], prédiction qui fût bientôt confirmée par des expériences [20] utilisant les méthodes du pompage optique pour créer cette polarisation. La seconde, plus surprenante encore, était qu'en présence de polarisation nucléaire le phénomène de diffusion de spin perdait son caractère purement dissipatif et devenait oscillant ; en d'autres termes, il existe des ondes de spin quantiques même dans un gaz dilué dont les propriétés d'équilibre restent classiques. Ces prédictions concernent aussi bien les fermions que les bosons, et furent bientôt vérifiées par des expériences dans l'hélium 3 et l'hydrogène atomiques polarisés [21, 22]. Ainsi les physiciens s'habituèrent-ils progressivement à manipuler ces gaz inhabituels et progressaient-ils vers l'obtention de la fameuse transition de Bose-Einstein !

### 3.3 La percée des alcalins

Au cours des vingt dernières années, le développement des méthodes de refroidissement des atomes par laser a ouvert une voie originale vers la préparation d'un condensat de Bose-Einstein. En tirant parti des échanges d'impulsion entre lumière et matière, il est désormais possible de diminuer considérablement l'agitation thermique d'une assemblée d'atomes, sans jamais la mettre en contact avec une paroi matérielle refroidie. Les températures qu'on atteint par refroidissement laser sont dans le domaine du microkelvin, et correspondent à une vitesse d'agitation de quelques millimètres par seconde seulement. Par comparaison, à température ordinaire, la vitesse d'agitation thermique des atomes est de plusieurs centaines de mètres par seconde.

Pour refroidir efficacement une vapeur atomique grâce à un faisceau laser, la seule contrainte est de disposer d'une source lumineuse relativement intense, de longueur d'onde voisine de la fréquence de résonance des atomes. Cette condition a conduit les physiciens à remettre en cause le choix naturel de l'atome d'hydrogène que nous avons évoqué au paragraphe précédent, pour s'intéresser plutôt aux atomes alcalins. La raie de résonance de ces atomes est en effet située dans le domaine visible ou infra-rouge proche du spectre

électromagnétique, pour lequel des sources laser fiables sont disponibles. Au contraire, la raie de résonance de l'hydrogène est située dans l'ultra-violet lointain, ce qui rend cet élément extrêmement difficile à manipuler par laser.

### 3.3.1 La manipulation d'atomes par laser

Plusieurs forces entrent en jeu dans la manipulation d'atomes par la lumière. La plus connue est la force de pression de radiation, qui avait été présentée par Kepler pour expliquer l'orientation de la queue des comètes par rapport au soleil. Un atome placé dans un faisceau laser est poussé comme une poussière dans un courant d'air. La force exercée par le laser est d'autant plus grande que sa fréquence est voisine de la fréquence de résonance de l'atome. Ceci est mis à profit dans le « refroidissement Doppler », le plus simple des mécanismes de refroidissement d'atomes par laser, proposé en 1975 par T. Hänsch et A. Schawlow pour des atomes neutres et par D. Wineland et H. Dehmelt pour des ions. Une description plus détaillée des divers aspects des méthodes de piégeage et de refroidissement laser peut être trouvée dans les conférences Nobel 1997 (voir « références générales » à la fin de ce texte).

La mise en œuvre du refroidissement Doppler est schématisée sur la figure 3.3. Pour simplifier, nous considérons le mouvement des atomes selon une seule direction. On éclaire les atomes par deux ondes laser de même fréquence et de même intensité, ces deux ondes se propageant en sens opposé. La fréquence des ondes lumineuses est choisie légèrement inférieure à la fréquence de résonance des atomes. Considérons d'abord un atome immobile ; par symétrie, les deux forces de pression de radiation sont de même intensité et de sens opposé. Elles se compensent et l'atome ne ressent aucune force résultante. Considérons maintenant un atome en mouvement, par exemple vers la droite. En raison de l'effet Doppler, l'atome « verra » l'onde venant à sa rencontre avec une fréquence augmentée, donc plus proche de résonance. La force de pression de radiation créée par cette onde sera donc plus grande que pour un atome au repos. En revanche, l'onde venant de la gauche, allant dans le même sens que l'atome, est vue par l'atome avec une fréquence plus basse que sa fréquence réelle. La force qu'elle engendre est diminuée par rapport à celle agissant sur un atome immobile. Au final, l'équilibre entre les deux forces

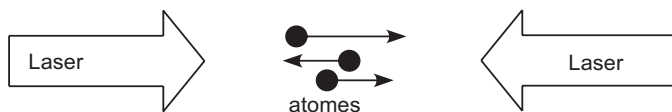


FIG. 3.3 – Principe du refroidissement d'atomes par laser, dans une géométrie unidimensionnelle. La fréquence des faisceaux laser est choisie légèrement inférieure à la fréquence de résonance atomique. En raison de l'effet Doppler, un atome en mouvement subit une force de pression de radiation opposée à sa vitesse, et cette force le freine.



est rompu pour un atome en mouvement, au profit du faisceau laser venant à la rencontre de l'atome. La force résultante est opposée à la vitesse atomique ; elle correspond donc à une force de friction. Il en ira de même si on considère un atome en mouvement vers la gauche, qui ressentira une force dirigée vers la droite.

Le raisonnement ci-dessus se transpose sans difficulté à trois dimensions. Les photons du laser créent pour l'atome un milieu visqueux, et les atomes y sont englués comme une cuillère dans un pot de miel. Ce milieu visqueux, dont le volume typique est de l'ordre du centimètre cube, a été baptisé « mélasse optique ». On y capture typiquement un milliard d'atomes, qui sont refroidis à une température de l'ordre d'une centaine de microkelvins en quelques millièmes de secondes. En disposant la mélasse optique dans un gradient de champ magnétique, et en choisissant convenablement la polarisation des faisceaux laser, on crée, en plus de la force de friction qui diminue la vitesse, une force de piégeage qui tend à accumuler les atomes au centre de la mélasse, là où le champ magnétique s'annule. Le dispositif correspondant s'appelle un piège magnéto-optique. Signalons également que d'autres mécanismes de refroidissement d'atomes par laser viennent renforcer le refroidissement Doppler décrit ci-dessus ; c'est le cas du refroidissement Sisyphé par exemple, ainsi baptisé car l'atome est placé dans une situation où il ne cesse de gravir des collines de potentiel jusqu'à ce qu'il s'immobilise. Les températures résiduelles sont alors de l'ordre du microkelvin.

Le mécanisme du refroidissement Doppler est une transposition presque directe d'un raisonnement d'Einstein publié en 1917. Einstein s'intéressait à l'époque aux échanges d'énergie et d'impulsion entre un atome et le rayonnement. Il avait introduit les trois processus de base de l'interaction lumière-matière : (i) l'absorption, quand un photon disparaît et l'atome est porté dans un niveau excité ; (ii) l'émission spontanée, quand un atome préparé dans un état excité en l'absence de tout champ extérieur, tombe dans un état moins excité en émettant un photon ; (iii) l'émission stimulée, où la même transition qu'au (ii) se produit pour l'atome, mais cette fois-ci en présence d'autres photons, ce qui augmente la probabilité de la transition. Pour justifier l'introduction de ces trois processus, Einstein s'était posé le problème du mouvement d'un atome dans le rayonnement du corps noir, c'est-à-dire un bain de photons à l'équilibre thermodynamique. Einstein, qui avait entière confiance dans la mécanique statistique, savait que les échanges d'impulsion entre l'atome et les photons devaient conduire à une thermalisation de l'atome. Il montra que c'était effectivement le cas ; le mouvement atomique est un mouvement brownien, avec une force de friction d'origine identique à celle vue plus haut : à cause de l'effet Doppler, l'atome en mouvement dans le rayonnement du corps noir absorbe préférentiellement des photons venant à sa rencontre ; il est donc freiné et sa vitesse moyenne s'annule. Le caractère aléatoire des processus d'absorption et d'émission entraîne quant à lui une diffusion de l'impulsion atomique. La compétition entre ces deux composantes, friction et diffusion, conduit à un état d'équilibre ; la répartition des vitesses atomiques est donnée

par la distribution de Maxwell-Boltzmann, de même température que le bain de photons. Ce raisonnement d'Einstein est un véritable tour de force, dans lequel il parvient à identifier les trois processus fondamentaux de l'interaction lumière-matière sans disposer de la théorie quantique du champ électromagnétique (cette dernière ne sera fondée qu'une dizaine d'années plus tard par Dirac).

### 3.3.2 Vers la condensation

L'utilisation exclusive du refroidissement laser n'a pas encore permis d'atteindre la condensation de Bose-Einstein. En effet, si on cherche à augmenter la densité atomique au sein des mélasses optiques pour atteindre le seuil de condensation, l'irradiation laser favorise la formation de molécules : quand deux atomes sont proches l'un de l'autre, l'absorption d'un photon peut conduire à la formation d'un dimère qui s'échappe ensuite de la mélasse optique. Fort heureusement, les progrès qui avaient été réalisés dans les années 1980–1990 pour la condensation d'un gaz d'hydrogène atomique ont indiqué la voie à suivre pour contourner cet obstacle : une fois la mélasse optique obtenue, on transfère les atomes dans un piège magnétique, on éteint tous les faisceaux laser, et on procède au refroidissement par évaporation pour atteindre la condensation.

Le piège magnétique utilisé pour les atomes alcalins est identique dans son principe à celui mis en œuvre pour l'étude de la condensation de l'hydrogène. On réalise un minimum local du champ magnétique  $\mathbf{B}$  et on oriente le moment magnétique  $\boldsymbol{\mu}$  des atomes dans la direction opposée au champ magnétique. L'énergie d'interaction entre le moment magnétique atomique et le champ magnétique,  $E = -\boldsymbol{\mu} \cdot \mathbf{B}$ , est minimale au même endroit que  $|\mathbf{B}|$  et les atomes sont confinés dans cette cuvette de potentiel.

Pour évaporer les atomes, on modifie la hauteur du puits de potentiel magnétique grâce à une onde radio de pulsation  $\omega$  ajustable. Cette onde fait basculer les moments magnétiques résonnants avec elle, c'est-à-dire ceux situés sur une surface de champ magnétique constant, telle que  $\hbar\omega = \mu B$ . On choisit une grande valeur initiale de  $\omega$  : ceci correspond à une profondeur élevée pour le puits de potentiel, et permet de confiner même des atomes d'énergie importante. Le refroidissement par évaporation se fait ensuite en diminuant progressivement  $\omega$ . La profondeur du puits de potentiel en fin d'évaporation est très faible, de l'ordre de quelques microkelvins seulement.

Il faut enfin détecter le condensat qu'on espère avoir formé. L'utilisation d'atomes alcalins facilite considérablement cette tâche par rapport aux expériences menées avec l'hydrogène. Il suffit en effet d'éclairer le nuage d'atomes par une brève impulsion lumineuse résonnante, comme indiqué sur la figure 3.4. On mesure l'absorption ou le déphasage de cette lumière par l'assemblée atomique grâce une caméra rapide. On a ainsi accès à la distribution spatiale des atomes dans le potentiel magnétique. On peut également couper le piège magnétique et laisser l'assemblée atomique s'étaler pendant une durée

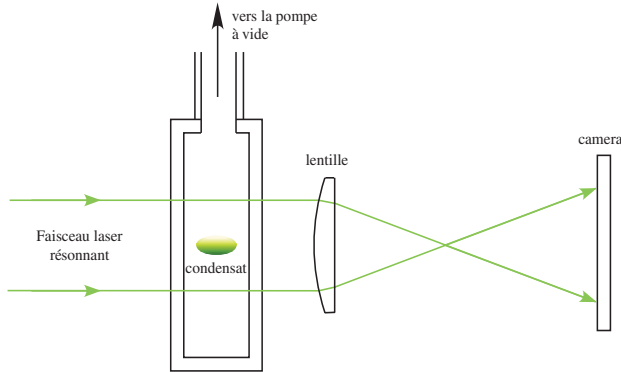


FIG. 3.4 – Principe de la détection d'un condensat. On envoie sur le nuage atomique une brève impulsion lumineuse, résonnante avec les atomes. On mesure grâce à une caméra très sensible l'ombre portée par le nuage atomique sur le faisceau laser. On peut alors en déduire les principaux paramètres du nuage : répartition spatiale, nombre d'atomes, température.

de quelques dizaines de millisecondes, avant d'envoyer l'éclair lumineux. De l'étendue du nuage atomique après étalement, on déduit la distribution en vitesse initiale.

La figure 3.5 montre un exemple de distributions en vitesse d'un gaz d'atomes de sodium piégés, pour différentes températures. À relativement haute température (Fig. 3.5a, quelques microkelvins), la distribution est donnée par la loi classique de Maxwell-Boltzmann. Quand on abaisse la température en dessous de la température critique  $T_C$  de condensation (Fig. 3.5b), une fraction significative des atomes s'accumule dans le pic central étroit correspondant au condensat de Bose-Einstein. On peut également choisir les paramètres du refroidissement par évaporation pour ne conserver que les atomes du pic condensé (Fig. 3.5c) : les atomes sont alors tous dans le même état, décrit par la même fonction d'onde  $\psi(\mathbf{r})$ . On dispose alors d'un condensat de Bose-Einstein pur.

C'est en 1995 que les premiers condensats atomiques ont été observés, à Boulder dans le groupe de E. Cornell et C. Wieman (atomes de rubidium), puis au MIT dans l'équipe de W. Ketterle (atomes de sodium) ; pour une description plus détaillée, voir les conférences Nobel de ces auteurs dans la rubrique « références générales » à la fin de ce texte. Ce type d'expérience a ensuite été reproduit avec plusieurs autres espèces atomiques. Tous les alcalins stables (lithium, sodium, potassium, rubidium, césium) ont été condensés ainsi que l'ytterbium. Un condensat d'atomes d'hélium, préparés dans un niveau électronique métastable, a également été produit. Enfin, signalons que la condensation de l'atome d'hydrogène a finalement été observée en 1998 [16] : c'était un juste retour des choses pour un système sur lequel plusieurs concepts

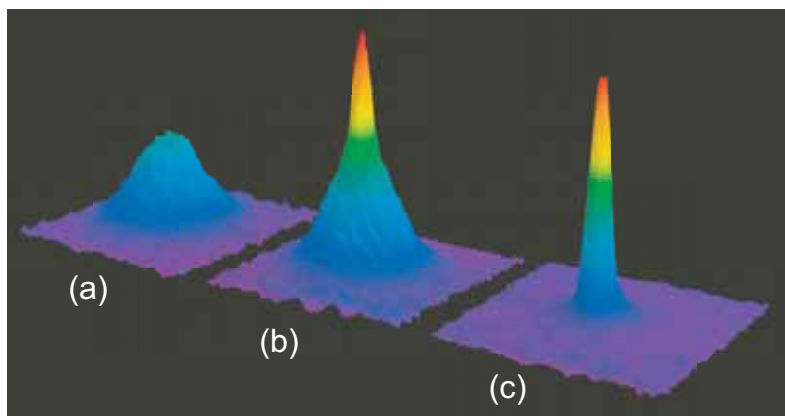


FIG. 3.5 – Évolution de la distribution en vitesse d'un gaz d'environ 1 million d'atomes de sodium lors du refroidissement par évaporation dans un piège magnétique. La photo (a) correspond à une température de quelques microkelvins, supérieure à la température de condensation ; la répartition des vitesses atomiques est proche d'une distribution de Maxwell-Boltzmann. La photo (b) correspond à une température légèrement inférieure à la température de condensation (de l'ordre du microkelvin). Le pic central étroit correspond aux atomes condensés ; ce pic est superposé à un piédestal plus large, correspondant aux atomes non condensés. La photo (c) a été obtenue en poussant encore plus loin le refroidissement par évaporation : le nuage atomique est alors quasi-totalement condensé (photographie fournie par W. Ketterle, MIT).

importants (piégeage magnétique, évaporation) avaient initialement été développés.

L'observation de nuages atomiques aussi froids que ceux de la figure 3.5 peut sembler choquante du point de vue de l'équilibre thermodynamique. À une température de l'ordre du microkelvin, l'état d'équilibre d'une assemblée d'atomes de sodium ou de rubidium est un solide. Comment est-il possible d'observer un gaz d'atomes dans ces conditions ? La réponse à cette question provient de la faible densité spatiale de ces assemblées d'atomes. Pour former des molécules ou des agrégats qui conduiront ultimement à un bloc de rubidium solide, il faut des collisions impliquant au moins trois partenaires. Ainsi, si on cherche à former un dimère dans une collision entre deux atomes, il faut qu'un troisième atome soit au voisinage des deux premiers (typiquement à quelques nanomètres, qui correspond à la taille du dimère) pour emporter l'énergie libérée lors de la formation de la molécule. Or, pour ces systèmes dilués, la probabilité d'avoir trois atomes au voisinage immédiat les uns des autres est très faible. Il s'ensuit que la formation de molécules est extrêmement lente. Elle ne constitue donc pas un obstacle à l'observation d'un condensat atomique gazeux : ce condensat n'est que métastable, mais sa durée de vie est

en pratique suffisamment longue pour qu'on puisse lui appliquer les principes habituels qui régissent l'équilibre thermodynamique.

### 3.3.3 Interférences de condensats

L'accumulation de la majorité des atomes dans le même état quantique confère aux condensats de Bose-Einstein des propriétés de cohérence spectaculaires. Le seul équivalent connu est la lumière laser, constituée d'un grand nombre de photons dans le même mode du champ électromagnétique.

Une très belle illustration de cette cohérence a été apportée par une expérience d'interférence entre condensats menée au MIT. En éclairant le centre du piège magnétique par une nappe de lumière (Fig. 3.6a), on peut réaliser un double puits de potentiel. La nappe lumineuse joue le rôle d'une barrière répulsive, qui empêche en pratique tout passage d'atomes condensés entre la partie gauche et la partie droite. On prépare par évaporation un condensat dans chacune de ces deux régions, puis on coupe le potentiel magnétique, ainsi que la barrière lumineuse. Les deux condensats s'étalent, se recouvrent, et on prend une photo de la distribution spatiale des atomes (Fig. 3.6b). Cette distribution présente des franges d'interférence avec un fort contraste (70 %), ce qui prouve la cohérence relative des sources atomiques constituées par chacun des deux condensats.

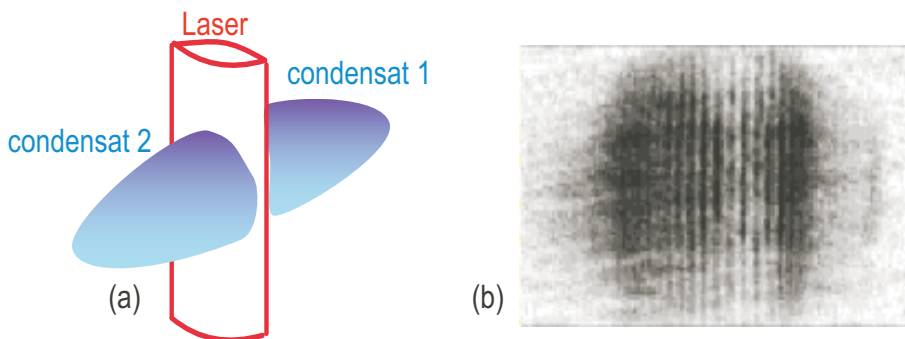


FIG. 3.6 – (a) On réalise un double puits de potentiel en disposant au centre du piège magnétique un faisceau lumineux, de fréquence supérieure à la fréquence de résonance atomique. Ce faisceau crée une barrière de potentiel, qui permet d'obtenir après évaporation deux condensats indépendants. (b) Lorsqu'on coupe le piège magnétique et la barrière lumineuse, chaque condensat s'étale. Dans la zone de recouvrement entre les deux condensats, on observe une figure d'interférence d'ondes de matière. Le contraste élevé des franges d'interférences est une conséquence directe de l'accumulation des atomes de chaque condensat dans un même état quantique (photographie fournie par W. Ketterle, MIT).

Pour rendre compte de manière quantitative de la figure d'interférence, on décrit chaque condensat par un champ d'onde de matière, comme on le fait en optique pour un faisceau lumineux monochromatique. L'évolution des deux ondes de matière après la coupure du piège est donnée par l'équation de Schrödinger, qui remplace ici les équations de Maxwell décrivant la propagation des faisceaux lumineux. La répartition spatiale des atomes à l'instant de la photo 3.6b s'obtient en additionnant les amplitudes des deux ondes, ce qui permet d'expliquer la figure d'interférence obtenue. Notons toutefois qu'on ne peut pas prévoir a priori la position de la frange brillante centrale. Celle-ci dépend de la phase relative initiale entre les deux condensats, qui fluctue d'une réalisation de l'expérience à une autre : si l'on refait plusieurs fois cette expérience, on obtiendra pour chaque réalisation un profil d'interférence semblable à celui de la figure 3.6b, mais son système de franges sera décalé aléatoirement.

### 3.4 Au-delà du gaz parfait

Dans cette partie, nous discutons la façon dont les interactions entre atomes modifient les propriétés physiques d'un gaz condensé. Ces interactions, lorsqu'elles sont répulsives, font apparaître la superfluidité. Cette dernière se signale à la fois par l'existence de courants permanents et celle de tourbillons quantifiés dans le fluide. Nous commençons par quelques considérations théoriques, puis décrivons ensuite quelques expériences.

#### 3.4.1 Description, manipulation et effets des interactions à l'équilibre

Pour les atomes d'un gaz qui est à la fois dilué et à basse température, la théorie générale des collisions en mécanique quantique se simplifie considérablement. On peut alors ne prendre en compte que les propriétés à longue distance des interactions binaires entre atomes, et utiliser le fait que leur énergie relative de collision est très faible. Toutes les propriétés de l'interaction binaire sont alors contenues dans une seule quantité, la « longueur de diffusion », traditionnellement notée  $a$  ; par exemple, la section efficace de collision entre atomes discernables est simplement  $4\pi a^2$ . Si  $a$  est positif, l'effet de l'interaction entre atomes sur les propriétés macroscopiques du gaz est répulsif (augmentation de la pression par rapport à celle du gaz parfait par exemple) ; si  $a$  est négatif, l'effet est attractif. Bien sûr, pour des atomes alcalins, le véritable potentiel d'interaction est en fait toujours attractif (sauf à très courte distance), ce qui permet l'existence de nombreux états moléculaires liés (pour un gaz dilué, nous avons déjà vu que la formation de ces états liés est fortement bloquée). Mais il suffit que  $a$  soit positif pour que l'effet global des interactions dans un gaz dilué soit le même que si le potentiel était réellement répulsif.

À la limite thermodynamique, on peut montrer qu'un gaz de Bose-Einstein devient fortement instable si  $a$  est négatif : il perd toute résistance à la compression et s'effondre sur lui-même (implosion). C'est la raison pour laquelle on ne s'intéresse généralement qu'aux valeurs positives de  $a$ , ou alors aux valeurs négatives uniquement dans le cas de systèmes suffisamment petits pour éviter l'instabilité. Prenons par exemple l'extension spatiale d'un condensat gazeux dans un piège harmonique. En l'absence d'interactions, son extension est simplement celle de l'état fondamental d'un atome unique placé dans le potentiel harmonique. Si  $a$  est positif et si un nombre suffisant d'atomes sont condensés, les interactions entre atomes du condensat entrent en jeu, et augmentent considérablement l'extension spatiale du condensat afin de réduire son énergie potentielle. Cet effet est calculable à partir de l'équation de Gross-Pitaevskii à laquelle obéit la fonction d'onde  $\Psi(\mathbf{r})$  du condensat ; on montre que la densité spatiale du condensat prend alors une forme caractéristique en parabole inversée, qui dépend de  $a$ . Si maintenant la longueur de diffusion  $a$  est négative, et si le gaz est contenu dans un piège qui le confine suffisamment, on montre que le condensat peut être stabilisé par le potentiel harmonique ; sous l'effet de l'attraction entre atomes, il prend une taille plus petite que pour le gaz parfait. Cependant, dès qu'on dépasse une certaine valeur critique du nombre total d'atomes dans le piège, le condensat implose et disparaît.

La longueur de diffusion ne doit pas être considérée comme une donnée intangible pour un gaz atomique donné : il est en effet possible d'utiliser diverses méthodes pour en contrôler la valeur. La plus efficace semble être l'utilisation d'une « résonance de Feshbach ». Cet effet met en jeu la dépendance en champ magnétique de la position des niveaux atomiques, ainsi que des résonances énergétiques entre niveaux moléculaires ; nous ne le décrirons pas ici en détail. Pour certaines valeur du champ  $B$ , il se produit des variations très rapides de  $a$ , et même une divergence pour une valeur précise de  $B$ . De chaque côté de cette valeur,  $a$  prend des valeurs qui sont grandes mais de signe opposé. Il est donc possible de contrôler le paramètre d'interaction des atomes et de le varier en fonction du temps ; on peut par exemple passer d'une situation de répulsion à une situation d'attraction, par la simple manœuvre d'un paramètre extérieur aussi accessible que le champ magnétique. C'est une très grande souplesse qui n'existe pas dans le cas des systèmes denses.

De plus, même pour une valeur donnée de  $a$ , plusieurs types de collisions peuvent se produire, suivant que chacun des deux atomes est, soit dans le condensat, soit dans le nuage thermique. En mécanique quantique, une façon de tenir compte de la statistique de Bose-Einstein est d'effectuer une opération de symétrisation du vecteur d'état (ou de la fonction d'onde) de particules numérotées. Partant de particules dans des états quantiques orthogonaux, on constate que cette symétrisation crée des corrélations entre les positions : deux particules de Bose ont plus de chances de se trouver proches l'une de l'autre que d'être éloignées, même en l'absence d'interactions. Il se trouve plus précisément que la probabilité de les trouver au même point est deux fois plus

grande qu'avant la symétrisation<sup>13</sup>. Comme les atomes doivent nécessairement être proches pour interagir, du fait de la courte portée de leur potentiel, cet effet double leur probabilité d'interaction<sup>14</sup>. En revanche, si les deux particules numérotées occupent le même état quantique individuel, la symétrisation n'a aucun effet particulier et ne change rien à leurs interactions.

Dans le nuage thermique, les atomes ont effectivement accès à un grand nombre d'états quantiques individuels, de sorte que leur probabilité d'interaction est effectivement doublée par cet effet. En revanche, des atomes dans le condensat occupent tous le même état quantique individuel de sorte que l'interaction reste inchangée. Enfin, pour les interactions entre deux atomes condensé et non condensé, l'augmentation d'un facteur 2 se produit également puisque les états quantiques individuels sont différents. On comprend ainsi que des interactions répulsives tendent à favoriser le condensat, puisque l'évaporation d'un atome hors du condensat le fait passer dans un état où son énergie d'interaction répulsive est plus élevée. En d'autres termes, l'argument entropique<sup>15</sup> du paragraphe 3.1.2 se double maintenant d'un argument énergétique ; l'occupation d'un seul état quantique par un nombre macroscopique d'atomes est favorisée. Dans un gaz dilué de densité donnée, on s'attend donc à ce que la population du condensat soit une fonction croissante de  $a$ .

De même que la pression d'un gaz de Bose dépend de  $a$ , c'est le cas des autres propriétés thermodynamiques à l'équilibre ; la température critique  $T_C$  est déplacée par rapport à la valeur  $T_C^0$  prévue par l'équation (3.9) du gaz parfait. Quelle est alors la correction d'ordre le plus bas en  $a$  à la température critique, analogue à la seconde correction du viriel pour la pression ? Il est curieux qu'un problème physique d'apparence aussi simple ait donné lieu à tant d'erreurs dans la littérature, conduisant à prédire des variations de la température critique selon des puissances différentes de  $a$ , et même avec des signes opposés. Le problème n'a en fait été résolu que récemment, ce qui montre bien que sa simplicité n'est qu'apparente.

Si l'on appelle  $\Delta T_C = T_C - T_C^0$  le déplacement de température critique introduit par les interactions à densité constante, le but de la théorie est de calculer le rapport sans dimension  $\Delta T_C/T_C$  en fonction d'un nombre sans dimension caractérisant les interactions. Pour ce dernier, il est naturel de choisir le rapport entre la longueur de diffusion  $a$  et la distance moyenne entre atomes, c'est-à-dire simplement  $a\rho^{1/3}$ . Lorsque ce paramètre est petit, le gaz est presque parfait ; on pourrait alors s'attendre a priori à ce que  $\Delta T_C/T_C$  puisse être exprimé comme une série de puissances de  $a\rho^{1/3}$ , comme

<sup>13</sup>Pour deux atomes, c'est un facteur 2! qui intervient, pour trois un facteur 3!, etc.

<sup>14</sup>Pour des fermions, c'est le contraire qui se produit, car ce n'est pas une symétrisation qu'il faut appliquer, mais une antisymétrisation. On voit alors que les fermions sont eux aussi corrélés statistiquement, mais d'une façon qui les empêche de se trouver au même point de l'espace. Ainsi leurs interactions à courte portée sont inhibées par la statistique quantique. Il en résulte des effets physiques fort intéressants qui sont brièvement mentionnés dans la conclusion.

<sup>15</sup>L'argument du paragraphe 3.1.2 porte sur le décompte des états accessibles au système physique. Un système physique possède une entropie d'autant plus grande qu'il a accès à un grand nombre d'états quantiques distincts.



habituellement en théorie des perturbations. Or une théorie détaillée [23] (menée pour un gaz homogène dans une boîte et à la limite thermodynamique) montre que ce n'est pas le cas. Certes, le premier terme de correction relative à la température critique s'écrit simplement sous la forme  $c_1 (a\rho^{1/3})$ , comme si c'était le début d'un développement limité en puissances de  $a$ . Mais il est possible de montrer que cette correction linéaire en  $a$  ne peut pas être obtenue dans le cadre d'une théorie des perturbations en puissances de  $a$ , ce qui se traduit en pratique par une grande difficulté à déterminer le coefficient  $c_1$  ; il a fallu des calculs numériques récents pour trouver  $c_1 \simeq 1,3$ . Le terme de correction suivant, lui, s'exprime sous une forme plus compliquée puisqu'il est proportionnel à  $(a\rho^{1/3})^2 \ln(a\rho^{1/3})$ , manifestant ainsi clairement la singularité du problème. On retrouve là une conséquence du fait qu'un gaz parfait, pour lequel  $a$  est nul, est en fait un système marginalement stable, puisqu'un gaz s'effondre (à la limite thermodynamique) dès que  $a$  prend une valeur négative, si petite soit-elle. Il n'est donc pas étonnant que le comportement du système donne lieu à des singularités mathématiques.

### 3.4.2 La superfluidité

Venons-en maintenant aux propriétés hors d'équilibre du gaz. Nous allons voir pourquoi un gaz condensé à basse température peut être le siège de courants permanents, sans amortissement par viscosité, en d'autres termes pourquoi il peut être superfluide. La présence d'interactions répulsives entre les atomes joue un rôle crucial dans ce phénomène. Le premier élément à noter est que cette répulsion tend à stabiliser le condensat contre son évaporation vers le nuage thermique, pour les raisons que nous venons de mentionner. Nous caractériserons l'état quantique unique occupé par les particules du condensat par une fonction d'onde  $\Psi(\mathbf{r})$ .

Cette tendance à occuper un état unique ne signifie pas que  $\Psi(\mathbf{r})$  soit rigide, par exemple qu'il doive toujours correspondre à l'état fondamental de l'hamiltonien à une particule ; au contraire, il reste une certaine souplesse dans le choix de  $\Psi(\mathbf{r})$ . Nous avons déjà vu que, pour un gaz à l'équilibre contenu dans un piège harmonique, l'état unique vers lequel s'accumulent les particules est affecté par les interactions, puisqu'il occupe un volume plus grand que l'état fondamental à une particule. Mais c'est encore plus vrai pour des états hors d'équilibre où l'on peut jouer, non seulement sur l'extension spatiale de  $|\Psi(\mathbf{r})|$ , mais aussi sur sa phase. Il est commode de poser :

$$\Psi(\mathbf{r}) = |\Psi(\mathbf{r})| \exp[i\theta(\mathbf{r})] \quad (3.15)$$

où  $\theta(\mathbf{r})$  est la phase de  $\Psi(\mathbf{r})$ . Un gradient de  $\theta(\mathbf{r})$  implique la présence d'un courant de masse associé à une vitesse d'écoulement :

$$\mathbf{v}(\mathbf{r}) = (\hbar/m) \nabla\theta(\mathbf{r}) \quad (3.16)$$

On voit ainsi que les atomes peuvent se condenser vers un état stationnaire, mais non-immobile puisqu'il est le siège de courants se refermant sur

eux-mêmes. Bien sûr, il ne faut pas imposer à  $\theta(\mathbf{r})$  des variations spatiales trop rapides, impliquant des vitesses élevées, car cela correspondrait à une forte élévation d'énergie cinétique. Mais supposer des variations lentes de  $\theta(\mathbf{r})$ , à l'échelle de l'extension spatiale du récipient contenant le gaz, reste relativement sans conséquences, car l'élévation d'énergie cinétique supplémentaire par atome du condensat est très faible<sup>16</sup>. Nous allons voir que les états physiques ainsi obtenus sont, non pas l'état fondamental du système, mais des états métastables de très grande durée de vie, ce qui implique l'existence de courants permanents.

Considérons pour simplifier un gaz contenu dans un récipient en forme de tore ; supposons que, par un moyen ou un autre, on ait mis le condensat en mouvement, et que la phase  $\theta(\mathbf{r})$  de sa fonction d'onde varie de  $2\pi$  lorsque le point  $\mathbf{r}$  effectue un tour dans le tore. Le nuage thermique est libre de relaxer à l'échelle microscopique, puisqu'il a accès à de très nombreux états individuels. Il se met donc rapidement à l'équilibre par rapport aux parois du récipient par un effet habituel de viscosité, comme le ferait un gaz ordinaire. Le condensat, quant à lui, est lié par la contrainte forte d'occuper un seul état, et ce dernier est décrit par une fonction d'onde unique  $\Psi(\mathbf{r})$  dont la valeur initiale  $\Psi_i(\mathbf{r})$  possède un gradient de phase – c'est l'origine du courant circulaire de particules. Comment dans ces conditions le condensat peut-il relaxer progressivement vers un état au repos ? On pourrait envisager de prendre pour  $\Psi(\mathbf{r})$  une superposition :

$$\Psi(\mathbf{r}) = \alpha \Psi_i(\mathbf{r}) + \beta \Psi_0(\mathbf{r}) \quad (3.17)$$

où  $\Psi_0(\mathbf{r})$  est une fonction d'onde de phase constante, nulle par exemple ; en faisant tendre progressivement  $\alpha$  vers zéro et  $\beta$  vers un, tout en maintenant la somme  $|\alpha|^2 + |\beta|^2$  constante pour conserver la norme de  $\Psi(\mathbf{r})$ , on peut mathématiquement ramener le condensat vers un état immobile. Mais il faut pour cela nécessairement passer par des situations intermédiaires où ni  $\alpha$  ni  $\beta$  ne sont nuls ; il existe alors dans le tore des points où les deux fonctions d'onde sont en phase, ce qui implique un maximum de densité, et d'autres (diamétralement opposés sur le tore) où elles sont en opposition de phase, ce qui implique un minimum. Ceci reste vrai quel que soit le choix de  $\alpha$  et  $\beta$  puisque, si l'on change la phase relative de ces deux coefficients, on déplace le maximum de densité sans changer sa valeur. Or, du fait de ses interactions répulsives, le gaz est peu compressible : une telle variation de densité entraîne nécessairement une variation positive d'énergie mécanique élastique. On voit ainsi qu'il existe une barrière de potentiel répulsive macroscopique, créée par la répulsion entre atomes, qui s'oppose à une relaxation progressive de la fonction d'onde  $\Psi_i(\mathbf{r})$  vers  $\Psi_0(\mathbf{r})$ . L'état initial est donc métastable, protégé par une barrière d'énergie dont le franchissement prend un temps d'autant plus grand que le système est grand, infini à la limite des tailles infinies. On comprend ainsi l'existence de courants permanents, caractéristiques de la superfluidité.

---

<sup>16</sup>Elle tend vers zéro à la limite thermodynamique.

Une propriété importante de ces courants permanents est la quantification de leur circulation. Dans le cas que nous avons considéré, la phase  $\theta(\vec{r})$  varie de  $2\pi$  lorsque le point  $\mathbf{r}$  fait un tour dans le tore, de sorte que l'équation (3.16) permet de montrer que l'intégrale de la vitesse sur tout chemin fermé faisant un tel tour vaut  $2\pi(\hbar/m)$ . Mais nous aurions également pu supposer que la variation par tour de la phase de la fonction d'onde du condensat est égale à  $2p\pi$ , où  $p$  est un nombre entier quelconque. La circulation de la vitesse est donc quantifiée en multiples entiers de  $2\pi(\hbar/m)$ , ce qui est très différent de la situation pour un fluide classique, où elle peut varier continûment et prendre n'importe quelle valeur.

### 3.4.3 Condensat en rotation et tourbillons quantiques

Considérons maintenant un tore comme le précédent, mais de section rectangulaire au lieu de circulaire ; rien ne change évidemment aux arguments précédents. À la limite où le rayon intérieur du tore tend vers zéro, l'état métastable superfluide que nous avons décrit acquiert une singularité sur l'axe du tore, et devient ce que l'on appelle un vortex, ou tourbillon. La circulation de la vitesse autour de la singularité, souvent appelée « cœur » du tourbillon, est quantifiée en unités  $2\pi(\hbar/m)$ . Comme le récipient qui contient le superfluide n'a plus de trou intérieur, le cœur est maintenant libre de se déplacer latéralement, mais cela ne fait pas disparaître le tourbillon. On peut ensuite ajouter une seconde singularité, et donc un second tourbillon, puis d'autres. Des arguments hydrodynamiques montrent que de tels tourbillons ne sont pas indépendants, mais interagissent entre eux, de sorte qu'ils tendent pour minimiser leurs interaction à se placer sur des réseaux réguliers, appelés « réseaux d'Abrikosov ».

Pour étudier ces tourbillons dans un condensat, on utilise un « agitateur », par exemple un faisceau laser. Le condensat est confiné dans un piège magnétique circulaire et l'agitateur laser tournant va mettre ce condensat en rotation, comme une cuillère fait tourner un liquide dans une tasse. Si la rotation de l'agitateur est suffisamment rapide, on crée des vortex, qu'on observe ensuite en mesurant la répartition spatiale des atomes. Puisque les vortex correspondent à des endroits où la densité du condensat s'annule, ils apparaissent comme des trous dans le nuage atomique. Un résultat caractéristique, obtenu au laboratoire Kastler Brossel à Paris, est indiqué sur la figure 3.7.

De nombreuses études ont été menées ces dernières années sur ces vortex. Outre le caractère spectaculaire intrinsèque des photos ainsi obtenues (on « voit » directement une fonction d'onde quantique), ces tourbillons quantiques posent de nombreuses questions intéressantes et difficiles, aussi bien sur le plan physique que mathématique : quelle est la forme d'une ligne de vortex, quels sont les paramètres du réseau régulier de la figure 3.7d, que se passe-t-il quand on tourne encore plus vite ? Une voie de recherche actuellement très explorée est le lien entre ces condensats en rotation et la physique de l'effet Hall quantique, qui se produit lorsqu'on place dans un grand champ magnétique

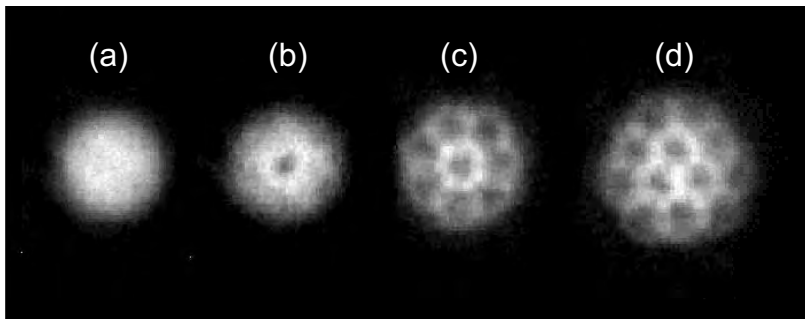


FIG. 3.7 – Profil de densité après expansion balistique d’un condensat mis en rotation avec un agitateur laser. (a) La fréquence de rotation est en dessous de la fréquence critique et rien ne se produit. (b) Juste au-dessus de la fréquence critique, le premier vortex apparaît. (c,d) La fréquence de rotation est notablement supérieure à la fréquence critique et un réseau régulier de vortex se forme (réseau triangulaire d’Abrikosov).

un matériau semi-conducteur confinant un gaz d’électrons bi-dimensionnel. Même si les ordres de grandeur sont très différents (la masse d’un atome est 100 000 fois plus grande que la masse d’un électron), les analogies entre les deux situations semblent fortes ; on peut donc espérer utiliser ces condensats en rotation pour modéliser des systèmes complexes de physique du solide, en tirant parti de l’universalité de la physique quantique à très basse température.

### 3.4.4 Du superfluide vers l’isolant de Mott

Jusqu’ici, nous nous sommes intéressés à un condensat de Bose-Einstein confiné dans un piège magnétique, bien décrit par un potentiel harmonique. Des faisceaux laser peuvent également être utilisés pour confiner les atomes. La palette des potentiels disponibles s’en trouve alors considérablement accrue. En particulier, on peut utiliser une onde lumineuse stationnaire, qui présente une répartition périodique de nœuds et de ventres d’intensité. Si la fréquence du faisceau laser est inférieure à la fréquence de résonance atomique, les atomes vont se localiser autour des zones où l’intensité lumineuse est maximale, c’est-à-dire les ventres de l’onde stationnaire. Ce « réseau optique » est un véritable cristal artificiel ; il constitue un système modèle pour étudier le problème de la conduction électrique, dans lequel un ou plusieurs électrons bougent dans le potentiel périodique créé par les atomes d’un solide.

Une expérience spectaculaire, menée à Munich, a consisté à placer un condensat de Bose-Einstein dans un réseau optique avec des nombres d’atomes et de sites similaires. Les chercheurs de Munich ont alors observé la transition entre l’état « condensat de Bose-Einstein » et l’état « isolant de Mott ».

Ce dernier est obtenu en disposant exactement un atome en chaque site du réseau périodique. C'est un isolant puisqu'aucun transport de matière n'est possible si on ne s'autorise pas à avoir au moins quelques sites doublement occupés et d'autres vides.

Pour explorer cette transition, il est essentiel de contrôler la facilité pour un atome de sauter d'un site à l'autre. Cette mobilité des atomes dépend de l'intensité lumineuse : pour une intensité faible, la barrière de potentiel entre deux sites adjacents n'est pas très haute et chaque atome peut facilement sauter d'un site à l'autre, grâce à l'effet tunnel. Dans ce cas, les interactions entre atomes jouent un rôle négligeable, et l'assemblée d'atomes reste dans l'état « condensat de Bose-Einstein », comme pour un confinement dans un piège harmonique. En particulier, l'ensemble des atomes est toujours décrit par une fonction d'onde unique, délocalisée sur l'ensemble du réseau.

Au contraire, si on augmente l'intensité lumineuse, l'effet tunnel est considérablement réduit. Les sauts entre sites deviennent plus difficiles et l'interaction entre deux atomes occupant le même site devient comparativement plus importante. Or, cette interaction a un prix : il est plus coûteux d'avoir deux atomes sur le même site et un site voisin vide, que d'avoir un atome dans chacun des deux sites. Plus précisément, il existe un seuil au-delà duquel il n'est plus favorable d'un point de vue énergétique de sauter d'un site à l'autre (voir Fig. 3.8). Au-delà de ce seuil, la cohérence collective du gaz d'atomes

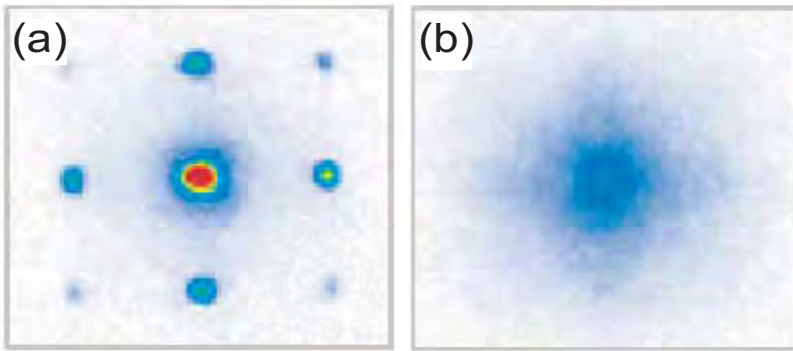


FIG. 3.8 – Observation de la transition condensat de Bose-Einstein-isolant de Mott pour un gaz d'atomes placé dans le potentiel périodique créé par un réseau optique. On mesure la distribution spatiale des atomes après expansion ballistique. La figure (a) correspond à la phase condensat, obtenue pour une faible valeur du potentiel périodique de confinement. Les différents pics d'interférence révèlent la cohérence de la fonction d'onde atomique sur plusieurs sites du réseau. La figure (b) correspond à la phase isolant de Mott, obtenue pour des sites deux fois plus profonds qu'en (a). Le réseau est alors rempli avec exactement un atome par site et la cohérence du gaz d'atomes sur l'ensemble du réseau est perdue (photographies fournies par I. Bloch, Munich).

disparaît, et la règle du « chacun chez soi » l'emporte : on a produit un isolant de Mott. Ce régime, qui comporte un nombre défini d'atomes sur chaque site du réseau, ouvre des perspectives très prometteuses dans le domaine du stockage et du traitement de l'information quantique.

### 3.5 Conclusion et perspectives

L'étude de la condensation de Bose-Einstein dans les gaz est ainsi devenue l'un des domaines de recherche les plus actifs en physique. Pour donner une idée du niveau de cette activité, mentionnons simplement qu'elle se traduit par la publication d'environ un millier d'articles par an. Les raisons d'un tel intérêt pour les gaz de bosons condensés sont simples à comprendre. L'effet des interactions entre atomes qui joue, comme nous l'avons vu plus haut, un rôle si important dans des effets quantiques macroscopiques comme la superfluidité, est beaucoup plus simple à décrire et à interpréter pour des systèmes gazeux que pour des phases liquides ou solides. On peut dans la plupart des cas se limiter à des interactions à deux corps (collisions élastiques entre atomes, bien séparées dans le temps), ce qui est évidemment impossible pour des systèmes plusieurs milliards de fois plus denses, comme l'hélium liquide. Nous avons mentionné plus haut (paragraphe 3.4.1) les simplifications de la description des collisions dans un gaz dilué à basse température, avec l'existence d'un seul paramètre pertinent, la longueur de diffusion, et la possibilité de le varier en utilisant des résonances de Feshbach. Il a été ainsi possible au cours des dernières années d'explorer de très nombreuses situations physiques différentes et des phénomènes d'une grande diversité. Les échanges constants entre théorie et expérience ont permis que s'élabore progressivement une compréhension plus satisfaisante de la physique des systèmes de bosons condensés. Il est incontestable que nous avons maintenant une vue beaucoup plus claire du rôle des interactions dans ces systèmes. Le phénomène de condensation prédit par Einstein pour un gaz parfait s'est finalement révélé intéressant essentiellement pour l'étude des effets nouveaux auxquels donnent naissance les interactions.

La moisson de résultats obtenus sur les gaz d'atomes bosoniques a bien sûr stimulé des études analogues sur les gaz d'atomes fermioniques. La statistique de Fermi-Dirac conduit à une répartition tout à fait différente des atomes dans les niveaux d'énergie du piège qui contient le gaz (supposé parfait) de fermions. Chaque état quantique ne peut contenir au maximum qu'un seul fermion. À température nulle, les  $N$  fermions remplissent les  $N$  premiers états d'énergie du piège. L'énergie  $E_F$  du niveau le plus excité qui est occupé est appelée énergie de Fermi et permet de définir la température de Fermi  $T_F$  par la relation  $E_F = k_B T_F$  (voir Fig. 3.9). Un gaz de fermions est dit dégénéré si sa température  $T$  est très inférieure à la température de Fermi  $T_F$ . La transition entre niveaux occupés et niveaux vides ne se fait plus brusquement comme

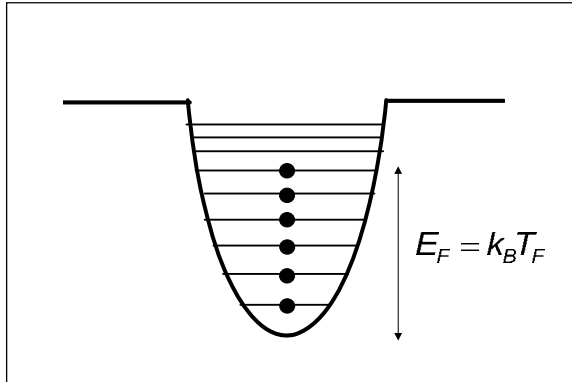


FIG. 3.9 – À température nulle  $T = 0$ , les atomes d'un gaz parfait de fermions occupent, à raison d'un seul atome par état, les niveaux d'énergie les plus bas du piège qui les contient. L'énergie du dernier niveau occupé est l'énergie de Fermi  $E_F$  qui permet de définir la température de Fermi  $T_F$  par la relation  $E_F = k_B T_F$ .

sur la figure 3.9, mais sur un intervalle d'énergie de l'ordre de  $k_B T$ , beaucoup plus petit que  $E_F$ .

Les premiers travaux réalisés sur des gaz de fermions ont essayé d'atteindre ce régime de dégénérescence en refroidissant les atomes à des températures aussi basses que possible. Le refroidissement laser peut bien sûr être utilisé pour les fermions mais, comme pour les bosons, il ne permet pas à lui seul d'atteindre le régime de dégénérescence quantique. C'est là qu'une difficulté nouvelle surgit quand on essaie d'étendre aux fermions le refroidissement par évaporation, si efficace pour les bosons. Les fermions sont polarisés dans le piège magnétique qui les contient et le principe de Pauli interdit à deux fermions dans le même état de spin de s'approcher trop près l'un de l'autre. Leur distance minimale d'approche est de l'ordre de la longueur d'onde de de Broglie, beaucoup plus grande que la portée du potentiel d'interaction entre atomes. Les interactions entre fermions identiques polarisés, et en particulier les collisions élastiques essentielles pour l'évaporation, sont donc inhibées par le principe de Pauli. Une solution ingénieuse a été trouvée pour résoudre cette difficulté. Elle consiste à mélanger le gaz de fermions avec un gaz de bosons, qui peuvent être refroidis par évaporation, et à utiliser les interactions entre les deux gaz pour refroidir les fermions par contact thermique avec les bosons ultrafroids<sup>17</sup>. Une telle méthode de refroidissement, déjà utilisée pour les ions piégés, est appelée « refroidissement sympathique ». Elle a permis d'atteindre des degrés de dégénérescence caractérisés par des valeurs de  $T/T_F$  aussi basses que 0,05.

<sup>17</sup>Il est également possible aussi d'utiliser des mélanges de fermions dans des états de spin différents, pour lesquels les collisions élastiques entre 2 fermions dans des états de spin différents ne sont pas inhibées et permettent l'évaporation.

Contrairement aux bosons, les fermions ne subissent pas une transition de phase spectaculaire quand ils sont dégénérés. En revanche, les interactions entre fermions ultrafroids de spins opposés donnent naissance à des effets nouveaux particulièrement intéressants. Les fermions de spins opposés peuvent s'apparier et les paires de fermions ainsi obtenues, qui sont des bosons, peuvent former des condensats de Bose-Einstein pour des températures suffisamment basses. Un premier exemple de telles études concerne les molécules formées à partir de deux fermions de spins opposés lorsque la longueur de diffusion est réglée à une valeur positive au moyen d'une résonance de Feshbach. Dans ce cas, il existe un état lié peu profond dans le puits de potentiel décrivant l'interaction entre les deux atomes, puits de potentiel dans lequel, sous l'effet de collisions à trois corps, les deux atomes peuvent tomber pour former une molécule faiblement liée. La bonne surprise est que les molécules ainsi formées à partir de deux fermions de spins opposés sont beaucoup plus stables que les molécules analogues formées à partir de deux bosons. L'origine d'une telle stabilité est, là encore, liée au principe de Pauli qui ne permet pas à un autre fermion, soit libre, soit lié dans une autre molécule, de s'approcher suffisamment de la molécule considérée pour la faire relaxer vers un état plus profond dans le puits de potentiel. En effet, quel que soit son spin, ce fermion trouvera toujours dans la molécule dont il essaie de s'approcher un autre fermion de même spin que lui. Cette réduction des collisions explique la grande durée de vie des molécules, et par suite la possibilité de les obtenir avec des densités élevées. Comme ces molécules, formées à partir de fermions ultrafroids, sont elles-mêmes ultrafroides, le seuil de condensation de Bose-Einstein a pu être atteint par plusieurs équipes. Ainsi, c'est à partir de travaux portant sur des fermions que les premiers condensats moléculaires gazeux ont pu être réalisés !

Une autre situation très étudiée actuellement concerne des gaz de fermions de spins opposés pour lesquels la longueur de diffusion est réglée à une valeur négative au moyen d'une résonance de Feshbach. Dans ce cas, il n'existe pas d'état lié peu profond dans le puits de potentiel d'interaction des deux atomes, mais les interactions effectives entre atomes sont attractives. Les atomes fermioniques s'apparient alors dans l'espace des impulsions. Les paires obtenues sont analogues aux paires de Cooper formées par deux électrons dans un métal sous l'effet d'une faible attraction due à l'échange de phonons. On sait que ces paires de Cooper peuvent, en dessous d'une certaine température critique, donner naissance à une phase condensée superfluide, à l'origine de la supraconductivité. C'est la transition BCS, voir note (9). De nombreux problèmes peuvent être étudiés dans le cadre des recherches actuelles sur les fermions ultrafroids. Existe-t-il une transition de type BCS pour les gaz d'atomes fermioniques ? Comment la caractériser ? Peut-on, en balayant la résonance de Feshbach et en passant continûment d'une valeur positive à une valeur négative de la longueur de diffusion, étudier la transition continue entre un condensat moléculaire et un superfluide de type BCS ? Dans la région centrale de la résonance de Feshbach, le module de la longueur de diffusion devient infini. Quelles approches théoriques peut-on utiliser pour analyser ce problème à  $N$  corps en



interaction forte ? Quels effets nouveaux peut-on espérer observer ? Toutes ces questions illustrent à quel point l'article d'Einstein de 1925 se révèle fécond par les développements auxquels il donne naissance 80 ans après sa parution.

### 3.6 Références générales

Pour un sujet aussi vaste, il semble impossible de donner une liste complète de références, qui serait bien trop longue. Mentionnons cependant quelques ouvrages généraux contenant eux-mêmes un grand nombre de références utiles :

C. Pethick, H. Smith, *Bose-Einstein condensation in dilute Bose gases*, Cambridge University Press (2002).

L. Pitaevskii, S. Stringari, *Bose-Einstein condensation*, Clarendon Press (2003).

J. Matricon, G. Waysand, *La guerre du froid*, Seuil (1994).

H.J. Metcalf, P. Van Der Straten, *Laser Cooling and Trapping*, Springer-Verlag (1999).

Conférences Nobel 1997 :

S. Chu, *The manipulation of neutral particles*, Rev. Mod. Phys. **70**, 685–706 (1998).

C. Cohen-Tannoudji, *Manipulating atoms with photons*, Rev. Mod. Phys. **70**, 707–719 (1998).

W. Phillips, *Laser cooling and trapping of neutral atoms*, Rev. Mod. Phys. **70**, 721–741 (1998).

Conférences Nobel 2001 :

E.A. Cornell, C.E. Wieman, *Bose-Einstein condensation in a dilute gas, the first 70 years and some recent experiments*, Rev. Mod. Phys. **74**, 875–893 (2002).

W. Ketterle, *When atoms behave as waves: Bose-Einstein condensation and the atom laser*, Rev. Mod. Phys. **74**, 1131–1151 (2002).

Voir également :

C. Cohen-Tannoudji, cours au Collège de France :  
<http://www.phys.ens.fr/cours/college-de-france/> cours des 5 années de 1997 à 2002.

À ces ouvrages, nous ajoutons ci-dessous une liste d'articles, soit pour leur caractère historique, soit parce qu'ils ne se trouvent pas référencés dans les ouvrages généraux ci-dessus.

- [1] S.N. Bose, *Z. Physik* **26**, 178 (1924).
- [2] A. Einstein, Preussische Akademie der Wissenschaften, Phys.-Math. Klasse, Sitzungsberichte, p. 261–267 (1924) et 3–14 (1925). Traduction française dans *Albert Einstein, oeuvres choisies*, tome 1, Quanta, F. Balibar, O. Darrigol et B. Jech eds., Editions du Seuil et CNRS, « Théorie quantique du gaz parfait monotaomique », pp. 172–179 et « Théorie quantique du gaz parfait », pp. 180–192 (1989).
- [3] W.A. Blanpied, *Satyendranath Bose: co-founder of quantum statistics*, *Am. J. Phys.* **40**, 1212–1220 (1972).
- [4] A. Pais, *Albert Einstein, la vie et l'oeuvre*, Interéditions (1993); voir particulièrement § 23.d.
- [5] A. Kastler, *On the historical development of the indistinguishability concept for microparticles* dans *Old and New Questions in Physics, Cosmology, Philosophy and Theoretical Biology*, van der Merve (ed.), New York, Plenum Press, p. 607 (1983).
- [6] G.E. Uhlenbeck, *Proceedings Einstein Centennial Symposium 1979* (H. Woolf dir.), Reading (Mass.), Addison Wesley (1980).
- [7] P. Nozières, M. Le Bellac, communications privées.
- [8] A. Mysyrowicz, *Bose-Einstein condensation of excitonic particles in semiconductors* dans *Bose-Einstein condensation*, edited by A. Griffin, D.W. Snoke, S. Stringari, Cambridge University Press, p. 330 (1995).
- [9] C.E. Hecht, *The possible superfluid behaviour of hydrogen atom gases and liquids*, *Physica* **25**, 1159–1161 (1959).
- [10] W.C. Stwalley, L.H. Nosanow, *Possible “new” quantum system*, *Phys. Rev. Lett.* **36**, 910–913 (1976).
- [11] I.F. Silvera, J.T.M. Walraven, *Stabilization of atomic hydrogen at low temperatures*, *Phys. Rev. Lett.* **44**, 164–168 (1980). Pour une discussion plus générale voir, par les mêmes auteurs, « Spin polarized atomic hydrogen », chap. 3 in *Progress in Low temperature physics*, D.F. Brewer ed., vol. X, North Holland (1986).
- [12] T.J. Greytak, D. Kleppner, *Lectures in spin-polarized hydrogen*, Les Houches summer school 1282, session XXXVIII, p. 1129, G. Grynberg and R. Stora eds., Elsevier (1984).
- [13] J.T.M. Walraven, I.F. Silvera, *The magnetization and density of spin polarized atomic hydrogen*, *J. Physique Coll.* **C7**, 147–149 (1980).
- [14] V.V. Goldman, I.F. Silvera, A.J. Leggett, *Atomic hydrogen in an inhomogeneous magnetic field: density profile and Bose-Einstein condensation*, *Phys. Rev.* **B24**, 2870–2873 (1981).
- [15] H. Hess, *Evaporative cooling of magnetically trapped and compressed spin-polarized atomic hydrogen*, *Phys. Rev.* **B34**, 3476–3479 (1986).
- [16] D.G. Fried, T.C. Killian, L. Willmann, D. Landhuis, S.C. Moss, D. Kleppner, T.J. Greytak, *Bose-Einstein condensation of atomic hydrogen*, *Phys. Rev. Lett.* **81**, 3811–3814 (1998).
- [17] V.J. Emery, *Spin diffusion in gases at low temperatures*, *Phys. Rev.* **133A**, 661–664 (1964).

- [18] C. Lhuillier, F. Laloë, *Transport properties in a spin polarized gas I and II*, J. Physique **43**, 197–224 et 225–241 (1982).
- [19] C. Lhuillier, F. Laloë, *L'hélium trois polarisé, un « nouveau » fluide quantique ?*, J. Physique **40**, 239–256 (1979).
- [20] M. Leduc, P.-J. Nacher, D. Betts, J. Daniels, G. Tastevin, F. Laloë, *Europhys. Lett.* **4**, 59 (1987). D.S. Betts, F. Laloë, M. Leduc, *Properties of strongly spin polarized Helium 3 gas*, Prog. Low Temp. Phys. vol XII, Elsevier (1989).
- [21] B.R. Johnson, J.S. Denker, N. Bigelow, L.P. Levy, J.H. Freed, D.M. Lee, *Observation of nuclear spin waves*, Phys Rev. Lett **52**, 1508–1511 (1984).
- [22] P.J. Nacher, G. Tastevin, M. Leduc, S.B Crampton, F. Laloë, *Spin rotation effects and spin waves in gaseous  $^3\text{He}$* , J. Phys. Lett. **45**, L-441–448 (1984).
- [23] M. Holzmann, J.N. Fuchs, G.A Baym, J.P. Blaizot, F. Laloë, *Bose-Einstein transition temperature in a dilute repulsive gas*, C.R. Physique **5**, 21–37 (2004).



# Chapitre 4

## De l'émission stimulée d'Einstein aux lasers aujourd'hui

Par *Jean-Paul Pocholle*

### Sommaire

<b>4.1</b>	<b>Introduction</b>	<b>131</b>
<b>4.2</b>	<b>Bref historique</b>	<b>132</b>
<b>4.3</b>	<b>Interaction d'un rayonnement électromagnétique avec un système atomique</b>	<b>133</b>
4.3.1	Coefficients d'Einstein	133
4.3.2	Description sommaire d'un laser	136
4.3.3	Systèmes à trois et à quatre niveaux	138
<b>4.4</b>	<b>Description de quelques réalisations de laser</b>	<b>141</b>
4.4.1	Diodes laser et génie quantique	141
4.4.2	Diodes laser à cascade quantique (QCL : Quantum Cascade Laser)	147
4.4.3	Rayonnement TeraHertz (THz)	150
4.4.4	Le renouveau des lasers solides	153
4.4.5	Laser femtosecondes	156
4.4.6	Le laser et l'optique non linéaire	159
<b>4.5</b>	<b>Applications</b>	<b>163</b>
4.5.1	Le domaine de la communication et de l'information	163
4.5.2	Le domaine de l'instrumentation	170
4.5.3	Lévitiation optique	181
4.5.4	Le domaine de l'énergie	182
4.5.5	Le laser et ses applications industrielles	185
<b>4.6</b>	<b>Conclusions</b>	<b>191</b>
<b>4.7</b>	<b>Références</b>	<b>196</b>



### **Jean-Paul Pocholle**

Jean-Paul Pocholle est actuellement expert scientifique senior à THALES Research and Technology, après avoir dirigé le laboratoire « Sources pour Optronique » au Laboratoire Central de Recherche de THOMSON-CSF. Ses activités de recherche appliquée l'ont conduit à déposer 70 brevets dans les domaines de l'optique guidée, des capteurs optoélectroniques et des structures à puits quantiques. Il a fait progresser la technologie des lasers solides et à semi-conducteurs, l'optique non linéaire et la physique des plasmas photo-induits.

# De l'émission stimulée d'Einstein aux lasers aujourd'hui

## 4.1 Introduction

L'apport conceptuel d'Albert Einstein à l'interaction lumière-matière est essentiel. Son approche, qui a consisté à introduire la notion d'absorption et d'émission stimulée du rayonnement, est à l'origine du processus d'émission laser, si l'on y ajoute l'inversion de population de milieux très divers soumis à une excitation.

En analysant les conditions d'équilibre thermique dans l'interaction du rayonnement électromagnétique avec la matière, Einstein a compris que la prise en compte de l'émission spontanée permet seulement de trouver la loi de répartition spectrale de Wien. En revanche la loi de Planck [1] ne peut être obtenue que si on postule l'existence d'un processus d'émission stimulée. C'est en cela que les travaux d'Einstein contiennent en germe le développement des sources de rayonnement électromagnétique cohérentes à partir de la parution de sa publication en 1917 [2].

En effet, les premières démonstrations d'une émission cohérente de rayonnement ont associé les spécialistes en physique atomique à ceux de l'électromagnétisme. Ceci a donné naissance au MASER (*Microwave Amplification by Stimulated Emission of Radiation*). C'est la conjugaison des travaux menés sur le pompage optique au Laboratoire de Spectroscopie Hertzienne de l'École Normale Supérieure par Alfred Kastler en 1949 [3] avec ceux plus anciens de Charles Fabry et Alfred Perot sur l'interférométrie à ondes multiples qui a permis à Arthur L. Schawlow et Charles H. Townes dès 1957 de réaliser les conditions d'obtention d'une émission cohérente de lumière. C'est dans une publication de 1958 [4] qu'ils définissent le schéma type d'un LASER (*Light Amplification by Stimulated Emission of Radiation*), tout en prédisant la création de rayonnements cohérents dans l'infrarouge et dans le visible. Cette date est à l'origine de multiples et incessantes avancées spectaculaires tant au niveau du développement des sources laser qu'à celles liées aux applications.

C'est l'évolution de cet outil aux multiples facettes et la réponse qu'il a su apporter à des problèmes posés qui ont contribué au développement de l'opto-électronique, de l'optronique et maintenant de la bio-photonique. Sur le plan

plus fondamental, c'est l'existence de cet instrument ou plus exactement de cette source de rayonnement lumineux qui présente un caractère non classique liée au processus stimulé de l'émission, qui a permis l'éclosion en particulier :

- de la physique des atomes ultra-froids ;
- de l'optique quantique ;
- de la physique de l'interaction lumière-matière sous éclairage intense avec les sources laser femtosecondes . . .

domaines qui sont aujourd'hui en pleine effervescence, en particulier en France.

## 4.2 Bref historique

De 1952 à 1954, l'emploi d'un milieu avec inversion de population inclus dans un résonateur hyperfréquence à haut facteur de qualité (fort coefficient de surtension  $Q$ ) fut proposé aux États-Unis et en URSS afin de réaliser de nouveaux standards de fréquence fondés sur l'emploi d'oscillateurs quantiques actifs (QFS : Quantum Frequency Standards). Quasi-simultanément, en 1954, N.G. Basov et A.M. Prokhorov [5] en URSS et J.P. Gordon, H.J. Ziger et C.H. Townes [6] aux États-Unis proposèrent l'amplification stimulée à partir de la molécule d'ammoniac  $\text{NH}_3$ . Le MASER était né [7]. Une période faste voyait se développer la spectroscopie micro-onde et une intense activité dans le domaine des standards de fréquence passif (faisceau atomique de césium, cellule de rubidium . . .) et actif (Maser  $\text{NH}_3$ , H . . .).

C'est en 1960 que la première émission laser fut annoncée dans une publication [8]. Cette émission impulsionnelle résulte du pompage optique d'un cristal de rubis (barreau  $\text{Cr}^{3+} : \text{Al}_2\text{O}_3$ ) par l'intermédiaire d'un flash hélicoïdal qui l'enveloppe. Ensuite, en quelques années, on observe une explosion du nombre de publications qui établissent les fondements de l'électronique quantique et qui décrivent des réalisations expérimentales qui vont constituer les prémices d'une activité industrielle.

Dans ce cadre, nous pouvons citer :

- la démonstration d'une émission laser en continu à partir d'un mélange gazeux He-Ne par A. Javan [9] ;
- l'émission laser à partir de matrices vitreuses dopées par l'ion néodyme et sous forme de fibre laser par E. Snitzer en 1961 [10] ;
- le déclenchement par modification du facteur de qualité de la cavité laser ( $Q$ -switch) par R. Hellwarth en 1961 [11] ;
- l'émission à 633 nm du laser He-Ne par White et Rigden en 1962 ;



- la proposition d'employer un pompage optique de matrices solides dopées à partir de composés semiconducteurs [12]. On doit noter qu'à cette époque, le concept d'hétérostructure était inconnu et l'injection de porteurs dans la micro plaquette de composés GaAs s'effectuait par l'intermédiaire d'un canon à électrons ;
- le laser CO<sub>2</sub> par K. Patel en 1964 [13] ;
- l'observation d'une émission laser à partir d'un cristal Nd:YAG [14] ;
- la démonstration d'interactions non linéaires et la description phénoménologique des processus d'optique non linéaire [15–17] ;
- les laser ioniques par B. Bridges en 1964 et à colorants par P. Sorokin en 1966 . . .

Toute cette histoire peut être explorée en consultant l'ouvrage de Mario Bertolotti [18].

## 4.3 Interaction d'un rayonnement électromagnétique avec un système atomique

### 4.3.1 Coefficients d'Einstein

Ce paragraphe présente de manière simple les relations qui ont conduit Albert Einstein à raccorder les processus d'émission et d'absorption stimulées à la théorie du rayonnement du corps noir établie par Max Planck [1].

La description élémentaire de l'interaction d'une onde électromagnétique avec des atomes peut être formulée en considérant un système à deux niveaux (Fig. 4.1). Soit un nombre total  $N$  d'atomes qui se répartissent selon les populations  $N_1$  et  $N_2$  respectivement sur deux niveaux d'énergie  $E_1$  et  $E_2$ .

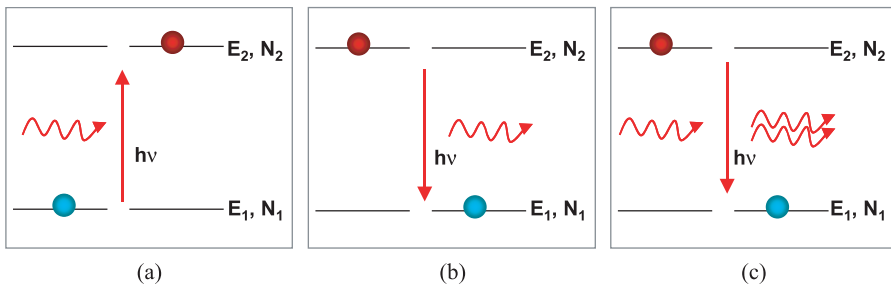


FIG. 4.1 – *Système à deux niveaux en interaction avec un rayonnement électromagnétique. (a) absorption stimulée, (b) émission spontanée, (c) émission stimulée.*

On observe une absorption résonnante du rayonnement électromagnétique<sup>1</sup> lorsque la fréquence  $\nu$  de ce rayonnement est égale à la différence d'énergie entre ces deux niveaux. L'énergie émise ou absorbée est alors liée à la relation :

$$E_2 - E_1 = h\nu \quad (4.1)$$

$h$  est la constante de Planck.

Le taux d'absorption dépend de la densité spectrale d'énergie du champ électromagnétique incident  $\rho(\nu)$ , de la population  $N_1$  du niveau inférieur de la transition et il est proportionnel à un facteur  $B_{12}$  qui traduit les propriétés du système atomique :

$$\left. \frac{dN_2}{dt} \right|_{\text{abs}} = N_1 B_{12} \rho(\nu) = - \left. \frac{dN_1}{dt} \right|_{\text{abs}} \quad (4.2)$$

De la même manière, on introduit le concept d'émission stimulée (induit par la densité d'énergie du champ incident) qui est analogue au processus d'absorption. En présence d'un rayonnement de même fréquence que celle de la transition, le système présente une certaine probabilité d'être désexcité pour retrouver son état fondamental ou initial :

$$\left. \frac{dN_2}{dt} \right|_{\text{stim}} = -N_2 B_{21} \rho(\nu) = - \left. \frac{dN_1}{dt} \right|_{\text{stim}} \quad (4.3)$$

Dans ce processus, le rayonnement émis présente les mêmes caractéristiques que le rayonnement incident qui stimule l'opération de désexcitation (fréquence, phase, état de polarisation ...). L'émission est dite cohérente.

À ces deux mécanismes se superpose l'émission spontanée qui est intrinsèque. Elle traduit le fait qu'un système excité revient à son état initial en absence de rayonnement avec un temps caractéristique : la durée de vie de l'état excité. Le taux de désexcitation est proportionnel à la population du niveau supérieur de la transition.

$$\left. \frac{dN_2}{dt} \right|_{\text{spont}} = -N_2 A_{21} \quad (4.4)$$

L'émission spontanée est un phénomène aléatoire en direction et en phase. Elle présente un caractère incohérent. Le coefficient  $A_{21}$  correspond au taux de transition spontanée entre les niveaux 2 et 1 alors que les coefficients  $B_{12}$  et  $B_{21}$  correspondent respectivement à l'absorption et à l'émission stimulées.

La conservation du nombre total d'atomes  $N_{\text{tot}} = N_1(t) + N_2(t)$  en interaction avec le rayonnement se traduit par la condition cinétique :  $\frac{dN_2}{dt} + \frac{dN_1}{dt} = 0$ , ce qui revient à écrire :

$$\rho(\nu) B_{12} N_1 = \rho(\nu) B_{21} N_2 + A_{21} N_2$$

---

<sup>1</sup>Rappelons qu'à l'époque le concept de photon n'existait pas, la notion de quanta de rayonnement ou de lumière était alors employée. Le terme photon a été introduit en 1926 par Gilbert N. Lewis (« The Conservation of Photons », *Nature* **118**, part 2, Dec. 18, 1926, p. 874).

soit :

$$\frac{N_2}{N_1} = \frac{B_{12}\rho(\nu)}{A_{21} + B_{21}\rho(\nu)} \quad (4.5)$$

La population des niveaux à l'équilibre thermodynamique suit une loi de distribution de Boltzmann donnée par la relation :

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} e^{-\frac{E_2 - E_1}{k_B T}} = \frac{g_2}{g_1} e^{-\frac{h\nu}{k_B T}} \quad (4.6)$$

$k_B$  est la constante de Boltzmann,  $g_j$  décrit la dégénérescence de l'état  $j$  ( $j = 1, 2$ ) et  $T$  est la température absolue (K). Comme  $E_2 > E_1$ , à l'équilibre thermique nous avons  $N_2 < N_1$ . Ainsi, pour un rayonnement de longueur d'onde  $\lambda = \frac{c}{\nu} = 1 \mu\text{m}$ , où  $c$  est la vitesse de la lumière, à la température ordinaire :  $T = 300 \text{ K}$ , en considérant  $g_2 = g_1$ , nous obtenons :  $\frac{N_2}{N_1} = e^{-\frac{hc}{\lambda k_B T}} = 1,48 \times 10^{-21}$ . En revanche, dans le domaine des ondes millimétriques, ce rapport est proche de l'unité.

En combinant les expressions (4.5) et (4.6), qui déterminent le rapport entre les populations des deux niveaux, nous obtenons une relation simple entre la densité spectrale de rayonnement et les probabilités de transitions :

$$\rho(\nu) = \frac{A_{21}}{\frac{g_2}{g_1} B_{12} e^{\frac{h\nu}{k_B T}} - B_{21}} \quad (4.7)$$

La comparaison de cette expression avec la loi de Planck, qui décrit la densité de rayonnement à l'équilibre dans une cavité de dimensions finies :

$$\rho(\nu)d\nu = \frac{8\pi\nu^2 d\nu}{c^3} \frac{h\nu}{e^{\frac{h\nu}{k_B T}} - 1} \quad (4.8)$$

permet d'obtenir une relation entre les coefficients  $A_{21}$ ,  $B_{12}$  et  $B_{21}$ . La relation de Planck, stipule que la densité de rayonnement par intervalle de fréquence correspond à la densité volumique de mode de rayonnement du corps noir  $\frac{8\pi\nu^2}{c^3}$ ,  $h\nu$  est l'énergie moyenne d'un quantum de rayonnement et  $\frac{1}{e^{\frac{h\nu}{k_B T}} - 1}$  représente la population des modes.

À partir des expressions (4.7) et (4.8) nous obtenons :

$$\frac{g_2}{g_1} = \frac{B_{12}}{B_{21}} \quad (4.9)$$

et :

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h}{c^3} \nu^3 \quad (4.10)$$

Ces coefficients sont dénommés les **coefficients d'Einstein**. C'est l'introduction des concepts d'émissions induite et spontanée qui permettent de décrire correctement, selon une approche semi-classique, l'interaction du champ électromagnétique avec l'atome [19].

En effet, faisons l'hypothèse qu'il n'existe que l'absorption stimulée et l'émission spontanée. Dans ces conditions, la distribution des populations s'écrit :

$$\frac{N_2}{N_1} = \frac{B_{12}\rho(\nu)}{A_{21}} = \frac{B_{12}}{A_{21}} \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{\frac{h\nu}{k_B T}} - 1}$$

qui est en contradiction avec la loi de Boltzmann. En incluant le processus d'émission stimulée, la statistique de Boltzmann est satisfaite.

Les coefficients d'Einstein sont associés à des paramètres mesurables : une mesure du spectre d'absorption, lié au coefficient  $B_{12}$  donne une information sur le coefficient d'émission stimulée  $B_{21}$  (Éq. (4.9)). Le taux d'émission spontanée  $A_{21}$  correspond au temps de vie d'un état excité  $A_{21} = 1/\tau_{\text{spont}}$  ce qui permet d'écrire :

$$B_{21} = \frac{c^3}{8\pi h\nu^3} \frac{1}{\tau_{\text{spont}}} \quad (4.11)$$

Cette description est bien entendu réductrice car les états d'un système atomique ne sont pas aussi simples. En particulier, les niveaux se décomposent souvent en sous-niveaux et sont soumis à un élargissement spectral des résonances sous l'effet :

- du décalage Doppler, de la pression et des collisions associées dans les gaz ;
- du champ électrique local dans les solides (champ cristallin) ;
- de la température ;
- des interactions avec le milieu ou entre atomes et ions actifs.

### 4.3.2 Description sommaire d'un laser

Pour un système en équilibre thermique, le niveau le plus bas d'une transition est toujours plus peuplé que le niveau supérieur. En conséquence, le milieu se comporte comme un absorbant en présence d'un rayonnement incident de même fréquence que celle de la transition entre les deux états. En revanche, si cet équilibre est modifié en sorte que le niveau supérieur soit nettement plus peuplé que le niveau inférieur, le système favorise le processus d'émission stimulée. On obtient ainsi un mécanisme d'amplification optique, qui est associé à l'inversion de population. Cette inversion est réalisée par l'intermédiaire d'un pompage optique que l'on doit à Alfred Kastler et à Jean Brossel.

Un laser est ainsi constitué en général :

- d'un milieu actif qui joue le rôle d'un amplificateur optique. Ce milieu peut être un gaz d'atomes, de molécules, des ions métalliques de transition ou de terres rares, un matériaux semiconducteurs, un faisceau d'électrons ... ;

- d'une source d'énergie primaire qui permet d'exciter le milieu actif et de réaliser en son sein une inversion de population par pompage optique (obtenu avec par une lampe, un flash ou un laser dit de pompe). L'inversion peut aussi résulter d'une excitation électrique (par décharge et création d'un plasma ou par injection de courant dans une jonction de matériaux semiconducteurs) ou encore d'une réaction chimique ou pyrotechnique ;
- d'une cavité résonnante linéaire ou en anneau dans laquelle est placé le milieu actif qui participe à l'amplification du rayonnement électromagnétique.

Généralement, cette cavité se compose de deux miroirs parallèles qui réfléchissent les ondes lumineuses qui interagissent avec le milieu actif par aller et retour successifs dans la cavité. Cette cavité se comporte comme un interféromètre à ondes multiples qui présente une réponse spectrale sélective sous la forme d'un peigne de fréquences régulièrement espacées. Un de ses miroirs présente un caractère semi-réfléchissant afin d'extraire une partie du rayonnement lumineux présent dans la cavité. Il impose le caractère unidirectionnel de l'émission. Ce rayonnement est émis sous la forme d'un faisceau dont la divergence dépend de la longueur de cavité et des rayons de courbure des miroirs employés.

L'effet de filtrage spectral de la cavité est la conséquence du mécanisme d'interférence constructive. La distance  $L_{\text{cav}}$  entre les deux miroirs de cavité doit être un multiple entier non nul de la longueur d'onde  $\lambda$  du rayonnement émis par le système actif.

$$L_{\text{cav}} = m \frac{\lambda}{2} \quad (4.12)$$

Cette relation traduit le fait que l'on doit avoir un déphasage  $m$  fois  $2\pi$  ( $m$  entier) sur un aller et retour dans la cavité soit :

$$\Delta\varphi = \frac{\omega_m}{c} 2L_{\text{cav}} = m2\pi$$

avec  $\omega_m = 2\pi\nu_m = \frac{2\pi c}{\lambda_m}$ . On attribue cette résonance longitudinale à un mode de la cavité vide. Si la distribution spectrale du gain est large, différents modes longitudinaux peuvent se développer. L'écart en fréquence entre deux modes longitudinaux adjacents est constant :

$$\Delta\nu = \nu_{m+1} - \nu_m = \frac{c}{2L_{\text{cav}}} \quad (4.13)$$

Pour obtenir une source monofréquence on peut jouer sur la distribution spectrale du gain par introduction de pertes sélectives (filtres biréfringents) ou en diminuant la longueur de cavité puisque l'écartement en fréquence est inversement proportionnel à  $L_{\text{cav}}$ .

En tenant compte de ces éléments, en l'absence d'un signal injecté, c'est à partir du bruit que se développe l'émission laser. Ce bruit est associé à l'émission spontanée qui, par nature, est incohérente temporellement et spatialement. Seuls, parmi ces photons spontanés, ceux pouvant se coupler à un mode spatial de la cavité laser participent à l'initiation du processus d'émission stimulée. Celle-ci, par multiples passages dans la cavité, contribue à l'obtention d'un gain d'amplification conséquent. Lorsque le gain, qui dépend du taux de pompage, est supérieur aux pertes sur un aller et retour dans la cavité, le seuil d'émission laser est franchi. Nous allons illustrer comment les coefficients d'Einstein peuvent être appliqués à un système atomique afin de définir les conditions d'apparition d'une inversion de population.

### 4.3.3 Systèmes à trois et à quatre niveaux

On discerne deux classes principales de systèmes atomiques donnant lieu à l'effet laser :

- le système à trois niveaux (Fig. 4.2a) qui décrit par exemple les mécanismes de transfert et les transitions des ions  $\text{Cr}^{3+} : \text{Al}_2\text{O}_3$  (rubis dont la couleur est liée à la présence d'ions  $\text{Cr}^{3+}$  substitués à des atomes Al) ou  $\text{Er}^{3+}$  dans les cristaux ou les verres ;
- le système à quatre niveaux (Fig. 4.2b) auquel est associé le mécanisme de transfert dans le mélange gazeux He-Ne ou de l'ion  $\text{Nd}^{3+}$  dans une matrice solide.

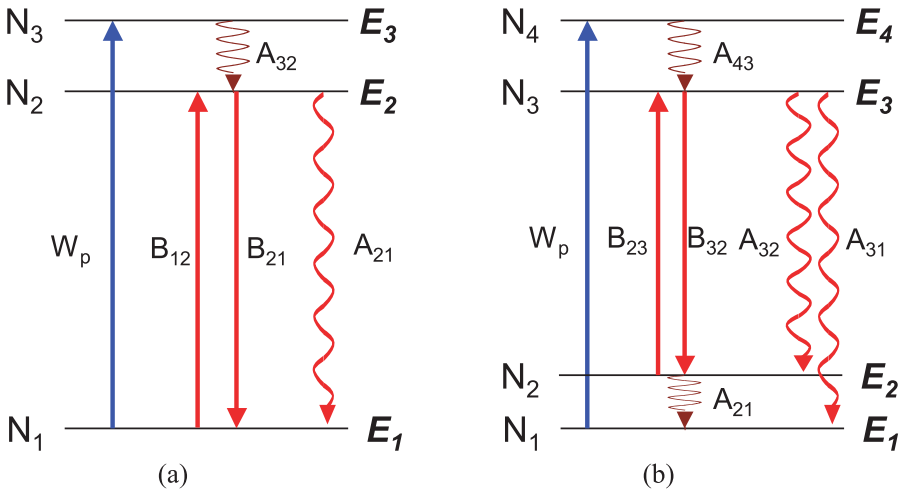


FIG. 4.2 – Schéma d'un système (a) à trois niveaux et (b) à quatre niveaux.

À titre d'exemple, tout en restant simple, nous allons écrire les équations d'évolution des populations pour un système à trois niveaux. Les coefficients

$A_{ji}$  correspondent au taux de transitions spontanées ou non rayonnantes entre les niveaux  $j$  et  $i$ , alors que les coefficients  $B_{ij}$  et  $B_{ji}$  correspondent respectivement à l'absorption et à l'émission stimulées.  $W_p$  est le taux de pompage.

Le principe de fonctionnement d'un système à trois niveaux est le suivant (Fig 4.2a) :

- on vient exciter le système à l'aide d'un rayonnement de pompe qui porte une partie de la population du niveau  $E_1$  vers le niveau  $E_3$  ;
- la population du niveau  $E_3$  relaxe rapidement vers le niveau  $E_2$  selon un processus non rayonnant (via une interaction de type phonon par exemple). On doit noter que l'on ne prend pas en compte le taux de transition  $A_{31}$  en faisant l'hypothèse que la durée de vie de cette transition est plus grande que la durée de vie associée au taux de transition  $A_{32}$  ;
- le niveau  $E_2$  correspond à un état métastable, c'est-à-dire qu'il présente une durée de vie assez grande pour réaliser une opération de stockage d'énergie et suffisamment courte pour que l'émission spontanée puisse se développer. Ce niveau est appelé le niveau supérieur de la transition ;
- enfin, sous l'effet conjugué de l'émission stimulée et de l'émission spontanée, le système retourne dans son état initial (niveau inférieur de la transition).

Ce cycle est traduit par le système d'équations couplées suivant :

$$\frac{d}{dt} \begin{vmatrix} N_3 \\ N_2 \\ N_1 \end{vmatrix} = \begin{vmatrix} -A_{32} & 0 & W_p \\ A_{32} & -B_{21}\rho(\nu) - A_{21} & B_{12}\rho(\nu) \\ 0 & B_{21}\rho(\nu) + A_{21} & -W_p - B_{12}\rho(\nu) \end{vmatrix} \begin{vmatrix} N_3 \\ N_2 \\ N_1 \end{vmatrix} \quad (4.14)$$

En régime d'excitation stationnaire (pompage), et pour un temps de transfert court entre les niveaux 3 et 2, nous avons :

$$\begin{aligned} \frac{dN_3}{dt} &= 0 & \text{soit : } W_p N_1 &= A_{32} N_3 \\ \frac{dN_1}{dt} &= -W_p N_1 + B_{21}\rho(\nu) \left( N_2 - \frac{g_2}{g_1} N_1 \right) + A_{21} N_2 \\ \frac{dN_2}{dt} &= -\frac{dN_1}{dt} \end{aligned}$$

avec la condition sur le nombre total de systèmes actifs :  $N_{tot} \approx N_1 + N_2$  selon l'hypothèse où la population du niveau  $E_3$  est transférée rapidement vers le niveau  $E_2$ .

En considérant l'inversion de population exprimée par :  $\Delta N = N_2 - \frac{g_2}{g_1} N_1$ , nous obtenons :

$$\frac{d\Delta N}{dt} = \left( 1 + \frac{g_2}{g_1} \right) (\Delta N B_{21}\rho(\nu) - W_p N_1 + A_{21} N_2) \quad (4.15)$$

En absence de rayonnement  $\rho(\nu) = 0$  et en régime d'équilibre  $d\Delta N/dt = 0$ , nous pouvons évaluer les populations des niveaux  $E_1$  et  $E_2$  :

$$N_1 = \frac{N_{\text{tot}} A_{21}}{A_{21} + W_p} \quad \text{et} \quad N_2 = \frac{N_{\text{tot}} W_p}{A_{21} + W_p}$$

L'inversion normalisée devient :  $\frac{\Delta N}{N_{\text{tot}}} = \frac{W_p - \frac{g_2}{g_1} A_{21}}{A_{21} + W_p}$ .

Tant que  $W_p < \frac{g_2}{g_1} A_{21}$ , nous avons  $\Delta N/N_{\text{tot}} < 0$ . En conséquence, la condition de seuil sur le taux de pompage est obtenue lorsque  $\Delta N/N_{\text{tot}} = 0$ ,

$$\text{soit :} \quad W_{p \text{ seuil}} = \frac{g_2}{g_1} A_{21} = \frac{g_2}{g_1 \tau_{21}} \quad (4.16)$$

où le taux d'émission spontanée  $A_{21}$  est l'inverse du temps de vie de fluorescence  $\tau_{21}$  de la transition  $E_2 \rightarrow E_1$ . Nous observons que le taux de pompage qui permet d'obtenir le seuil de l'inversion de population :  $\frac{N_2}{g_2} = \frac{N_1}{g_1}$  est d'autant plus faible que la durée de vie de l'état métastable est grande.

Pour un système à quatre niveaux (Fig. 4.2b), le système d'équations couplées peut se mettre sous la forme :

$$\frac{d}{dt} \begin{pmatrix} N_4 \\ N_3 \\ N_2 \\ N_1 \end{pmatrix} = \begin{pmatrix} -A_{43} & 0 & 0 & W_p \\ A_{43} & -B_{32}\rho(\nu) - A_{32} - A_{31} & B_{23}\rho(\nu) & 0 \\ 0 & B_{32}\rho(\nu) + A_{32} & -B_{23}\rho(\nu) - A_{21} & 0 \\ 0 & A_{31} & A_{21} & -W_p \end{pmatrix} \begin{pmatrix} N_4 \\ N_3 \\ N_2 \\ N_1 \end{pmatrix}$$

En considérant la durée de vie du niveau 2 comme très courte (le niveau inférieur de la transition émettrice de photons est quasiment vide), en absence de rayonnement  $\rho(\nu)$  et en régime stationnaire, l'inversion de population ne présente pas de seuil. Pour un tel système, l'inversion en valeur relative s'écrit :

$$\frac{\Delta N}{N_{\text{tot}}} \approx \frac{W_p}{A_{32} + W_p} = \frac{W_p \tau_{32}}{1 + W_p \tau_{32}} \quad (4.17)$$

Ce sont les pertes de cavité (absorption, diffusion, diffraction et transmission des miroirs) qui déterminent la condition sur le seuil permettant l'obtention d'une émission cohérente de lumière.

Cette cohérence se manifeste par les propriétés spécifiques du faisceau émis. Un très grand nombre de photons, donc l'énergie, peuvent être localisés dans un mode unique du rayonnement électromagnétique par opposition au rayonnement thermique où ils sont répartis sur une multitude de modes. Ainsi, il est directif avec une faible divergence (cohérence spatiale), ce qui permet son emploi dans le transport d'énergie ou de l'information à distance. Le caractère monochromatique de l'émission laser (cohérence temporelle) contribue au développement de la métrologie, soit comme source insérée dans des montages sophistiqués (programme VIRGO et LIGO servant à détecter des ondes gravitationnelles) soit en spectroscopie à haute résolution.



## 4.4 Description de quelques réalisations de laser

Une multitude d'émissions de rayonnement laser a été obtenue en exploitant :

- des mélanges gazeux (atomes neutres, ioniques, molécules) ;
- des liquides dopés par des terres rares (chélates) ou des colorants ;
- des matériaux semiconducteurs ;
- des cristaux ou des verres dopés par des ions actifs ;

en volume ou en optique guidée. De ce point de vue, les résultats obtenus récemment sur des fibres optiques dopées par des terres rares en émission laser sont remarquables.

Selon l'application désirée on favorise :

- l'énergie émise dans une impulsion unique ;
- une puissance laser élevée en continu ou en mode impulsionnel à haute cadence de répétition ;
- la finesse spectrale ou au contraire une émission et une accordabilité sur une large bande.

Grâce au développement des diodes laser, et plus particulièrement des diodes laser de puissance, on assiste à un renouveau des sources cohérentes « tout-état-solide » avec les avantages inhérents à ces milieux par rapport aux milieux gazeux ou liquides.

Les matrices solides peuvent être des cristaux ou des verres dopés par des terres rares ou par des ions métalliques de transition. La densité d'ions actifs, même en faible dopage, est bien supérieure à la densité d'atomes qu'il est possible d'exciter dans un mélange gazeux (généralement sous faible pression).

L'inventaire de tous les types de laser et une description détaillée des mécanismes mis en jeu nécessiterait l'écriture d'un volume. Cependant, nous pouvons illustrer quelques réalisations qui ont un impact sur notre vie au quotidien ou qui sont représentatives de l'actualité du sujet.

### 4.4.1 Diodes laser et génie quantique

Les avancées les plus remarquables ont été observées avec les développements des diodes laser, tant par le domaine spectral couvert que par les caractéristiques obtenues. Elles doivent ces résultats aux propriétés des structures à multi-puits quantiques (MPQ). Dans de telles structures, les porteurs (électrons et trous) sont confinés spatialement et les niveaux d'énergie pouvant être pris par les électrons dans la bande de conduction et les trous dans la bande de valence sont quantifiés. En jouant sur les propriétés du puits, sa

largeur et/ou sa profondeur, on est à même de définir une énergie de bande interdite et donc une longueur d'onde d'émission.

Par ailleurs, au confinement des porteurs est associée une fonction d'onde (fonction propre) localisée au niveau du puits. La force d'oscillateur (i.e. le coefficient de gain optique), qui est proportionnelle à l'intégrale de recouvrement entre les fonctions d'onde associées aux électrons et aux trous pour une transition optique, est plus élevée que celle pouvant être obtenue dans un matériau massif.

C'est le développement des techniques de croissance d'alliages de composés semiconducteurs qui a permis de réaliser des hétérostructures à base de MPQ de plus en plus complexes. En effet, avec les techniques d'épitaxie par jet moléculaire (EJM ou MBE Molecular Beam Epitaxy) ou par MOCVD (Metal Organic Chemical Vapor Deposition ou ses variantes CBE Chemical Beam Epitaxy . . .), on sait contrôler le dépôt de matériaux semiconducteurs monocouche par monocouche atomique sur un substrat.

La figure 4.3 illustre l'assemblage de doubles hétérostructures à puits quantiques. Cette configuration est obtenue par croissance d'un matériau à grande énergie de bande interdite (matériau de barrière) de part et d'autre d'un matériau à plus faible énergie de bande interdite (matériau de puits).

Si l'épaisseur du matériau de puits  $L_z$  est de l'ordre de grandeur de la fonction d'onde associée à un porteur libre dans le matériau semiconducteur, cette perturbation du milieu va imposer au porteur un déplacement quantifié selon la direction  $z$  avec des états discrets d'énergie. Cet axe correspond à l'axe de croissance de l'empilement. Ces états électroniques dépendent de la largeur et de la hauteur du puits de potentiel. La hauteur du puits varie selon les alliages de composés semiconducteurs employés. En revanche, la particule possède la propriété de pouvoir se déplacer librement sur les deux autres axes dans le plan du puits.

Cette propriété fait que l'on parle d'un gaz d'électrons bi-dimensionnel (2D) pour une structure à puits quantique. Le mouvement d'un électron se trouve quantifié dans une direction. Dans un matériau massif, le gaz d'électrons est tri-dimensionnel (3D). La recherche d'un gain d'amplification plus élevé, d'un faible courant de seuil et d'une finesse spectrale à l'émission expliquent les travaux menés autour des structures à plus basse dimensionnalité (1D : fil quantique et 0D : boîte quantique).

Typiquement, la largeur du puits est de l'ordre de 10 nm. On parle, pour ces structures, de matériau artificiel dans la mesure où l'on définit une énergie de bande interdite qui n'est plus liée aux caractéristiques physiques du matériau massif mais à celles du puits quantique.

L'émission optique d'une diode laser conventionnelle résulte de la recombinaison d'une paire électron-trou. Cette recombinaison est de type inter bandes puisque l'on traite des transitions entre la bande de conduction et la bande de valence du matériau.

Ainsi, en jouant sur la largeur du puits, nous pouvons modifier la position du niveau d'énergie quantifié du premier état électronique dans la bande de

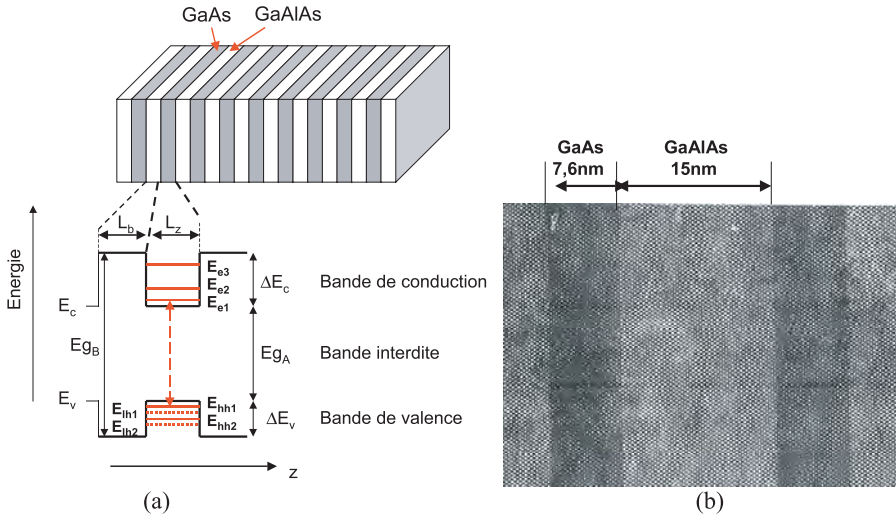


FIG. 4.3 – (a) Illustration d'une structure à multi-puits quantiques et niveaux d'énergie pouvant être occupés par les électrons dans la bande de conduction et par les trous dans la bande de valence. (b) Image sous fort grossissement de deux structures à puits quantiques.  $L_z$  et  $L_b$  représentent respectivement les largeurs du matériau de puits, d'énergie de bande interdite  $E_{gA}$  et la largeur du matériau d'énergie de bande interdite  $E_{gB}$ . Les niveaux d'énergie quantifiés dans la bande de conduction sont représentés par  $E_{e1}, E_{e2}, \dots$  et ceux dans la bande de valence par  $E_{hh1}, E_{hh2}, \dots$  pour les « trous lourds » (heavy-holes) et par  $E_{lh1}, E_{lh2} \dots$  pour les « trous légers » (light-holes) [20]. La transition optique ou l'énergie de bande interdite de la structure correspond à  $E_{gA} + E_{e1} + E_{hh1}$ .

conduction ( $E_{e1}$  sur la figure 4.3a) et celle associée à un trou ( $E_{hh1}$ ). L'énergie de bande interdite liée au puits quantique  $E_{eh}$  (qui correspond à une transition optique de longueur d'onde  $\lambda = \frac{1,24}{E_{eh}}$  avec  $\lambda$  en  $\mu\text{m}$  et l'énergie exprimée en électron Volt) est comprise entre celle du matériau massif qui constitue le puits ( $E_{gA}$ ) et celle du matériau de barrière ( $E_{gB}$ ). Elle est égale à :

$$E_{eh} = E_{gA} + E_{e1} + E_{hh1} \tag{4.18}$$

Cette propriété fait que l'on peut réaliser des émetteurs de lumière dont la longueur d'onde peut être ajustée à la conception et imposée par la structure du puits quantique. Elle est par exemple largement mise à profit dans des expériences fondamentales qui reposent sur l'emploi de sources laser dont l'émission est affinée spectralement et ajustée sur une raie atomique.

Pour réaliser la zone de recombinaison, on utilise deux matériaux dopés par des atomes donneurs (n) et accepteurs (p) d'électrons entre lesquels se trouve une zone active non dopée qui contient les puits quantiques (Fig. 4.4). Cette quasi-jonction p-n à double hétérojonction permet d'injecter par un courant

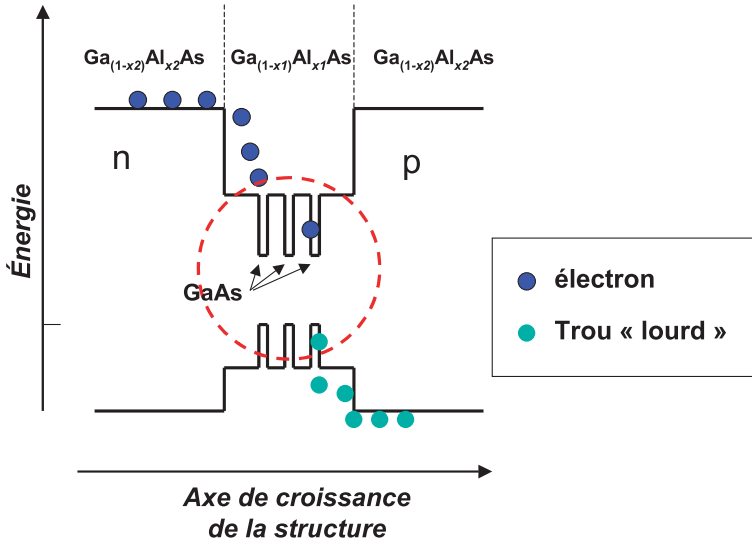


FIG. 4.4 – Diagramme de bande pour une structure de diode laser GaAs/GaAlAs.  $x_1$  et  $x_2$  correspondent à la teneur en aluminium du ternaire GaAlAs avec  $x_2 > x_1$ . Les flèches signalent la localisation des puits quantiques en GaAs. Le cercle en tirets rouge simule la localisation du faisceau lumineux.

en sens direct des porteurs qui viennent se localiser dans la structure à multi-puits quantiques. Cette zone est alors le siège du mécanisme de recombinaison des paires électron-trou avec émission d'un photon. La zone non dopée fait également office de guide d'onde optique. Ainsi, le pompage de ce type de laser s'effectue par l'injection de porteurs.

Initialement, ces structures étaient développées à partir de composés d'arséniure de gallium (GaAs et AlGaAs). Cette ingénierie est maintenant appliquée à tous les composés semiconducteurs, que ce soit en électronique (transistors hyperfréquences) ou pour fabriquer des diodes laser.

Le coefficient de gain d'amplification est important dans les semiconducteurs. À titre d'exemple, nous avons porté sur la figure 4.5 l'évolution du coefficient de gain en fonction de la densité de porteurs injectés et en fonction de la longueur d'onde au voisinage de la longueur d'onde associée à l'énergie de bande interdite.

On observe le fort gain d'amplification optique pouvant être obtenu dans les structures à puits quantiques (2D), comparativement à un milieu massif (3D), ainsi que la différence de comportement spectral en fonction de la nature du matériau et de la densité de porteurs injectés. Les gains obtenus à partir des matériaux semiconducteurs entraînent que l'effet laser peut être engendré en utilisant la réflectivité naturelle des facettes réalisées par clivage du cristal. Cette réflectivité résulte de la forte valeur de l'indice de réfraction optique

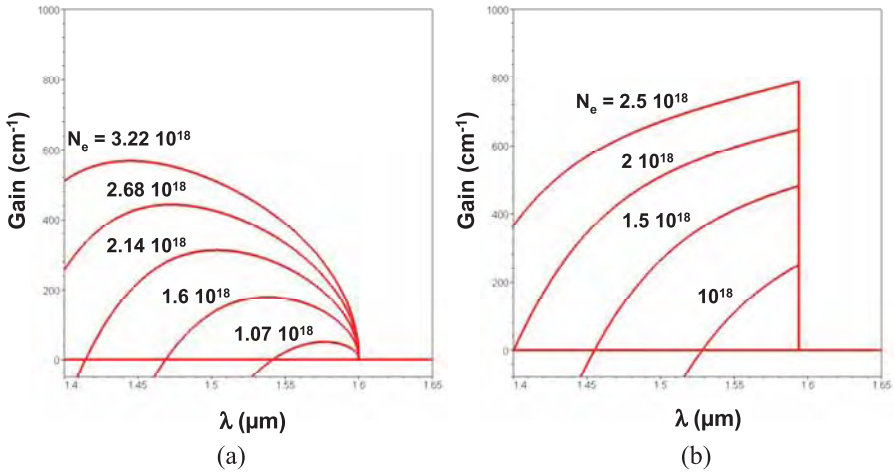


FIG. 4.5 – Distribution spectrale du gain d'amplification ( $\text{cm}^{-1}$ ) : (a) d'un matériau semiconducteur massif InGaAsP sur substrat InP, (b) d'un puits quantique InGaAs/InP (matériaux de puits/barrière) d'épaisseur 10 nm, en fonction de la densité  $N_e$  de porteurs injectés ( $\text{cm}^{-3}$ ).

( $n \sim 3,5$ ), ce qui permet, grâce à cette opération de clivage, de réaliser une cavité monolithique avec émission laser.

Le rendement à la prise de tels composants, qui correspond au rapport entre la puissance optique émise et la puissance électrique appliquée, est supérieur à 50 %. Même avec ce bon rendement, on doit garder à l'esprit qu'une source laser délivrant 50 W en régime continu impose de savoir résoudre l'évacuation de 50 W thermique. La figure 4.6 illustre l'évolution de la puissance optique émise par une diode unique à ruban large en fonction du courant injecté. Une puissance de l'ordre de 7 W peut être émise en régime continu et 10 W en régime quasi-continu (durée impulsion : 50  $\mu\text{s}$  à 100 Hz).

En multipliant le nombre de ces éléments sur un même substrat, on obtient une barrette de diodes dont la puissance totale à l'émission est la somme des puissances émises par chaque émetteur. C'est ce qui est présenté sur les figures 4.7a et 4.7b.

Les propriétés d'émission couvrent le domaine spectral qu'occupent les bandes d'absorption (d'excitation) des terres rares. C'est l'émergence de ces composants qui a grandement contribué à donner un nouveau souffle au développement des lasers solides au niveau des architectures et des matériaux en se substituant aux lampes ou aux flashes pour le pompage optique.

Un des gros avantages que présentent les diodes laser est l'aptitude de pouvoir émettre sur une plage de longueur d'onde définie par les alliages de matériaux semiconducteurs employés ou par la distribution des hétérostructures. La miniaturisation, le traitement collectif des processus mis en jeu au

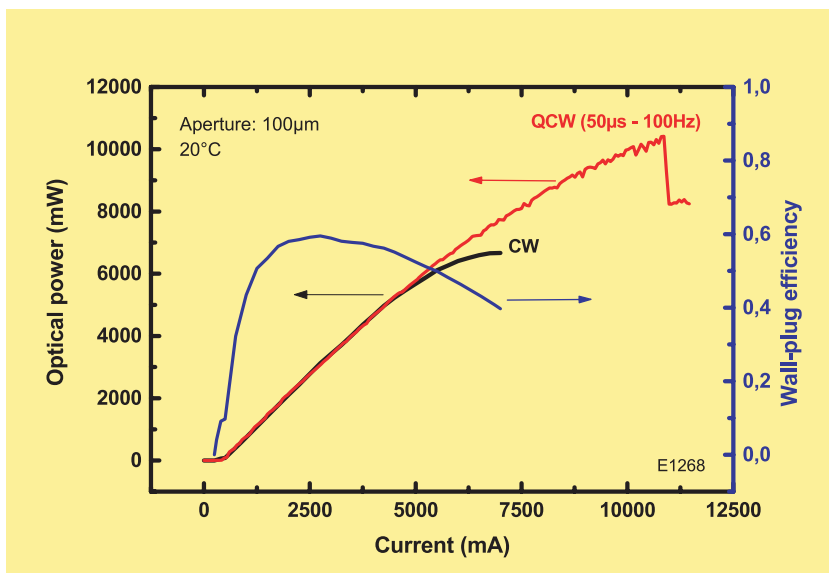


FIG. 4.6 – Puissance optique délivrée par une diode laser GaAs/GaAlAs à puits quantique à  $\lambda = 800 \text{ nm}$  en régime continu (CW) ou quasi continu (QCW) et rendement électro-optique à la prise électrique. Longueur de cavité 1 mm, largeur du ruban  $100 \mu\text{m}$ . (THALES Research & Technology.)

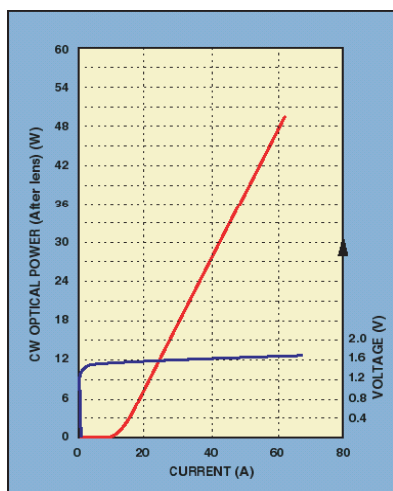


FIG. 4.7 – (a) Puissance émise à  $843 \text{ nm}$  en régime continu (courbe rouge) en fonction du courant avec une barrette de 1 cm de large et une longueur de cavité de  $900 \mu\text{m}$  pour chaque émetteur. Caractéristique  $V = f$  (courant d'injection) courbe bleue. (b) Photographie d'une barrette de diodes intégrée dans son embase avec refroidissement. (THALES Research & Technology, THALES Laser Diodes.)

niveau technologique, propre à la micro-électronique, le rendement de conversion électrique/optique font de la diode laser le composant le plus répandu au niveau des applications et le plus important en terme de couverture de marché. Le stockage optique et la transmission de données sont les meilleures illustrations de la concrétisation des idées émises en physique il y a de cela cent ans et qui marquent notre société.

Des structures plus complexes peuvent être réalisées, en particulier si les puits quantiques sont proches l'un de l'autre (largeur de barrière de l'ordre de grandeur de celle du puits), on obtient un super-réseau. Cette configuration est équivalente à un ensemble d'oscillateurs couplés qui se traduit par une levée de dégénérescence des niveaux d'énergie de chaque état quantique et une localisation de la particule sur l'ensemble du super-réseau. C'est cette propriété qui est mise à profit pour réaliser des lasers à cascade quantique.

#### 4.4.2 Diodes laser à cascade quantique (QCL : Quantum Cascade Laser)

Le principe de l'émission laser de ces composants repose sur l'emploi de transitions intra-bandes. Le mécanisme de transport par effet tunnel résonnant et l'émission intra-bande dans les hétérostructures étaient proposés par R. Kazarinov et R.A. Suris en 1971 [21]. Ces transitions exploitent les propriétés d'un seul type de porteur dans le semiconducteur, en l'occurrence les électrons, d'où la dénomination de laser unipolaire.

Comparativement à une transition inter-bande, une transition intra-bande peut être assimilée à un système à trois niveaux (quasi quatre niveaux) comme le montre la figure 4.8.

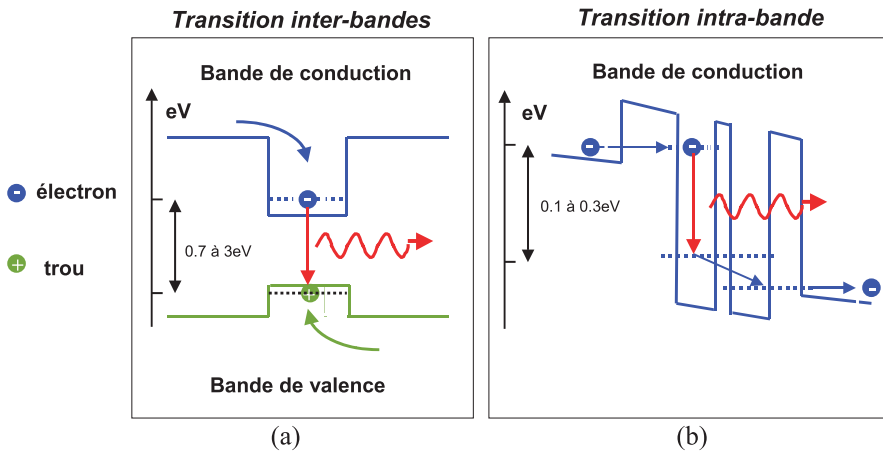


FIG. 4.8 – Représentation schématique (a) d'une transition inter-bande (rayonnement émis par la recombinaison d'une paire électron-trou) et (b) d'une transition intra-bande. On note la différence d'énergie entre les niveaux quantifiés.

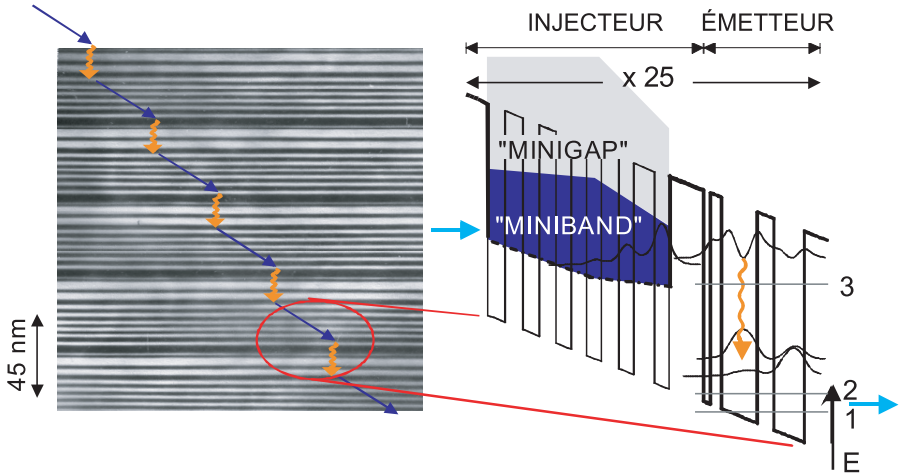


FIG. 4.9 – Structure de l'empilement d'hétérojonctions et définition du mécanisme d'émission de photons infrarouges.

En jouant sur la structure des puits (largeur, hauteur ou structure couplée), il est possible d'avoir plusieurs niveaux d'énergie pouvant être pris par un électron. À titre d'exemple, la figure 4.9 illustre cette propriété. La hauteur du puits de potentiel dans la bande de conduction va définir le domaine spectral d'émission. En effet, la différence d'énergie entre deux niveaux ne peut excéder cette hauteur. En fonction des alliages semiconducteurs mis en œuvre pour réaliser la barrière et le puits, cette hauteur est plus ou moins importante. Pour le couple GaAs/Ga<sub>0,67</sub>Al<sub>0,33</sub>As (puits/barrière), elle vaut typiquement 295 meV, alors que l'emploi du composé Ga<sub>0,47</sub>In<sub>0,53</sub>As/Al<sub>0,48</sub>In<sub>0,52</sub>As est caractérisé par une hauteur de 520 meV. En conséquence, il sera possible de réaliser des sources dans le lointain infrarouge (LIR) avec GaAs/GaAlAs et dans le moyen IR (MIR) avec le second alliage. Les compositions de ces alliages sont déterminées par l'adaptation en maille cristalline des composés ternaires ou binaires entre eux.

C'est en 1994 que F. Capasso et ses collaborateurs [22] démontrent pour la première fois une émission laser dans l'infrarouge en exploitant les transitions intra-bandes dans une hétérostructure. Cette réalisation est suivie de nouvelles avancées démontrant une émission laser impulsionnelle puis en régime continu dans le moyen et dans le lointain infra-rouge à des températures supérieures à celle de l'azote liquide [23] et maintenant proches de la température ambiante. L'effet de cascade et d'injection peut être décrit de la manière suivante (voir Fig. 4.9). Tout d'abord on réalise un super réseau (multiples puits couplés entre eux) qui est caractérisé par un grand nombre d'états quantiques pouvant être pris par l'électron. C'est ce que l'on appelle une mini-bande qui traduit la levée de dégénérescence des niveaux d'énergie quantifiés. En présence d'un



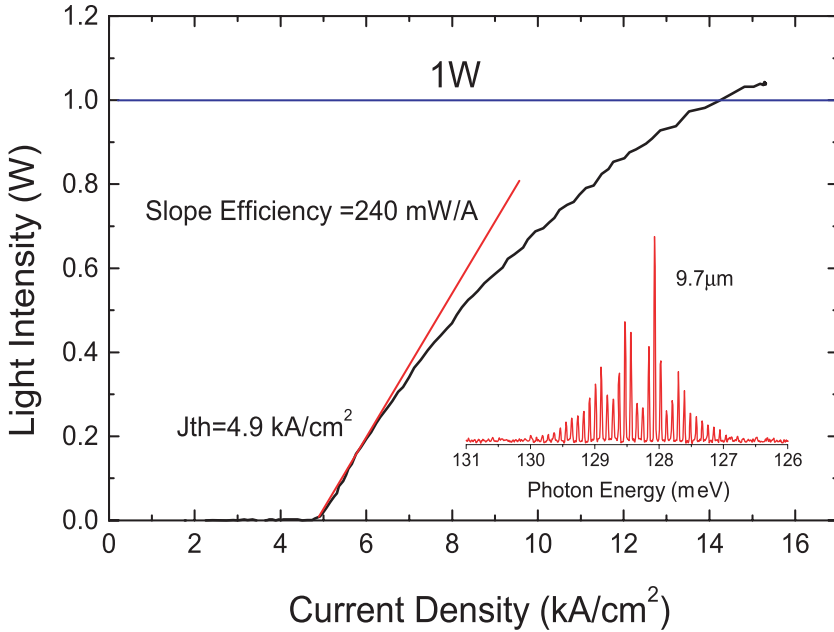


FIG. 4.10 – Puissance émise par une diode QCL à la longueur d'onde de  $9,7 \mu\text{m}$  à la température de  $77 \text{ K}$ . Le rendement de conversion est de  $240 \text{ mW/A}$  et la densité de courant de seuil est de  $4,9 \text{ kA/cm}^2$ . (THALES Research & Technology.)

champ électrique, on déforme l'ensemble des puits de potentiel. L'injection d'un électron s'effectue à travers la mini-bande via un des niveaux d'énergie. Par effet tunnel, il se couple au niveau 3 de la structure à puits quantiques couplés qui est le siège de la transition optique (zone émetteur). L'électron passe sur le niveau 2 en émettant un photon IR puis via les phonons il rejoint le niveau 1. La fonction d'onde de cet état se trouve couplée à un des états de la mini-bande suivante et le processus de transfert et d'émission se poursuit de proche en proche. Ainsi avec un électron on est capable de générer autant de photons qu'il y a d'éléments injecteur-émetteur intégrés dans la structure épitaxiée. Ce mécanisme explique l'effet de cascade.

Un résultat type est reproduit sur la figure 4.10 qui donne le niveau de puissance optique émis par une diode laser à cascade quantique à  $9,7 \mu\text{m}$  et à la température  $T = 77 \text{ K}$ . L'émission d'une puissance optique de  $1 \text{ W}$  est obtenue sur un seul élément, ce qui ouvre de nouvelles perspectives d'emploi des diodes laser en infrarouge. On pourra se reporter à la référence [24] pour une analyse plus approfondie des mécanismes mis en jeu et des caractéristiques de l'émission laser délivrée par ce type de composant semiconducteur.

L'intégration d'une rétroaction distribuée par réseau de Bragg gravé (DFB : Distributed Feedback laser) permet d'obtenir une émission monofréquence comme le montre la figure 4.11. Cette rétro réflexion in situ, qui

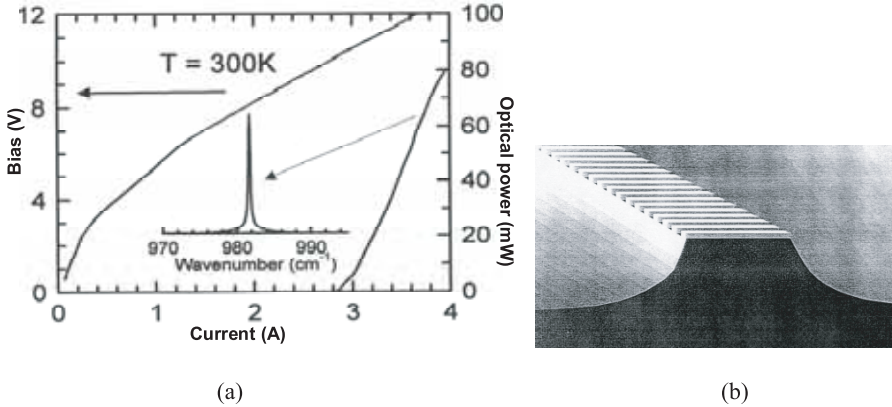


FIG. 4.11 – (a) Caractéristiques électro-optique (puissance =  $f$  (courant d'injection) et  $V = f(I)$ ) et spectrale obtenues sur un laser à cascade quantique au voisinage de  $10 \mu\text{m}$ . Au centre, spectre d'émission. (b) Photographie du réseau distribué intégré à la structure du laser ruban.

affecte le mode transverse guidé dans la structure, est équivalente aux propriétés de réflexion que présente un empilement de lames quart d'onde à deux indices de réfraction.

L'accordabilité du spectre d'émission peut être obtenue en contrôlant le courant d'injection (ou la température). Cette disposition donne à ce type de source une excellente aptitude à la détection d'espèces moléculaires sous forme de trace par absorption dans l'infrarouge. Une telle source couplée à une cellule photoacoustique offre des sensibilités de détection inférieures au ppb (part par milliard).

C'est donc l'application du génie quantique et la maîtrise de la croissance de composés semiconducteurs qui permet de concevoir des sources de lumière cohérentes miniaturisées sur un domaine spectral de plus en plus large. À l'opposé de celui couvert par les diodes à cascade quantique ( $\lambda = 4$  à  $18 \mu\text{m}$ ), une émission laser dans le domaine des courtes longueurs d'onde ( $\lambda \sim 0,3 \mu\text{m}$ ) est obtenue à partir d'hétérostructures de composés semiconducteurs à grande bande d'énergie interdite (GaN, AlN : nitrure de gallium, et d'aluminium). Cette tendance est illustrée par les investigations menées, sur le domaine spectral relativement vierge, au-delà du lointain infrarouge.

#### 4.4.3 Rayonnement TeraHertz (THz)

Les ondes THz ( $10^{12}$  Hz ou  $\lambda = 299,79 \mu\text{m}$  dans le vide) couvrent un domaine spectral à la frontière entre l'optique et les micro-ondes (Fig. 4.12). La génération de ce rayonnement fait l'objet d'un récent intérêt scientifique avec les premières démonstrations réalisées à partir d'impulsions laser sub-picosecondes venant interagir avec un photoconducteur.

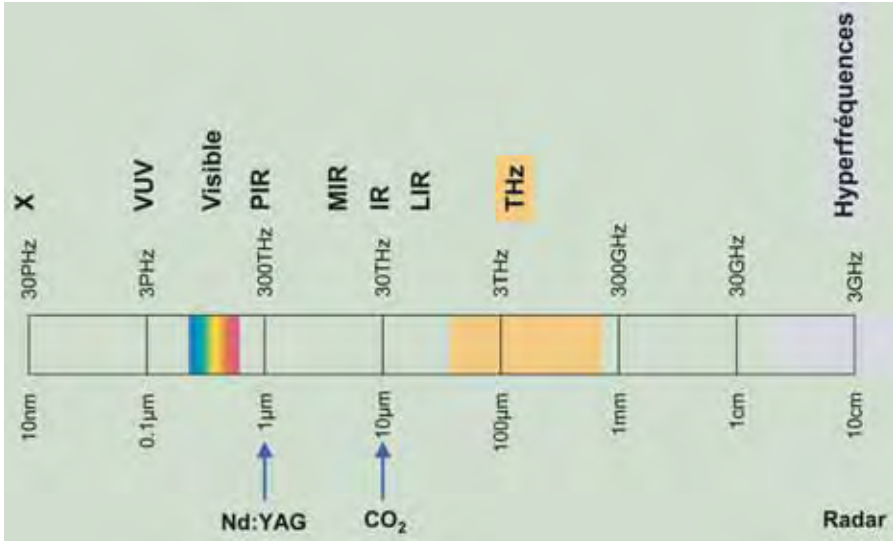


FIG. 4.12 – Positionnement du rayonnement THz dans l'échelle des longueurs d'onde (fréquence) et de deux sources laser conventionnelles. (PHz =  $10^{15}$  Hz, THz =  $10^{12}$  Hz). VUV : ultraviolet profond, PIR : proche infrarouge, MIR : moyen infrarouge, IR : infrarouge, LIR : lointain infrarouge.

Il existe deux méthodes pour produire un rayonnement THz. La première consiste à engendrer un large spectre d'émission en exploitant le phénomène d'onde de choc optique associé à une rectification optique. Ce mécanisme repose sur le fait que la polarisation induite dans un milieu ne suit pas celle du champ inducteur. La rectification optique est la manifestation d'un effet non linéaire du second ordre. En présence d'une variation rapide de l'enveloppe d'un champ optique, la polarisation non linéaire induite dans un milieu non centro-symétrique varie également rapidement (voir appendice A et paragraphe 4.6). Cette polarisation est non stationnaire et elle se déplace à la vitesse de groupe de l'impulsion optique. La contribution basse fréquence de la réponse diélectrique, liée aux vibrations du réseau cristallographique, fait que la vitesse de déplacement de la source excède la vitesse du rayonnement. On obtient un effet équivalent à l'effet Cerenkov. En conséquence la polarisation du milieu émet un rayonnement étendu qui couvre celui des THz [25].

Une seconde approche consiste à employer deux faisceaux laser décalés en fréquence de quelques THz qui viennent interagir avec un élément non linéaire tel qu'un photoconducteur. En effet, le processus de photodétection est par nature non linéaire puisque le photocourant engendré est proportionnel à l'éclairement incident, donc quadratique en champ. Ainsi, le photo-courant contient une composante à la différence de fréquence des deux porteuses optiques incidentes. Cette propriété explique l'emploi de la technique d'hétérodynage

optique en réception cohérente qui exploite les propriétés de réduction de bruit par auto-filtrage spectral. Avec un photodétecteur quantique on dispose donc d'un composant qui convertit un grain de lumière en une paire électron-trou et qui fait office de mélangeur de fréquences.

Ainsi, soient deux champs  $E^{\omega_1} = E_1 e^{i\omega_1 t}$  et  $E^{\omega_2} = E_2 e^{i(\omega_2 t + \phi)}$ . Le photocourant engendré est proportionnel à l'éclairement  $i_d = \eta |E|^2 = \eta |E_1 + E_2|^2$  où  $\eta$  est un coefficient de proportionnalité incluant le rendement quantique du photodétecteur et un facteur reliant l'amplitude du champ électrique à l'éclairement optique. On obtient un photo-courant :

$$i_d = \eta \left( |E_1|^2 + |E_2|^2 + 2E_1 E_2 \cos((\omega_1 - \omega_2)t + \phi) \right)$$

qui présente une composante à la différence de fréquences (fréquence de battement) pouvant être ajustée sur le domaine THz. Le photoconducteur comprend une antenne adaptée qui rayonne cette composante. Pour peu que l'on sache réaliser une source bi-spectrale avec une faible largeur de raie pour chaque fréquence il devient possible d'engendrer un rayonnement THz en continu. Le rendement de conversion favorise l'emploi de photons IR par rapport au visible. En effet, l'interaction est équivalente à l'interaction de deux photons d'énergie  $\hbar\omega_1$  et  $\hbar\omega_2$  qui engendrent un photon THz. Dans ces conditions, le rendement de conversion correspond à :

$$\eta_{\text{THz}} = \frac{\omega_{\text{THz}}}{\omega_1 + \omega_2}$$

L'utilisation d'une source laser bi-fréquences émettant au voisinage de la longueur d'onde de  $1 \mu\text{m}$  pour la génération d'une onde à 1 THz présente un rendement maximum de conversion de 0,17 %. Si l'on opère plutôt dans l'infrarouge à  $10 \mu\text{m}$ , on gagne naturellement un ordre de grandeur.

Avec l'emploi de sources primaires à  $1 \mu\text{m}$ , les propriétés de réponse en fréquence de photoconducteurs GaAs dont la croissance s'effectue à basse température (LT-GaAs) font de ces composants les éléments clés d'une source de rayonnement THz.

L'association de diodes laser à cascade quantique à émission bi-spectrale ou émettant directement un rayonnement sub-millimétrique [26] à des détecteurs à multiples puits quantiques IR (QWIP : Quantum Wells Infrared Photodetector) ou à des photoconducteurs HgCdTe représente une architecture plus favorable pour la génération d'ondes THz.

De nombreuses applications se développent autour de ce type de rayonnement. En particulier dans le domaine de la détection et l'identification de composés moléculaires dans l'atmosphère, en biotechnologie et dans l'imagerie biomédicale. En effet, les composés liquides présentent des bandes d'absorption importantes et discrètes liées aux modes de rotation moléculaire. De plus, la grande longueur d'onde des ondes THz rend possible leur propagation dans les milieux diffusants ou turbulents. Un autre domaine où le rayonnement THz peut trouver des applications concerne celui de la sécurité. La figure 4.13



FIG. 4.13 – Image millimétrique servant à la détection et à l'identification d'objets masqués. (THALES Sensor, Basingstokes.)

illustre une application de l'imagerie millimétrique pour la détection d'objets masqués ou le contrôle d'objets empaquetés.

#### 4.4.4 Le renouveau des lasers solides

On a assisté ces vingt dernières années à une multiplication de systèmes laser ou de produits à base de matériaux actifs à l'état solide dopés par des ions qui constituent le milieu actif amplificateur.

Une large diversité d'ions de terres rares ou de métaux de transition a bénéficié du pompage par diodes (ions  $\text{Nd}^{3+}$ ,  $\text{Er}^{3+}$ ,  $\text{Ho}^{3+}$ ,  $\text{Tm}^{3+}$ ,  $\text{Yb}^{3+}$ ,  $\text{Cr}^{3+}$ ,  $\text{Cr}^{4+}$ ,  $\text{Cr}^{2+}$ ,  $\text{Ti}^{3+}$ , ...) et une grande variété de matrices solides a fait l'objet de nombreux travaux. Pour les cristaux on utilise surtout soit des oxydes ( $\text{Al}_2\text{O}_3$ ,  $\text{BeAl}_2\text{O}_3$ ,  $\text{YAlO}_3$ ,  $\text{CaWO}_4$ ,  $\text{YVO}_4$ ,  $\text{Y}_3\text{Al}_5\text{O}_{12}$ ,  $\text{LiNdP}_4\text{O}_4$ , ...), soit des fluorures ( $\text{LiYF}_4$ ,  $\text{LiCaAlF}_4$ ,  $\text{LiSrAlF}_6$ , ...).

Les verres sont également utilisés (borates, phosphates, silicates, fluoroberyllates, fluorozirconates). Les matériaux amorphes, comme les verres, peuvent être fortement dopés d'ions actifs. Toutefois ils ont le désavantage d'avoir une mauvaise conductivité thermique. En revanche, le dopage des cristaux est limité par la création de défauts, ils ne se prêtent pas à des dopages élevés; toutefois ils ont de meilleures propriétés thermo-mécaniques.

Le mode de pompage de ces milieux est optique et les développements observés autour des diodes laser, qui sont adaptées au pompage des ions actifs, ont largement participé au renouveau des laser solides. Pour le pompage des ions insérés dans des matrices solides on remplace le flash ou la lampe par une barrette ou un empilement de barrettes de diodes laser dont la longueur d'onde d'émission est ajustée spectralement à une bande d'absorption de l'ion actif qui est inséré dans une matrice solide.

Deux modes de pompage peuvent exister : transverse ou longitudinal. Dans la seconde configuration, il est possible d'ajuster la zone excitée à celle occupée spatialement par le mode de cavité. Cette architecture permet d'obtenir des rendements de conversion optique/optique de l'ordre de 50 %. Le rendement global, incluant le rendement de conversion électrique-optique d'une diode, est alors de 25 %. Un tel rendement est bien supérieur à celui obtenu avec l'emploi de lampes (< 0,1 %) ou de flashes (< 1 %) pour le pompage.

Le pompage localisé ou distribué par diodes laser s'adapte bien au pompage optique de nouvelles géométries pour les milieux actifs (barreau, disque, tube, plaque, cube, fibre optique et optique guidée, planaire et 2D...).

Les principaux avantages associés à l'emploi de lasers solides sont :

- une stabilité de fréquence, liée à leur caractère monolithique (cavité intégrée au milieu actif) ;
- un accès commode à la cavité (pour la modulation ou pour réaliser des interactions non linéaires intra-cavité) ;
- une bonne voire excellente qualité de faisceau (en particulier avec les laser à fibres optiques) ;
- un stockage d'énergie lié à la grande durée de vie de l'état métastable des ions excités.

Le pompage par diodes a pour conséquence :

- d'améliorer le rendement électro-optique à la prise ;
- d'augmenter la durée de vie du système de pompage ;
- de gagner sur la taille des alimentations électriques (basse tension) ;
- d'obtenir une bonne reproductibilité de l'émission laser d'une impulsion à l'autre ou une faible fluctuation de la puissance délivrée en régime continu ;
- d'avoir des sources compactes ;
- de pouvoir travailler en régime de haute cadence ;
- de disposer d'une bonne stabilité du faisceau à l'émission.

En fait, un milieu solide dopé par des terres rares est un accumulateur d'énergie (équivalent à un condensateur) lié au stockage des ions dans un état métastable à grande durée de vie de fluorescence (100  $\mu\text{s}$  à 10 ms selon l'ion et la matrice). Ainsi, en régime déclenché, une grande partie de l'énergie stockée est produite dans une impulsion de courte durée, ce qui permet d'obtenir une puissance crête instantanée élevée.

C'est également un convertisseur de mode spatial, car il transforme un faisceau de pompe multimode transverse en un faisceau monomode transverse à une autre longueur d'onde.

L'efficacité ultime d'un laser est naturellement gouvernée par le rendement quantique, qui correspond au rapport entre l'énergie du photon émis et l'énergie d'un photon de pompe. En fait, le paramètre de qualification d'une source laser est sa luminance. Elle est reliée à la surface d'émission  $S$  et à ses propriétés de divergence (demi-angle au sommet du cône de lumière associé au faisceau). Ces deux derniers paramètres déterminent l'étendue géométrique du faisceau qui, en optique, est un invariant dans un système sans pertes.

La luminance exprimée en  $\text{W}/\text{m}^2/\text{sr}$  est donnée par l'expression :

$$L = \frac{P}{S \times \Omega} \quad (4.19)$$

avec la puissance  $P$  en Watt, la surface  $S$  en  $\text{m}^2$  et  $\Omega$  en stéradian.  $\Omega$  est l'angle solide défini par le demi-angle au sommet de la divergence du faisceau.

À titre d'exemple, on considère une fibre laser qui émet un rayonnement à 1  $\mu\text{m}$  et une puissance de 10 W. Le rayon du mode dans la fibre est de 20  $\mu\text{m}$  (cœur large). La divergence du faisceau étant limitée par la diffraction, nous obtenons une luminance de l'ordre de 1  $\text{GW}/\text{cm}^2/\text{sr}$ . Le record de luminance en émission laser est détenu par une fibre dopée  $\text{Yb}^{3+}$  fournissant une puissance de 610 W en continu à 1,1  $\mu\text{m}$ , avec une qualité de faisceau pratiquement limitée par diffraction (1,3 fois la limite de diffraction) [27] et un rendement optique de pompage de 80 %.

Bien que le rendement quantique puisse être faible dans certains cas, la conversion de mode permet de compenser ce défaut par un accroissement considérable de la luminance associée au faisceau laser émis par rapport à celle de la diode de pompage.

Enfin, un tel laser est un convertisseur de fréquence caractérisé par sa largeur spectrale de gain, qui peut être inférieure à celle d'un milieu semiconducteur et qui transforme un laser de pompe multimode longitudinal en un faisceau laser monofréquence.

Un schéma type de laser solide pompé transversalement ou/et longitudinalement est illustré par la figure 4.14.

Le laser saphir titane  $\text{Ti} : \text{Al}_2\text{O}_3$  [28] a tout particulièrement contribué au développement d'un nouveau domaine de l'instrumentation et il a enrichi l'approche physique de l'interaction lumière-matière. Sa grande largeur spectrale de gain qui s'étend de 0,7 à 1,1  $\mu\text{m}$ , et ses propriétés liées à son état

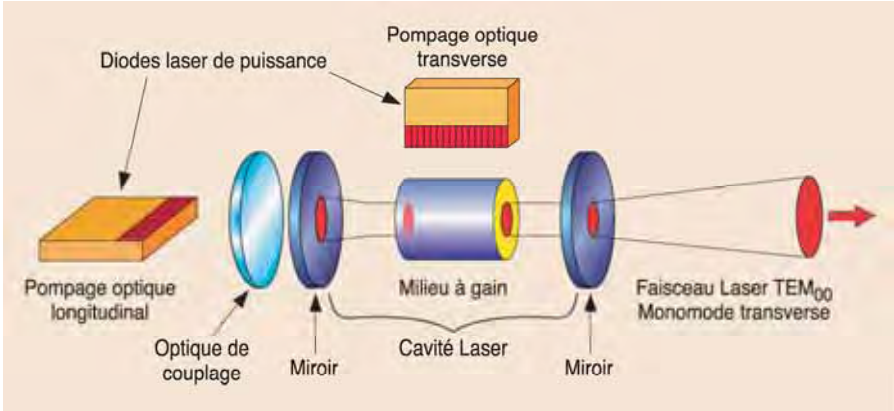


FIG. 4.14 – Schéma d'un laser solide pompé par diodes : le pompage est soit longitudinal, soit transversal.

solide l'ont conduit à se substituer aux laser à colorants pour la réalisation de sources accordables en longueur d'onde dans de très nombreux cas.

Toutefois, la bande d'absorption qui participe à l'excitation de l'ion  $Ti^{3+}$  est malencontreusement localisée entre 0,4 et 0,65  $\mu m$ . On ne dispose pas de diodes laser de puissance sur cette fenêtre. Par ailleurs la faible durée de vie de fluorescence du niveau excité  $Ti^{3+}$  (3,8  $\mu s$ ) rend son excitation par flash peu efficace. C'est donc en le pompant en continu par un laser solide comme le Nd:YAG pompé par diode et doublé en fréquence, ou par ce même laser pompé par flash en régime impulsif, que l'on a exploité ses propriétés à l'émission laser. Ce matériau dispose d'autres atouts tels que sa haute conductivité thermique et sa tenue mécanique.

#### 4.4.5 Laser femtosecondes

La grande largeur spectrale d'émission du laser saphir titane a naturellement conduit à son utilisation pour générer des impulsions ultra-courtes, en exploitant le mécanisme de verrouillage de phase entre modes longitudinaux [29, 30]. À partir des propriétés de la transformée de Fourier, on a une relation entre la largeur spectrale à mi-hauteur  $\Delta\nu_{1/2}$  d'un signal et sa largeur temporelle, qui pour une distribution gaussienne s'écrit :

$$\Delta\nu_{1/2}\Delta\tau = \frac{\sqrt{2\log 2}}{\pi} \quad (4.20)$$

On observe qu'un large spectre de fréquences permet d'obtenir une impulsion ultra-courte. Les propriétés de fluorescence de l'ion  $Ti^{3+}$  dans le saphir répondent à ce critère, ce qui explique les courtes durées d'impulsions obtenues. Le franchissement de la frontière des 10 femtosecondes (1 fs =  $10^{-15}$  s) [31] fut



réalisé dans les années 1990, suivi par l'obtention de quelques cycles optiques dans la fonction enveloppe temporelle d'une impulsion [32]. La propagation d'une impulsion de 10 fs équivaut au déplacement dans l'espace d'une allumette longue de  $3 \mu\text{m}$  à la vitesse de la lumière.

Un schéma de principe de ce type de laser est présenté sur la figure 4.15. Les prismes insérés dans la cavité ont pour fonction de compenser les effets de dispersion chromatique sur la phase des différents modes longitudinaux pouvant se développer sur la large courbe de gain du milieu amplificateur.

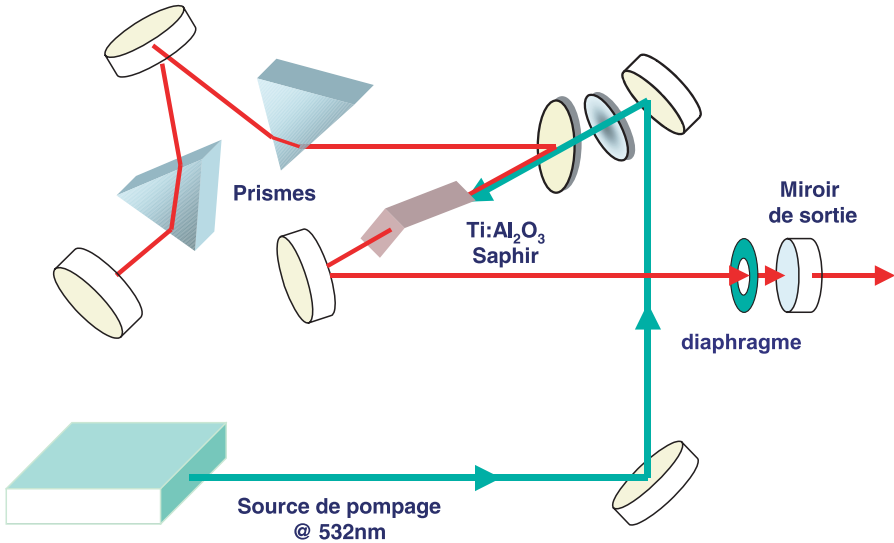


FIG. 4.15 – Schéma type d'un laser  $\text{TiAl}_2\text{O}_3$  délivrant des impulsions femtosecondes. Les prismes insérés dans la cavité permettent de compenser la dispersion spectrale de la vitesse de groupe qui affecte la propagation de l'onde optique à large spectre dans la cavité. Le verrouillage en phase des différents modes longitudinaux est réalisé par la création d'un effet Kerr optique dans le cristal  $\text{Ti} : \text{Al}_2\text{O}_3$ . L'effet de lentille associé à l'auto-focalisation optique contribue à réduire les pertes intra-cavité associées à la présence d'un diaphragme. Il en résulte l'émission d'une impulsion ultra-courte.

Pour bien des applications, il est nécessaire d'amplifier ces impulsions ultracourtes. C'est la technique d'étirement temporel, d'amplification et de compression qui permet de réaliser cette opération, tout en évitant une destruction des milieux optiques liée au fort éclaircissement (appendice B).

La technique d'amplification à dérive de fréquence consiste à injecter une impulsion courte dans une fibre optique dispersive qui assure la fonction d'étalement temporel. Chaque composante de fréquence se propage avec un temps différent lié à la dispersion chromatique du milieu. Ensuite on amplifie cette impulsion en l'injectant dans un amplificateur régénératif multi-passages. Lorsque l'on obtient le maximum d'extraction d'énergie, l'impulsion

est couplée à l'extérieur pour subir une opération de compression via une paire de réseaux diffractants. Ces réseaux ont pour fonction d'égaliser les chemins optiques des différentes composantes spectrales. Cette technique de compression d'impulsion, mise à profit dans le domaine Radar a été transposée dans le domaine optique avec succès [33].

En fonction de la durée de l'impulsion, la puissance crête d'un laser Terawatt peut correspondre à une énergie plus ou moins grande. Ainsi, pour une impulsion de 30 fs et une puissance crête de 1 TW, l'énergie contenue dans l'impulsion est de l'ordre de la trentaine de mJ. Ceci suppose qu'un soin particulier soit apporté à la conservation de la durée de l'impulsion durant les phases d'amplification et de la propagation. À partir d'un laser saphir-titane avec un encombrement réduit, l'ensemble laser et amplificateur peut être monté sur une table d'expérience (Fig. 4.16a).



FIG. 4.16 – *Source laser Alpha 10 : laser saphir titane délivrant 20 TW (1,2 J en 55 fs à la cadence de 10 Hz) (Thales Laser). (a) L'ensemble du dispositif sur une table optique, (b) cloche sous vide contenant le compresseur d'impulsion.*

En revanche si la durée de l'impulsion est plus élevée, par exemple 100 fs, l'énergie contenue dans l'impulsion doit être plus grande pour le même niveau de puissance crête. Pour une puissance crête de 100 TW, obtenue avec des lasers verre dopé à l'ion néodyme l'énergie doit être de 10 J. Dans ce cas, la source laser devient plus volumineuse et elle rejoint la classe des grands instruments. À titre indicatif, la focalisation d'un faisceau à distribution gaussienne, sur une surface délimitée par un rayon de 100  $\mu\text{m}$  et véhiculant une puissance crête de 100 TW permet d'obtenir un éclairage de 637  $\text{PW}/\text{cm}^2$ .

La vitalité de la communauté scientifique dans ce domaine a permis à certains industriels de développer des produits, dont la figure 4.16 est une illustration.

Les avancées observées autour des milieux solides dopés à l'ytterbium, à large spectre d'émission de fluorescence ouvrent de nouvelles perspectives de développement aux sources femtosecondes, car ils se prêtent à un pompage direct par diode laser.

Les principe de l'amplification d'impulsions à dérive de fréquence, de compression d'impulsions, permettent aujourd'hui de concevoir des sources laser Terawatt ( $10^{12}$  W) et Petawatt ( $10^{15}$  W), voire dans le futur Zettawatt ( $10^{18}$  W) et Exawatt ( $10^{21}$  W). L'existence de tels outils vient enrichir des expérimentations de plus en plus sophistiquées [34, 35] :

- accélération de particules ;
- source de rayonnement X ;
- ionisation multiphotonique dans les gaz ;
- initiation de fusion thermonucléaire par allumage rapide ;
- plasmas ultra-chauds et ultra-denses (ou comment réaliser sur terre l'équivalent de l'explosion d'une super-Nova) ...

En France, les travaux menés dans les laboratoires LULI de l'École Polytechnique et LOA de l'École Polytechnique-ENSTA ainsi que ceux du CEA sont des références dans ce domaine.

#### 4.4.6 Le laser et l'optique non linéaire

Un an après la démonstration expérimentale de l'effet laser à partir d'un cristal de rubis, P. Franken et ses collègues [15] observent le doublement de fréquence optique dans un cristal de quartz. Cette réalisation ouvre la voie à l'exploration d'un nouveau domaine : l'optique non linéaire [17]. Nous allons décrire de façon simple, donc non rigoureuse les mécanismes associés à ce type d'interaction.

L'interaction d'une onde électromagnétique avec un milieu induit une polarisation électrique :

$$\vec{P}_{\text{pol}}(t) = \varepsilon_0 \left[ \chi^{(1)} + \chi^{(2)} \vec{E}(t) + \chi^{(3)} \vec{E}(t) \cdot \vec{E}(t) + \dots \right] \cdot \vec{E}(t) \quad (4.21)$$

$\varepsilon_0$  est la constante diélectrique du vide,  $\chi^n$  représente la susceptibilité du  $n$ -ième ordre, qui est un tenseur de rang  $n + 1$  à  $3^{(n+1)}$  éléments. Pour une faible amplitude du champ électrique  $\vec{E}$ , équivalent à un régime de petit signal, cette polarisation induite suit le champ exciteur et rayonne une onde de même fréquence. Le terme  $\chi^{(1)}$ , susceptibilité linéaire complexe du milieu, décrit l'absorption et l'indice de réfraction optique du milieu. On décrit ainsi le régime de l'optique linéaire auquel s'applique le principe de superposition des champs sans transfert d'énergie entre eux. En présence d'un champ électrique important, donc pour un fort éclairage optique, les composantes non linéaires contribuent à modifier cette polarisation électronique. Les termes  $\chi^{(2)}$  et  $\chi^{(3)}$  représentent les susceptibilités non linéaires du second et du troisième ordre. La susceptibilité du second ordre est non nulle pour les matériaux non centro-symétriques. En revanche, tous les milieux présentent une non linéarité

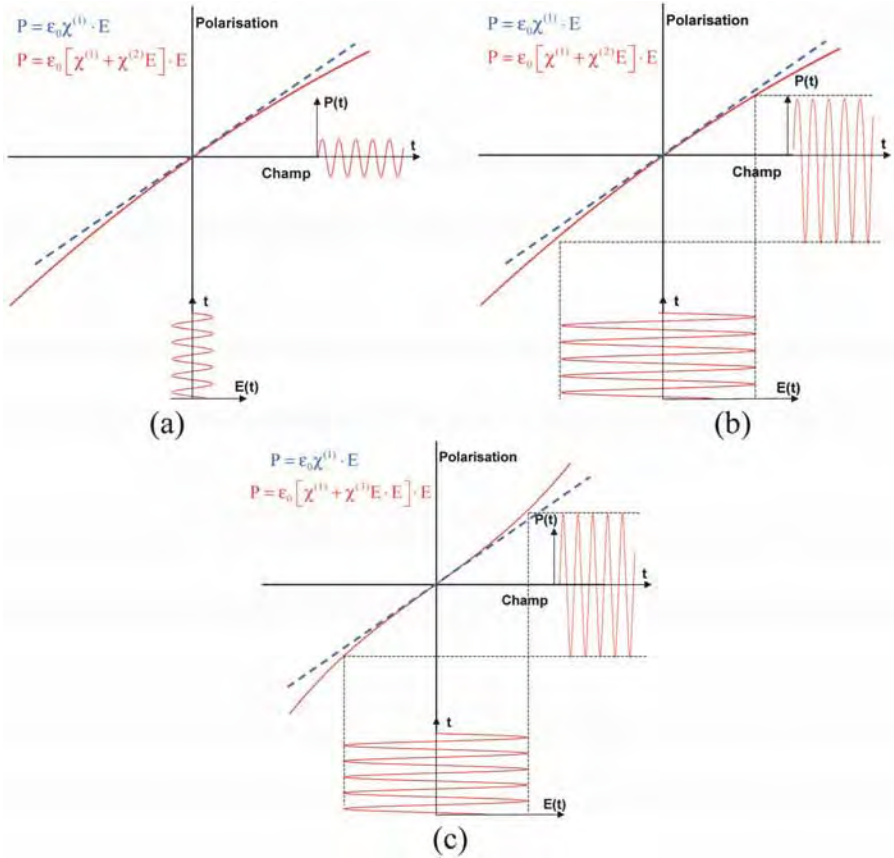


FIG. 4.17 – Dépendance de la polarisation induite en fonction de l'amplitude du champ optique excitateur en régime linéaire (a) et en régime non linéaire (b) pour des matériaux non centro-symétriques, (c) ou isotropes et centro-symétriques.

du troisième ordre dont les effets peuvent être observés sous fort éclaircissement (effet Kerr optique, auto-modulation de phase, conjugaison de phase ...). Ces non linéarités contribuent à un transfert d'énergie entre les ondes qui interagissent dans le milieu de propagation, qui se traduit par du mélange de fréquences. Pour une non linéarité d'ordre  $n$ , nous avons  $n + 1$  ondes qui peuvent interagir entre elles. La figure 4.17 illustre l'évolution de la polarisation induite en fonction de l'amplitude du champ électrique associée au champ optique.

À titre d'exemple, on considère deux ondes :  $E_1(t) = E_0 \cos(\omega_1 t)$  et  $E_2(t) = E_0 \cos(\omega_2 t)$  en interaction avec un milieu qui présente une non linéarité du second ordre. La polarisation non linéaire est proportionnelle au

champ total  $E(t) = E_1(t) + E_2(t)$  :

$$P_{\text{NL}}(t) \propto \chi^{(2)} E(t)^2 = \frac{\chi^{(2)} E_0^2}{2} (2 + \cos(2\omega_1 t) + \cos(2\omega_2 t) + \cos((\omega_1 + \omega_2)t) + \cos((\omega_1 - \omega_2)t))$$

On observe la génération d'une composante de pulsation nulle qui correspond à une rectification optique, une composante harmonique pour chacun des champs incidents, associée à un doublement de fréquence, une composante de pulsation  $\omega_1 + \omega_2$  qui traduit une sommation de fréquences et une composante  $\omega_1 - \omega_2$  qui décrit une différence de fréquences.

Ces interactions se doivent de respecter les conditions de conservation de l'énergie et de la quantité de mouvement. Par exemple une opération de sommation de fréquence impose :

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_s \quad (4.22)$$

deux photons d'énergie  $\hbar\omega_1$  et  $\hbar\omega_2$  donnent naissance à un photon  $\hbar\omega_s$ , et les conditions d'accord de phase (conservation de la quantité de mouvement) :

$$\frac{n_1\omega_1}{c} + \frac{n_2\omega_2}{c} = \frac{n_s\omega_s}{c} \quad (4.23)$$

où  $n_j$  ( $j = 1, 2, s$ ) représente l'indice de réfraction à la pulsation  $\omega_j$ . L'accord de phase revient à obtenir les conditions d'interférences constructives entre les ondes incidentes et les champs rayonnants engendrés par l'interaction non linéaire. Cette condition est généralement obtenue en utilisant les propriétés de dispersion de biréfringence des matériaux ou en réalisant un quasi-accord de phase en inversant périodiquement l'orientation de domaines ferro-électriques ou les axes cristallographiques. Le quasi-accord de phase revient à inverser périodiquement le signe du coefficient non linéaire afin de compenser périodiquement le déphasage accumulé entre la polarisation induite et l'onde engendrée.

L'effet paramétrique peut également être réalisé. Il correspond à un processus de mélange à trois ondes. Il consiste à coupler au matériau non linéaire un faisceau de pompe de pulsation  $\omega_p$  qui amplifie une onde signal ( $\omega_s$ ) tout en engendrant une onde complémentaire ( $\omega_c$  : « idler »). Les conditions sur l'énergie et sur l'accord de phase s'écrivent :

$$\begin{aligned} \hbar\omega_p &= \hbar\omega_s + \hbar\omega_c \\ n_p\omega_p &= n_s\omega_s + n_c\omega_c \end{aligned}$$

Ce mécanisme de mélange de fréquences à trois champs correspond à la transformation d'un photon de pompe en deux autres photons, dont les fréquences respectives peuvent être modifiées en jouant sur les conditions d'accord de phase. Ce processus d'amplification paramétrique, en absence de signal incident, démarre sur du bruit. L'insertion du cristal dans un cavité permet de

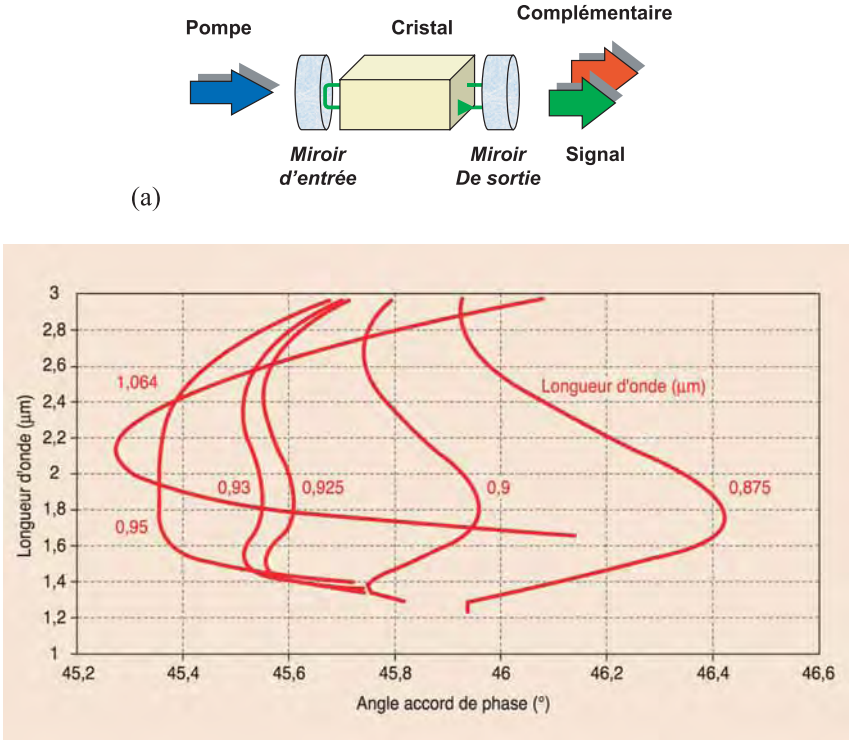


FIG. 4.18 – (a) Schéma d'un oscillateur paramétrique optique simplement résonnant. Le miroir d'entrée présente un maximum de transmission à la longueur d'onde de pompe et un maximum de réflectivité à la longueur d'onde signal. Le miroir de sortie présente un coefficient de réflexion maximum pour l'onde de pompe, ce qui permet un recyclage de celle-ci, et un coefficient de transmission ajusté à la longueur d'onde signal afin d'extraire de l'énergie de la cavité. L'onde complémentaire est libre, ce sont les conditions de résonance sur l'onde signal et la polarisation induite par l'onde de pompe qui génèrent cette onde complémentaire. (b) Longueurs d'onde émises en fonction de la longueur d'onde de pompe (rouge) et de l'angle d'accord de phase interne  $\theta$  en degré pour un OPO intégrant un cristal  $\text{LiNbO}_3$ . Pour chaque valeur de l'angle d'excitation du cristal, et pour une longueur d'onde de pompe donnée, on émet deux longueurs d'onde spécifiques. Pour ce matériau, on remarque la réalisation d'un accord de phase plat en utilisant un faisceau de pompe à la longueur d'onde de  $0,95 \mu\text{m}$ . Dans ces conditions, on transforme l'onde de pompe en un rayonnement à spectre large ( $1,7$  à  $2,3 \mu\text{m}$ ) par interaction paramétrique.

réaliser un oscillateur paramétrique optique (OPO, voir Fig. 4.18) qui présente la particularité d'avoir une émission spectralement accordable. Il suffit pour cela, de modifier l'orientation du cristal non linéaire, pour une taille cristallographique donnée, par rapport au faisceau de pompe, ou de faire varier la

température du cristal. Ces opérations reviennent à modifier les valeurs des indices de réfraction effectifs pour les trois ondes.

C'est le développement des laser pompés par diodes qui a permis une réelle exploitation des interactions nonlinéaires pour l'instrumentation scientifique et l'intégration dans des systèmes. La haute qualité spatiale du faisceau, la stabilité temporelle et la reproductibilité à l'émission de la source de pompage interagissant avec le milieu nonlinéaire, ont participé à dynamiser la mise en œuvre de ces méthodes pour réaliser des sources accordables sur un large domaine spectral ou pour effectuer des opérations de transposition de fréquences à haute efficacité [36].

En optique quantique, l'interaction paramétrique est un outil qui a permis l'observation de paires de photons corrélés, l'analyse des propriétés quantiques de photons jumeaux dont les états comprimés du champ [37].

## 4.5 Applications

### 4.5.1 Le domaine de la communication et de l'information

Les récentes applications des laser illustrent parfaitement comment un domaine émerge et se transforme en bénéficiant des avancées technologiques qui marquent l'évolution dans d'autres secteurs d'activités.

#### 4.5.1.1 Laser et stockage optique de données

Les premiers travaux sur le disque laser (vidéo disque) furent menés à THOMSON-CSF au Laboratoire Central de Recherches à Corbeville à la fin des années 1960 et au début des années 1970 [38]. Le concept mis en avant exploitait l'émission d'un laser He-Ne à 633 nm.

Dans cette mouvance, la société Spectra-Physics aux États-Unis développa une unité industrielle pour la fabrication en grande série de ce type de laser, qui sera par la suite reprise par la société Uniphase, devenue ensuite Uniphase-JDS.

La capacité d'inscription d'une information est directement liée à la longueur d'onde  $\lambda$ . Le rayon minimum d'une onde plane diffractée par une lentille de focalisation de diamètre  $D$  et de distance focale  $f$  (ou d'ouverture numérique O.N.) est le suivant :

$$r_{\min} = 1,22 \frac{f \lambda}{D} \approx \frac{0,61 \lambda}{\text{O.N.}} \quad (4.24)$$

Plus la longueur d'onde est courte plus la capacité d'inscription d'information est importante. On favorisera donc l'emploi de sources cohérentes qui émettent dans le visible.

C'est à partir de 1977 que le concept de CD (Compact Disc) prend forme avec les travaux initiés chez Philips et au Japon qui prennent comme élément

actif une diode laser GaAs/GaAlAs. Celle-ci a par ailleurs bénéficié d'un développement considérable avec l'émergence des télécommunications par fibres optiques. En 1979 on assiste à la première démonstration publique du CD comme nouveau support de diffusion d'œuvres artistiques. Une alliance est signée en 1980 entre Philips et Sony définissant un standard (notion importante en milieu industriel qui garantit un succès ou un échec) qui sera par la suite dénoncée.

À l'époque, le fait que l'on puisse lire une œuvre inscrite mais sans pouvoir l'enregistrer ne rend pas ce support compétitif par rapport au lecteur de vidéocassettes magnétiques. En revanche, la capacité de stockage, la durée de vie de stockage des informations inscrites, la compacité du support, l'inscription et la transcription numérique, la rapidité d'accès à une plage désirée rendent rapidement populaire ce nouveau moyen de diffusion d'œuvres musicales et par la suite cinématographiques. C'est en fait le développement des ordinateurs personnels qui a initialement contribué au lancement commercial du CD.

En 1983 le CD ROM (CD-Read Only Memory) est donc mis sur le marché, puis en 1993 apparaît le CD-R (CD-Recordable), suivi en 1997 par le CD-RW (CD-Rewritable). Les caractéristiques du support sont les suivantes :

- diamètre : 12 cm ;
- épaisseur : 1,2 mm ;
- vitesse de rotation angulaire : 200 à 500 t/minute ;
- capacité de stockage : 650 à 750 Mb/s.

La durée d'accès à un fichier inscrit et le débit de transfert pour un CD-ROM sont passés de 600 ms et 150 kb/s à 100 ms et 4,8 Mb/s pour un CD 32x (150 kb/s  $\times$  32 = 4,8 Mb/s) en une vingtaine d'années.

La technique d'enregistrement consiste à graver un film en polycarbonate en inscrivant une petite dépression à l'interface avec un film métallique en aluminium, qui fait office de réflecteur. Celui-ci est ensuite recouvert d'une fine couche de matériau acrylique qui vient le protéger de l'oxydation et des moisissures. Les disques gravés et commercialisés sont généralement réalisés par réplique sous pressage à partir d'un disque maître.

Le faisceau laser d'inscription émis par une diode laser à la longueur d'onde  $\lambda = 780$  nm fournit une puissance de quelques mW et la dimension du faisceau focalisé est de l'ordre du micron. En fonction de la rotation du disque, le faisceau d'inscription ou de lecture suit un déplacement en spirale. En lecture, le faisceau vient sonder l'interface métal-plastique et la variation de réflectivité traduit l'inscription d'une information binaire élémentaire.

Différentes techniques d'inscription ont été proposées dont l'effet magnéto-optique maintenant utilisé avec le DVD (Digital Versatile Disc). Il consiste à déposer sur un substrat un film de matériau magnétique amorphe (alliage de métaux de transition de terres rares TbFeCo, GdTbFe ...) caractérisé par une température de Curie de l'ordre de 200 °C. Ce matériau présente un



champ magnétique coercitif élevé à haute température et inversement à basse température. Ainsi, en présence d'un champ magnétique externe et au dessus du point de Curie, on obtient au préalable une orientation de tous les dipôles magnétiques dans le film. Cette orientation est ensuite conservée en repassant au-dessous de cette température de transition. On inverse ensuite l'orientation du champ magnétique. En présence d'une élévation de température localisée liée à l'illumination optique, on obtient un basculement de l'orientation des dipôles, qui est par la suite conservé en revenant à l'équilibre thermique. On inscrit ainsi une information dont les dimensions correspondent grossièrement à la géométrie du faisceau laser au point focal. La lecture optique de cette information photo-induite s'effectue en mesurant l'état de polarisation de la lumière réfléchie, sachant que le faisceau incident, émis par une diode laser, est naturellement polarisé. C'est l'asymétrie de la structure de guidage optique qui favorise l'existence d'un mode transverse électrique (TE). On exploite alors la variation de réflectivité qui dépend de l'orientation magnétique de la surface analysée. Généralement on mesure des variations inférieures à  $1^\circ$ . Cette technique permet d'inscrire, d'effacer et de réinscrire des motifs et de procéder à leur lecture.

Avec la maîtrise des diodes laser émettant dans le rouge (650 nm puis 635 nm), un forum entre industriels s'est tenu en 1995 et en 1997 pour le lancement du DVD. Cette filière permet non seulement d'accroître la capacité de stockage mais également de visualiser une image haute définition. La diffusion à grande échelle suivit, avec le développement du DVD Vidéo, du DVD Audio, du DVD ROM, du DVD R et enfin du DVD RAM, pour aboutir à la définition et à la conception du DVD-Multi standard en 2002 dont le diamètre est de 12 cm avec une capacité de stockage comprise entre 4,7 et 17 Go.

Un DVD pré-enregistré est composé de deux demi-disques assemblés ayant chacun une épaisseur de 0,6 mm. Pour un DVD-ROM, la capacité de stockage est de 4,7 Go par couche, avec un maximum de deux couches par face, soit pour les deux faces une capacité de 17 Go exploitable. Avec une optique de focalisation à grande ouverture (courte distance focale), la dimension du faisceau au point focal est de  $0,4 \mu\text{m}$ .

Le paradoxe veut que trente années après les premières démonstrations à partir d'un laser He-Ne, on aboutisse à l'industrialisation et à la commercialisation d'un produit qui exploite les propriétés de stockage à la même longueur d'onde de  $0,63 \mu\text{m}$ , maintenant.

D'autres méthodes sont également employées utilisant le changement de phase cristalline d'un matériau par voie thermique. Les DVD RW (Re-Writable), comme les CD RW réinscriptibles utilisent un alliage Ge, Sb, Te déposé sous la forme d'un film mince poly-cristallin qui fait office de couche inscriptible. Dès que l'on atteint la température de  $600^\circ\text{C}$  et en présence d'un refroidissement rapide, on obtient la formation d'une structure amorphe dont le pouvoir réflecteur est plus faible que celui associé à l'état polycristallin. Pour effacer l'information inscrite, le faisceau laser focalisé porte le film à une température de  $200^\circ\text{C}$  sur une durée plus élevée que celle associée à

l'écriture. Cette opération permet à l'alliage amorphe de retrouver son état polycristallin.

Après la démonstration d'une émission laser dans le bleu en 1993, la génération suivante (BD Blu-ray Disc ou HD-DVD (High Definition-DVD)) sera mise sur le marché grand public en 2006. Un consortium dénommé Blu-Ray comprenant les grands industriels du secteur de l'électronique grand-public (Hitachi, LG, Matsushita, Panasonic, Pioneer, Philips, Samsung, Sharp, Sony et Thomson) a jeté les bases du HD-DVD. Celui-ci utilisera une diode laser à 405 nm, un codage MPEG-2 et un débit de transfert des données compris entre 10 et 30 Mb/s. La capacité de stockage sera de 23 à 27 Go sur une seule face et sur une mono-couche. Le codage permettra de stocker un programme HDTV (haute définition) d'une durée de deux heures et plus de treize heures d'un programme TV standard. Si l'on transpose ce qui est développé autour du DVD en multipliant le nombre de couches et sur les deux faces du disque, on atteindra les capacités de 50 et 100 Go.

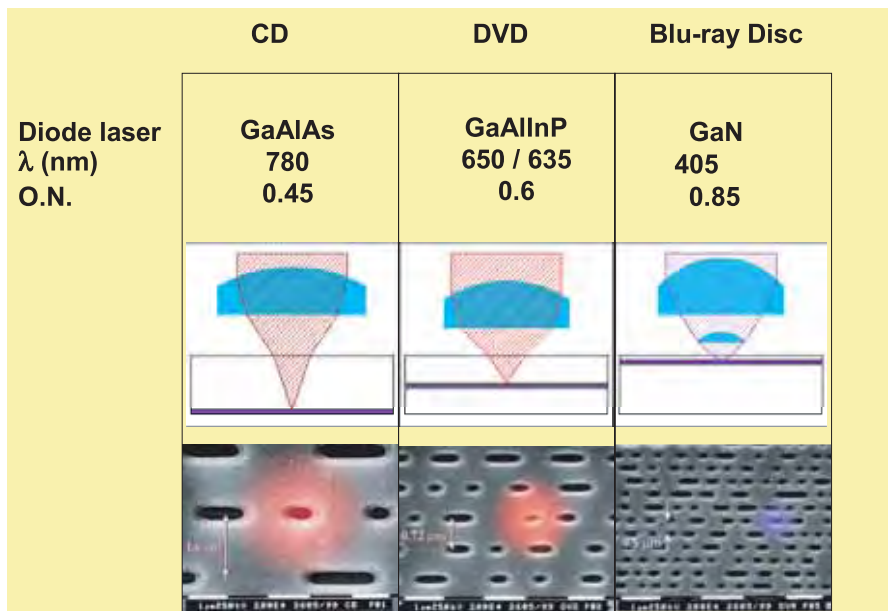


FIG. 4.19 – Caractéristiques comparées montrant l'évolution des différentes générations de systèmes de stockage optique (à partir de données Philips Research).  $\lambda$  est la longueur d'onde d'émission de la diode laser et O.N. est l'ouverture numérique de l'optique de focalisation.

L'évolution naturelle de cette filière consistera à passer du bleu à l'ultra-violet (UV), domaine accessible avec les alliages de matériaux AlGaIn.

Dans le futur, les travaux menés actuellement sur les disques à multiples couches de molécules fluorescentes (FMD-ROM Fluorescent Multi-layer Disc)

laissent entrevoir l'obtention d'une capacité de stockage de 50 à 140 Go par une autre méthode jugée plus simple sur une vingtaine de couches pour une épaisseur de 2 mm. L'holographie en volume est également, depuis plus de trente ans, une technique toujours étudiée qui permettrait de stocker 1 Téraoctets dans un volume ( $1 \text{ cm}^3$ ), correspondant à un morceau de sucre.

Avant de conclure cette partie, nous devons mentionner quelques caractéristiques associées au disque dur magnétique que l'on trouve dans les ordinateurs personnels et qui ont permis, par une fertilisation mutuelle, le développement des lecteurs et graveurs optiques.

La vitesse de défilement d'un disque dur (à vitesse angulaire constante) par rapport à la tête de lecture est entre 3600 tours/min et 7200 tours/min. La vitesse correspondante est de l'ordre de 100 km/h. On mesure l'exploit lorsque l'on prend en compte la densité d'information stockée qui dépend de la distance entre la tête de lecture et le support magnétique. Pour atteindre les densités de plusieurs dizaines de Go, cet espacement est entre 10 et 20 nm (cinq mille fois plus petite que celle d'un cheveu de diamètre 50 à 100  $\mu\text{m}$ ).

On ne doit pas ignorer que d'autres mécanismes physiques peuvent être mis en jeu pour le stockage à très grande densité. Parmi ceux-ci on peut mentionner les travaux menés à IBM et qui exploitent pour le stockage d'information les techniques dérivées de la microscopie à force atomique, les mémoires moléculaires qui représentent le stockage ultime en traitant deux états distincts (diamagnétique ou paramagnétique, paramagnétique ou ferromagnétique) pouvant être pris par une molécule unique sous l'effet d'un champ magnétique ou sous l'effet d'une illumination.

Enfin l'état solide avec les MRAM (Magnetic Random Access Memory) ayant pour fondement le contrôle des propriétés de transport des électrons via l'interaction entre leur spin et l'aimantation d'un matériau ferromagnétique, a déjà démontré son aptitude à augmenter les capacités de stockage des disques durs. Cette application résulte de travaux sur la magnétorésistance géante [39] créant une nouvelle discipline la « spintronique » qui ne fait que commencer son parcours avec les matériaux nanostructurés et qui doit ouvrir de nouvelles perspectives dans les domaines du stockage, du traitement de l'information et des capteurs.

#### 4.5.1.2 Communications en propagation libre

Les propriétés de faible divergence et de courte longueur d'onde font des sources laser d'excellent candidats pour réaliser des liaisons de transmissions de données entre satellites ou pour transmettre des informations montantes ou descendantes avec des sondes satellisées.

L'avantage attaché à l'utilisation de faisceaux laser, par rapport aux faisceaux hyperfréquences, est lié à l'utilisation d'une antenne (téléscope) de plus petite dimension que celle devant être déployée avec une onde hyperfréquence. On obtient également un gain sur la masse et sur l'encombrement du système

embarqué, tout en réduisant les effets d'interférences avec les autres canaux de communication.

Dans l'espace, l'absence de phénomènes de turbulence atmosphérique contribue à préserver la qualité du plan d'onde du faisceau optique émis. En outre, il n'y a pas à prendre en compte les critères de sécurité oculaire qui, sur Terre impliquent des restrictions sur le choix de la longueur d'onde pour réaliser des liaisons point à point.

C'est le 20 novembre 2001 qu'était réalisée pour la première fois une liaison optique entre le satellite de télécommunication ARTEMIS (Advanced Relay and TEchnology MISSION développé par l'Agence Spatiale Européenne ESA), localisé en orbite quasi-géostationnaire à 31 000 km autour de l'équateur avec le satellite d'observation SPOT 4 (Satellite Probatoire d'Observation de la Terre développé par le CNES) situé en orbite basse à 832 km. Les données recueillies par SPOT sont communiquées par voie laser, via le système SILEX (Semiconductor-laser Inter-satellite Link Experiment, développé par ASTRIUM), vers ARTEMIS qui transmet par voie conventionnelle les informations vers une station au sol.

Le système SILEX est composé de deux terminaux optiques : PASTEL (PASsager TELEcom) sur SPOT 4 et OPALE (Optical PAYload for inter satellite Link Experiment) sur ARTEMIS (Fig. 4.20).



FIG. 4.20 – *Vue artistique de la liaison entre le satellite géostationnaire ARTEMIS et le satellite d'observation de la terre SPOT 4 (ESA-CNES).*

Une diode laser GaAs/GaAlAs émettant au voisinage de  $0,82 \mu\text{m}$  est modulée à 50 Mb/s (dans le sens SPOT vers ARTEMIS) et fournit une puissance moyenne de 60 mW. La liaison ARTEMIS vers SPOT s'effectue avec un débit

en ligne de 2 Mb/s. La masse du système est de 150 kg et le diamètre du télescope est de 25 cm.

La liaison a permis de recueillir une image vidéo haute définition numérisée à 50 Mb/s avec un taux d'erreur inférieur à  $10^{-9}$ . La précision sur le pointage qui doit être maintenue entre l'optique d'émission et celle de réception est de l'ordre du  $\mu\text{rad}$  ( $\pm 30$  m à 30 000 km soit la dimension d'une pièce de monnaie de 1 cm à 10 km). On mesure la difficulté qui devait être surmontée lorsque l'on voit que le temps de parcours du signal sur cette distance est de l'ordre 0,1 s et que le déplacement relatif entre les deux satellites s'effectue à la vitesse de 7 km/s.

Plus récemment, en septembre 2003, une expérience a réussi à tester une liaison optique terre-satellite géostationnaire sur une distance de 38 000 km entre ARTEMIS et une station de l'ESA à Tenerife dans les îles Canaries en utilisant le système LUCE (Laser-Utilizing Communications Equipment) du programme OICETS (Optical Inter-Orbit Communications Engineering Test Satellite) de l'Agence d'exploration spatiale japonaise.

Une expérience similaire classifiée avait été menée en 2001 aux États-Unis avec le satellite géostationnaire GEOLite.

Historiquement, dès l'apparition du laser, son emploi pour transmettre des données à distance et en vue directe a suscité un immense intérêt. Très vite se sont posés des problèmes liés à la propagation d'un faisceau optique dans l'atmosphère (effets de turbulence, diffusion liée aux aérosols ...). Le laser  $\text{CO}_2$ , avec sa longueur d'onde d'émission dans l'infra-rouge à  $10,6 \mu\text{m}$  semblait assez bien adapté vis-à-vis des effets liés aux turbulences et à la diffusion tout en bénéficiant d'une excellente fenêtre de transmission de l'atmosphère ; toutefois la faible sensibilité des systèmes de réception à cette longueur d'onde venait réduire la distance des liaisons pouvant être imaginées. On doit rappeler que les travaux menés au CNET dans les années 1970 sur l'optique non linéaire avaient pour but de réaliser les premières démonstrations de multiplexage spectral en doublant la fréquence d'un laser  $\text{CO}_2$  (opération remise au goût du jour par la suite en devenant incontournable avec le développement des transmissions sur fibres optiques). On disposait ainsi de deux porteuses aux longueurs d'onde  $10,6 \mu\text{m}$  et  $5,3 \mu\text{m}$ .

L'arrivée des liaisons par fibres optiques a bien entendu mis un terme à de tels projets. Dans ce cas, le milieu de transmission, tout en offrant de faibles pertes, est relativement insensible à l'environnement. Cependant, avec les développements observés autour des diodes laser à cascade quantique et des détecteurs à multiples puits quantiques (QWIP : Quantum Well IR Photodetector), ce mode de transmission dans l'infrarouge et en propagation libre est revenu d'actualité [40]. En particulier, dans les villes ou quartiers qui présentent un environnement à forte concentration d'habitations, les coûts d'installation de liaisons filaires ou les difficultés d'accès représentent un problème. Dans certaines configurations architecturales, les effets de réverbération entre immeubles perturbent la propagation et la réception d'ondes radio ou hyperfréquences. Pour ces raisons, les communications optiques offrent une

alternative séduisante et la liaison point à point en propagation libre (Télégraphe de Chappe optoélectronique ou « last mile by laser ») émerge à nouveau. Le caractère discret de la liaison est aussi mis à profit pour certaines applications dans le domaine de la défense.

## 4.5.2 Le domaine de l'instrumentation

Le caractère monochromatique de l'émission laser a grandement influé sur la métrologie, d'une part pour la spectroscopie haute résolution en physique atomique et moléculaire, d'autre part pour des expériences complexes nécessitant une très grande précision, comme la détection des ondes gravitationnelles par interférométrie (programmes VIRGO ou LIGO). Le domaine des capteurs a également bénéficié des propriétés de cohérence temporelle des lasers.

### 4.5.2.1 La télémétrie

Les sources laser présentent la particularité d'avoir un rayonnement lumineux émis d'une grande pureté spectrale (monochromaticité) par leur cohérence temporelle et une grande directivité par leur cohérence spatiale. En conséquence, selon le caractère du faisceau émis, une grande diversité d'applications a vu le jour.

Ainsi, le caractère directif d'un faisceau laser permet de réaliser :

- le fil à plomb laser. Un fil à plomb un peu particulier dans la mesure où il n'est pas gouverné (au premier ordre) par les effets de la gravitation. On se sert alors du caractère directif pour réaliser un alignement qui facilite la construction de grandes infrastructures (l'élévation de la Tour Montparnasse à Paris a utilisé quatre faisceaux laser He-Ne), qui établit un trajet rectiligne grâce au niveau laser (guidage pour le creusement du tunnel sous la Manche) ;
- des opérations de télémétrie qui permettent d'évaluer la distance  $L$  qui sépare l'émetteur d'une cible coopérative. Plusieurs méthodes peuvent être employées. La première consiste à mesurer le temps de vol d'une impulsion laser. La différence de temps  $\Delta\tau$  entre le temps référence lié à l'émission de l'impulsion et celui correspondant à la détection de l'impulsion rétrodiffusée par la cible correspond à la relation :

$$\Delta\tau = \frac{2Ln}{c} \quad (4.25)$$

où  $n$  est l'indice de réfraction optique moyen du milieu de propagation et  $c$  est la célérité de la lumière dans le vide.

La mesure de la distance terre-lune s'effectue de cette manière à l'Observatoire de la Côte d'Azur (Fig. 4.21), avec une précision de 3 mm en utilisant les réflecteurs catadioptriques (coin de cube) déposés sur notre plus proche planète au cours des différentes missions spatiales lunaires américaines (Apollo



FIG. 4.21 – Photographie du site de l'Observatoire de Cote d'Azur et du dôme accueillant le télémètre laser mesurant la distance Terre-Lune (laser Nd:YAG doublé en fréquence).

XI, XIV, XV) et soviétiques (Lunakhod 1 et 2). En moyenne on détecte un photon tous les cent tirs. En considérant que la distance moyenne Terre-Lune est de 384 400 km, l'intervalle de temps entre l'émission et le retour d'une impulsion laser est de  $\sim 2,563$  s.

Dans ce même domaine, on trouve d'autres applications en géodésie (mesure de la dérive des continents, du positionnement de satellites, altimétrie des

océans avec le programme JASON-I (NASA-CNES), évolution de la croûte terrestre ...).

Les cinémomètres laser déployés en sécurité routière exploitent également les faisceaux lumineux réfléchis par un véhicule, ici en mouvement. Ils utilisent une diode laser impulsienne qui émet un rayonnement à la longueur d'onde de 904 nm. La mesure d'une variation temporelle de  $\Delta\tau$ , lorsque la distance  $L$  entre le véhicule et le détecteur varie, permet de remonter à la vitesse  $V$ .

$$\frac{d}{dt}\Delta\tau = \frac{2n}{c} \frac{dL}{dt} = \frac{2n}{c} V$$

Cette méthode présente un caractère rustique (par rapport à une méthode anémométrique qui exploite l'effet Doppler), tout en restant compatible avec l'objectif visé (résolution 0,1 km/h et gamme de mesure entre 0 et 250 km/h).

Une autre concept de télémétrie utilisant le laser consiste à mesurer la variation de phase entre le signal optique émis modulé en amplitude par une sinusoïde de pulsation  $\Omega$  avec un taux de modulation  $m$  :

$$P_e(t) = P_0(1 + m \sin(\Omega t + \phi))$$

et celui détecté après un parcours aller et retour :

$$P_r(t) = \alpha P_0(1 + m \sin(\Omega(t - \Delta\tau) + \phi))$$

où  $\alpha$  caractérise les propriétés d'albedo de la cible, les pertes à la propagation du faisceau dans l'atmosphère et l'efficacité de l'optique de collection au niveau du récepteur. La différence de marche entre les signaux émis et reçu introduit une différence de phase :

$$\Delta\phi = \Omega\Delta\tau = \Omega \frac{2Ln}{c} \quad (4.26)$$

On observe que la mesure présente une ambiguïté car  $\Delta\phi$  varie de  $2\pi$  chaque fois que  $L \times n = \frac{\pi c}{\Omega} = \frac{\Lambda}{2}$ .  $\Lambda$  représente la longueur d'onde associée à la fréquence de du signal de modulation. Cette relation montre qu'il est possible de mesurer un déphasage entre deux signaux que modulo  $2\pi$ . En conséquence, la portée du télémètre sera gouvernée par la fréquence de modulation du faisceau laser. Pour une fréquence de modulation de 100 kHz, un déphasage de  $2\pi$  correspond à une distance de 1500 m. Au-delà de cette distance, et pour cette fréquence de modulation on obtient une ambiguïté sur la localisation de la cible.

Une autre méthode consiste à analyser la fréquence instantanée d'un signal optique modulé. La source optique émet un signal modulé linéairement en fréquence dans le temps entre les valeurs  $f_1$  et  $f_2$  sur une période  $T$  (Fig. 4.22). La fréquence de modulation instantanée en retour de la cible  $f_r(t)$  est décalée par rapport à la fréquence émise  $f_e(t)$ , d'un intervalle de temps  $\Delta\tau$ . Le photodétecteur à la réception délivre donc un photocourant dont le spectre



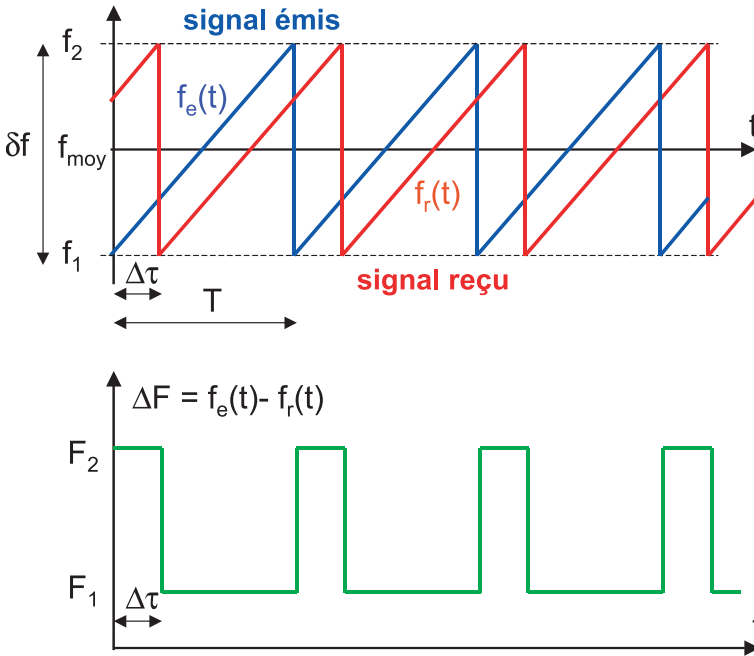


FIG. 4.22 – Principe de fonctionnement d'une télémétrie à modulation de fréquence.

en fréquence contient un signal de battement  $\Delta F = |f_e(t) - f_r(t)|$  qui varie avec le temps entre deux valeurs  $F_1$  et  $F_2$ .

Tant que  $t$  est dans l'intervalle  $[\Delta\tau, T]$  avec  $t$  modulo  $T$ , nous avons :

$$\Delta F = F_1 = \delta f \frac{\Delta\tau}{T}$$

$$L = \frac{c}{2n} T \frac{F_1}{\delta F}$$

Pour atteindre de grandes sensibilités, sur de courtes distances, ces techniques sont transposées aux fréquences optiques en modulant le courant d'une diode laser et en analysant le signal de battement optique.

Avec les moyens de calcul et de stockage de données de plus en plus performants, la télémétrie par temps de vol voit son domaine d'application étendu à la profilométrie et à la restitution tridimensionnelle de scènes. Ainsi, la mission MOLA (Mars Observer Laser Altimeter), intégrant un laser Nd:YAG pompé par diodes opérant en mode déclenché, a permis de réaliser un relevé topographique de la planète Mars. La restitution du pôle Nord de Mars (Fig. 4.23) a nécessité le traitement de 2,6 millions de tirs laser [41, 42].

Une autre illustration de l'emploi de la télémétrie laser concerne le relevé topographique de terrain par balayage de faisceau laser. Cette technique permet de restituer un relief qui contribue à observer l'évolution éventuelle d'un

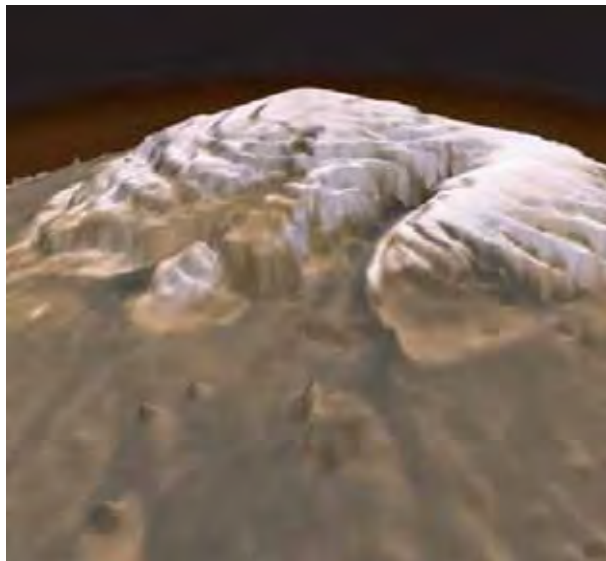


FIG. 4.23 – Reconstitution topographique 3D du pôle Nord de la planète Mars par télémétrie laser (laser solide Nd:YAG pompé par diodes). Mission MOLA-2 (Mars Orbiter Laser Altimeter) (NASA).

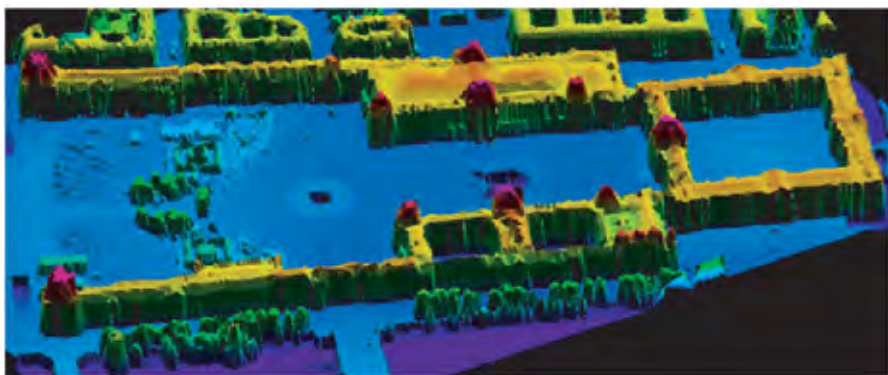


FIG. 4.24 – Relevé 3D du Louvre par profilométrie laser aéroportée [43]. (laser Nd:YAG pompé par diodes à fréquence de répétition de quelques kHz). (IGN).

environnement. Le positionnement absolu du laser aéroporté est déterminé en couplant les données d'un système GPS (Global Positioning System) et celles d'une centrale inertielle (gyromètres). L'ensemble donne par échométrie laser les coordonnées tri-dimensionnelles du relief (Fig. 4.24) [43].

Les forces spéciales peuvent également disposer de banques de données restituant l'extérieur et l'intérieur d'un bâtiment ou d'une pièce particulière

pouvant les aider à la neutralisation d'un forcené ou d'une action terroriste par exemple [44].

### 4.5.2.2 Anémométrie laser

L'anémométrie laser est une méthode de mesure non-invasive de vitesse d'écoulement d'un fluide. Le principe repose sur la formation en un point de l'espace d'un champ de franges d'interférences. En présence de particules animées d'une vitesse, la traversée de ce réseau de franges s'accompagne d'une modulation de la lumière diffusée. Ce réseau d'éclairement est obtenu en divisant un faisceau laser en deux sous-faisceaux, qui se superposent à la distance souhaitée sous un angle  $2\theta$  (Fig. 4.25).

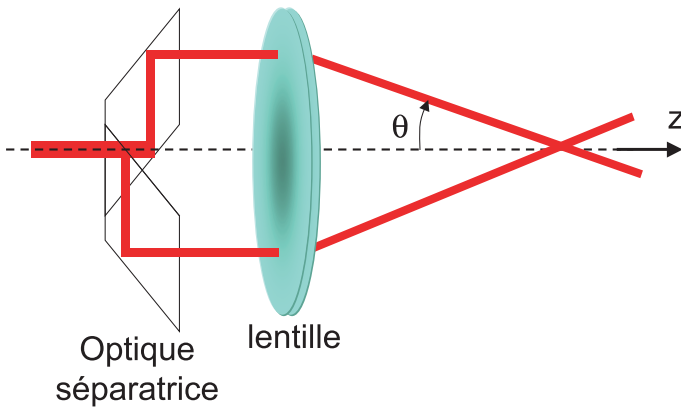


FIG. 4.25 – Structure élémentaire d'un montage optique permettant de réaliser à distance et localement un réseau transversal de franges d'éclairement.

Nous allons considérer deux faisceaux issus d'une source monofréquence avec une distribution transverse gaussienne du champ. Le champ électrique associé à un faisceau gaussien peut s'écrire [45] :

$$E_j(x', y', z') = E_{0_j} \left( \frac{q_{0_j}}{q_{0_j} + z'} \right) \exp \left( -ik \left( \frac{x'^2 + y'^2}{2(q_{0_j} + z')} \right) + i(\omega t - kz') \right) \tag{4.27}$$

$j = 1, 2$  est attaché à chacun des faisceaux,  $q_{0_j} = i\pi w_{0_j}^2 / \lambda$  est lié à la longueur de Rayleigh,  $w_{0_j}$  est le col (« waist ») du faisceau gaussien initial à  $1/e^2$ ,  $k = 2\pi/\lambda$ . La transformation sur le système de coordonnées :

$$\begin{aligned} x' &= x \\ y' &= y \cos \theta - z \sin \theta \\ z' &= z \cos \theta + y \sin \theta \end{aligned} \tag{4.28}$$

permet d'établir la distribution du champ total en appliquant le principe de superposition en présence de deux ondes polarisées selon l'axe  $x$ . En présence

de ces deux ondes, on obtient la distribution d'éclairement moyennée sur le temps :

$$I(x, y, z) = \langle (E_1 + E_2)(E_1^* + E_2^*) \rangle \quad (4.29)$$

Si la longueur d'interaction des deux faisceaux gaussiens  $l_{\text{int}} = 2w_0/\sin(2\theta)$  est plus petite que la longueur de Rayleigh  $|q_0|$  nous pouvons considérer que  $q_{0j} + z' \approx q_{0j}$ . Dans ces conditions, l'éclairement dans l'air s'écrit :

$$I(x, y, z) = \frac{I_0}{2} \left( \cos \left( \frac{4\pi}{\lambda} y \sin \theta \right) + \cosh \left( \frac{2yz \sin(2\theta)}{w_0^2} \right) \right) \times \exp \left( -2 \left( \frac{x^2 + y^2 \cos^2 \theta + z^2 \sin^2 \theta}{w_0^2} \right) \right) \quad (4.30)$$

qui définit un volume sur lequel on obtient une modulation spatiale de l'éclairement (Fig. 4.26). À partir de l'argument de la fonction cosinus, on observe que l'interfrange (espacement entre franges) selon l'axe  $y$  est donné par l'expression :

$$\Lambda = \frac{\lambda}{2 \sin \theta} \quad (4.31)$$

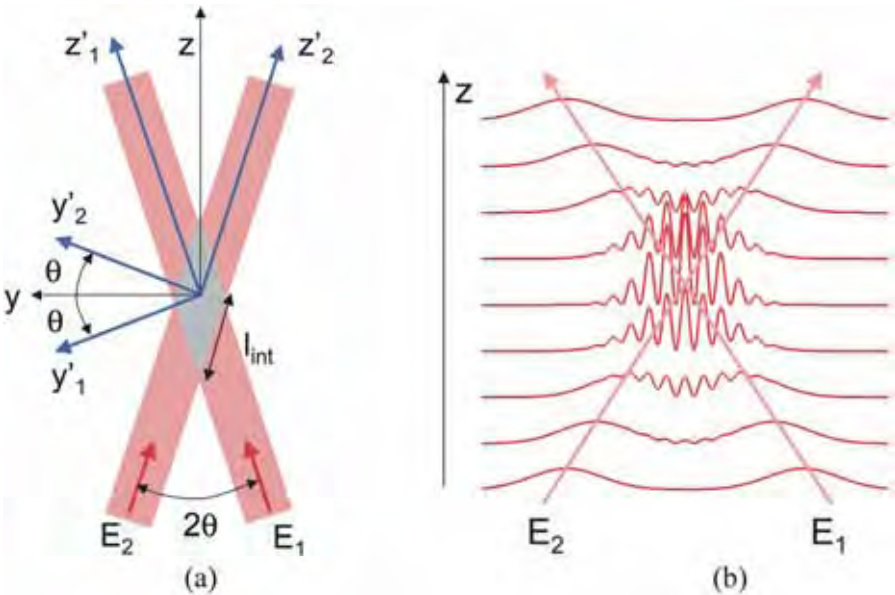


FIG. 4.26 – (a) Illustration de la figure d'interférence créée par le croisement de deux ondes sous une incidence de demi-angle  $\theta$ . (b) Distribution de l'éclairement dans un plan  $xy$  perpendiculaire à l'axe  $z$ , pour différentes valeurs de  $z$ .

En présence d'une particule qui traverse le champ d'interférences avec une vitesse  $v$ , on obtient une modulation temporelle de la lumière diffusée ou

diffractée. Lorsque la particule en suspension dans le fluide traverse la zone d'interférences, elle voit alternativement des franges sombres et brillantes. Dans une zone brillante elle diffuse la lumière et lorsqu'elle se trouve dans une zone sombre la diffusion lumineuse est réduite. La variation du signal détecté par une photodiode est caractérisée par une fréquence de modulation  $F_m$  qui, avec l'interfrange fixée par construction, donne la composante selon  $y$  de la vitesse de la particule (projection de la vitesse de la particule sur la normale au plan des franges) :

$$v_y = F_{\text{mod}} \Lambda \quad (4.32)$$

La gamme des vitesses que l'on peut mesurer est étendue, allant de quelques microns jusqu'à des centaines de mètres par seconde. L'inconvénient d'un tel vélocimètre laser à franges est de ne pas indiquer le sens de passage de la particule. Toutefois, en couplant le mécanisme de création de franges d'interférences avec l'effet Doppler qui affecte la lumière diffusée par la particule, on peut accéder au signe de la vitesse de déplacement de la particule.

Ce type d'instrument permet de cartographier la distribution de vitesse d'un fluide en l'ensemencant de particules. Il est également utilisé en avionique pour avoir accès au déplacement d'un aéronef ou pour étalonner des capteurs embarqués sur avion (tube de pitot). Dans ce cas, on considère le déplacement des aérosols naturels d'environnement comme étant quasi-stationnaire par rapport au mobile.

#### 4.5.2.3 Gyrométrie laser

Le mécanisme d'un gyrolaser est fondé sur l'étude des modifications de la polarisation d'un milieu actif inséré dans une cavité optique résonnante dans laquelle se développent deux ondes contra-propagatives.

Cette architecture permet de mettre à profit l'effet Sagnac [46,47] qui relie une variation de fréquence à une rotation du système interférométrique qui, dans ce schéma, inclut un milieu laser. Rappelons que dès l'avènement des sources cohérentes optiques, le gyromètre laser a fait l'objet de travaux expérimentaux [48–50]. Le schéma type d'un gyromètre laser à gaz est représenté sur la figure 4.27.

Le principe de base repose sur le fait que dans une cavité en anneau immobile, la pulsation du champ associée à un mode longitudinal est donnée par la relation :

$$\omega_m = 2\pi m \frac{c}{L} \quad (4.33)$$

La longueur de cavité est un multiple entier de la longueur d'onde :  $L/\lambda_m = m$ .

Si la cavité est animée d'un mouvement de rotation de vitesse angulaire  $\Omega$  autour d'un axe perpendiculaire au plan de la cavité, les ondes co-propagative et contra-propagative subissent l'effet Sagnac. Cet effet revient à modifier la distance parcourue par les deux ondes ( $+\delta L$  pour l'onde copropageante et  $-\delta L$  dans l'autre direction). Il s'accompagne d'une modification des pulsations

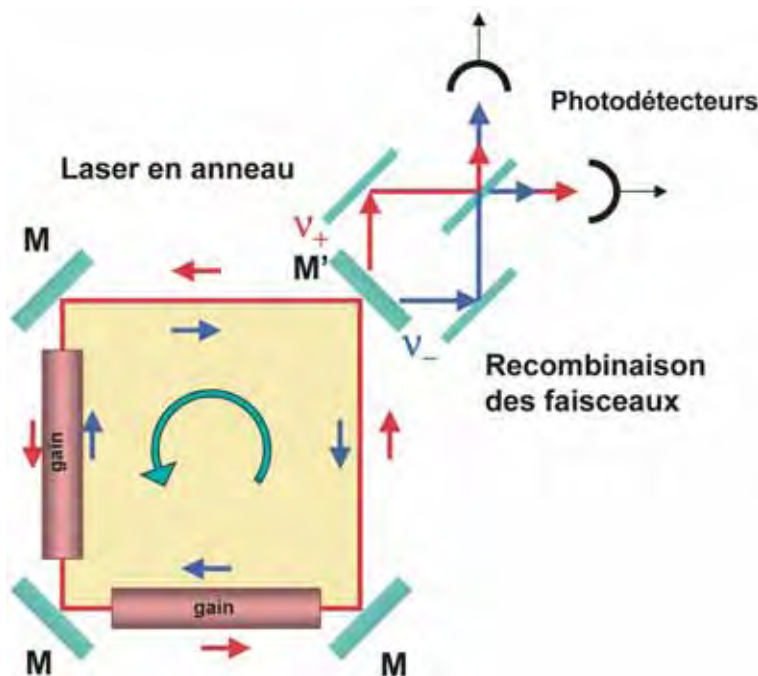


FIG. 4.27 – Schéma de principe d'un gyromètre laser. Il se compose d'un résonateur en anneau constitué de quatre miroirs, dont  $M'$  est semi-transparent de manière à légèrement coupler l'émission laser bi-directionnelle vers l'extérieur. Dans deux bras se trouve un milieu à gain. Les deux ondes qui se propagent dans les deux sens sont recombinées et donnent une figure d'interférence. Le battement de fréquence entre les deux ondes qui circulent dans les deux sens mesure la rotation du système.

associées aux champs selon leur sens de propagation dans la cavité (augmentation de la pulsation dans la direction opposée à la rotation et diminution de la pulsation dans la direction du mouvement de rotation) :

$$\Delta\omega = \omega_m^- - \omega_m^+ = \frac{2\omega_m\delta L}{L} = \frac{4S\omega_m}{Lc}\Omega \quad (4.34)$$

où  $S$  est la surface délimitée par le parcours suivi par l'onde et  $L$  est la longueur optique de la cavité (périmètre de la cavité). La fréquence de battement entre les deux ondes diffère alors de :

$$\Delta\nu = \frac{4S}{L\lambda}\Omega \quad (4.35)$$

La fréquence de battement donne la vitesse angulaire, alors qu'un comptage du nombre de franges qui défile donne la position angulaire. Ainsi, pour un gyrolaser de géométrie carré avec un périmètre de 14 cm, nous obtenons un

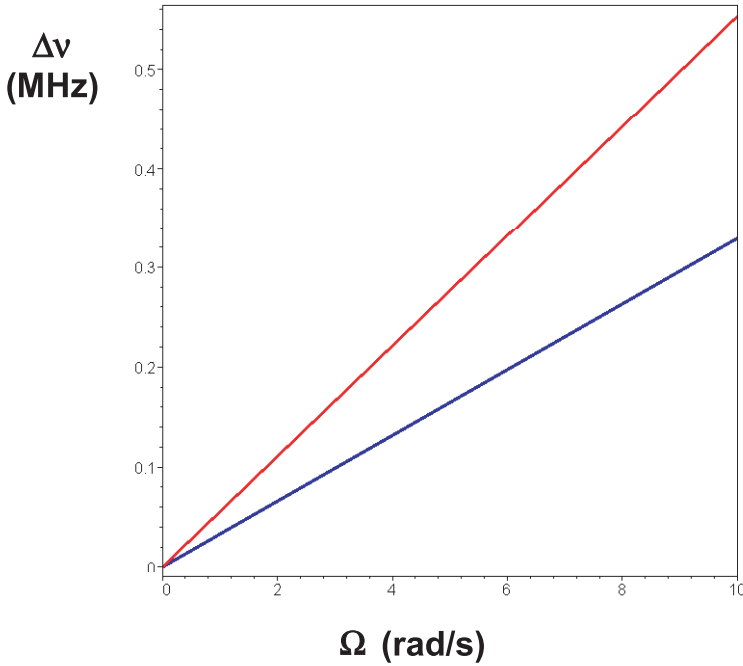


FIG. 4.28 – Fréquence de battement (MHz) en fonction de la rotation du gyrolaser (rad/s) et des longueurs d'onde à l'émission ( $\lambda = 0,633 \mu\text{m}$  (laser He-Ne, courbe rouge) et  $1,064 \mu\text{m}$  (laser Nd:YAG, courbe bleue)).

décalage en fréquence en fonction de la rotation du système interférométrique actif qui est reporté sur la figure 4.28.

Il est à noter que la rotation terrestre ( $15^\circ/\text{h} = 72,72 \mu\text{rad/s}$ ) se traduit par un décalage en fréquence de 4 Hz à  $6328 \text{ \AA}$  que l'on doit comparer à la fréquence de l'onde optique qui est de 473,755 THz. La détection de la rotation de la Terre implique donc une mesure de fréquence laser à mieux que  $10^{-14}$ .

À partir de la relation (4.35), nous définissons un facteur d'échelle :

$$F = \frac{\Delta\omega}{\Omega} = \frac{8\pi S}{\lambda L} \tag{4.36}$$

qui est un paramètre sans dimension permettant de mesurer l'effet d'amplification lié au choix de la longueur d'onde et de la structure de cavité. Pour une structure en anneau carré ( $S_{\square} = \frac{L^2}{16}$ ) nous obtenons le facteur d'échelle suivant :

$$F_{\square} = \frac{\Delta\omega}{\Omega} = \frac{\pi L}{2\lambda}$$

Pour une cavité en anneau carré de périmètre  $L = 14 \text{ cm}$ , le facteur d'échelle à  $\lambda = 0,6328 \mu\text{m}$  vaut :  $F_{\square} = 3,5 \times 10^5$ . Ces grandeurs sont rappelées afin d'avoir



FIG. 4.29 – (a) Gyrolaser uni-axe He-Ne, (b) centrales inertielles tri-axes He-Ne (THALES Avionics).

à l'esprit les ordres de grandeurs associés aux montages de gyromètres laser. La photographie 4.29a illustre un gyrolaser uni-axe He-Ne et la figure 4.29b une centrale gyrolaser tri-axes. Cette dernière configuration permet d'avoir accès au mouvement tri-dimensionnel d'un mobile.

On mesure ainsi les précisions à atteindre et à conserver durant les différentes phases de conception et de fabrication de tels capteurs inertiels. On comprend l'importance des « petits » effets physiques à maîtriser pour obtenir les précisions souhaitées (sensibilité supérieure à la mesure de la rotation de la terre). Tout ceci fait du gyrolaser l'objet le plus élaboré des dispositifs laser sur les plans à la fois technologique et théorique [51–53].

Sa réalisation a largement contribué à enrichir divers domaines de la physique dans ses différentes branches telles que :

- la physique atomique et spectroscopie ;
- la physique du laser (cavité, interaction résonnante lumière-milieu dilué...);
- la physique des décharges et des plasmas sous faible pression ;
- les techniques de dépôts (couches multidiélectriques optiques à faibles pertes et à haut pouvoir réflecteur lié au faible gain) ;
- la technologie optique (usinage, polissage...) et la métrologie

en y associant l'apport des techniques de traitement du signal. On doit également noter que l'optique non linéaire a bénéficié des études menées autour du gyrolaser par l'analyse des effets du troisième ordre dans les milieux dilués et résonnants. La raison est que tout effet non-réciproque est à bannir dans un tel montage (une perte différentielle induit un fonctionnement unidirectionnel). Une diffusion contribue à coupler les deux ondes contra-propagatives, ce



qui se traduit par un verrouillage de la fréquence des deux ondes et par voie de conséquence par la création d'une zone aveugle. Cette zone correspond au domaine fréquentiel sur lequel les deux ondes ont la même fréquence, d'où l'absence d'un signal de battement.

Il semble que ces études sur la gyrométrie puisse dans le futur se perpétuer avec l'émergence des gyromètres à ondes de matière, fondés sur l'exploitation des mécanismes qui gouvernent l'optique atomique utilisant des atomes refroidis par voie laser [54]. Les sensibilités susceptibles d'être atteintes par les gyromètres à atomes pourraient être de plusieurs ordres de grandeurs supérieures à celles des gyrolaser ou des gyromètres à fibre optique.

Dans un schéma utilisant un interféromètre de type Mach-Zehnder, pour une même surface, le rapport sur les variations de phase entre un système photonique et un système atomique est donné par la relation :

$$\frac{\Delta\Phi_{\text{atome}}}{\Delta\Phi_{\text{photon}}} \sim \frac{mc^2}{\hbar\omega} \sim 10^{10} \quad (4.37)$$

où  $m$  est la masse de l'atome considéré et  $\omega$  est la pulsation associée à la longueur d'onde de  $1 \mu\text{m}$ . On peut penser que les précisions pouvant être atteintes par de tels instruments permettront de tester certains fondements de la physique [55].

### 4.5.3 Lévitiation optique

Un faisceau lumineux est généralement utilisé par ses propriétés d'éclairage, ses qualités de transport énergétique ou encore ses caractéristiques spectrales. En outre lorsque ce faisceau entre en interaction avec un objet, il lui applique une force appelée pression de rayonnement. En effet, un faisceau lumineux transporte une quantité de mouvement qu'il peut transmettre à un objet.

La force exercée par un faisceau laser sur une sphère dont la dimension est grande devant la longueur d'onde optique a permis de mettre en évidence l'effet de lévitation optique [56]. En présence d'une distribution non uniforme de l'éclairage, cette force se décompose en une composante axiale (direction du faisceau laser) et une composante transverse dirigée perpendiculairement à cette direction. Selon le type de micro-bille considéré (diélectrique ou métallique) la composante transverse tend à repousser ou à piéger l'objet dans le faisceau.

Par exemple, en employant une fibre optique unimodale on dispose d'une distribution d'éclairage quasi-gaussienne, qui correspond au profil radial du mode fondamental. En ajustant la puissance optique véhiculée par ce guide d'onde optique, il est possible de maintenir en lévitation une sphère réfringente avec une position d'équilibre qui dépend de sa masse, donc de sa taille. C'est ce qui est illustré sur la figure 4.30. En utilisant un laser émettant en continu à  $1 \mu\text{m}$ , pour une puissance de quelques centaines de mW délivrée en extrémité de fibre, on observe la lévitation de microbilles [57].

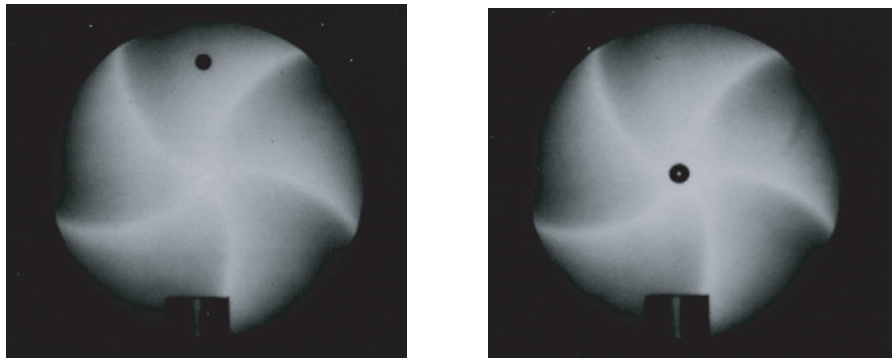


FIG. 4.30 – Illustration de la mise en lévitation d'une sphère réfringente de diamètre  $30\ \mu\text{m}$  à gauche et  $45\ \mu\text{m}$  à droite à partir d'un faisceau laser émergent d'une fibre optique unimodale.

On utilise cet effet pour auto-centrer la microbille par rapport à un faisceau ou pour analyser et pour mesurer les propriétés physiques d'un objet sans contact.

Ce sont les propriétés de la pression de rayonnement, appliquées aux atomes qui ont donné lieu au développement de concepts plus raffinés : le ralentissement et le piégeage d'atomes par la lumière (voir chapitre 2).

#### 4.5.4 Le domaine de l'énergie

L'énergie rayonnée par le soleil et par un grand nombre d'étoiles est issue de réactions de fusions nucléaires. Sur terre, la maîtrise de réactions thermonucléaires de fusion permettrait de disposer d'une source d'énergie électrique quasi-inépuisable. Bien que les difficultés soient nombreuses, c'est cette perspective qui anime une grande partie de la communauté scientifique, soutenue par la mise en place de grands projets internationaux, dont le projet ITER (International Thermonuclear Experimental Reactor), qui doit démontrer la maîtrise du plasma en combustion. La fusion consiste à combiner deux noyaux légers pour donner naissance à un noyau plus lourd. Cette opération s'accompagne d'une importante libération d'énergie. Elle est cependant difficile à obtenir car la répulsion électrostatique, ou barrière coulombienne, qui se manifeste entre les noyaux des atomes empêche leur rapprochement pour fusionner (les forces nucléaires sont à courte distance, alors que les forces électriques sont à longue distance). Les trois conditions pour effectuer des opérations de fusion sont :

- la réalisation d'une densité élevée de noyaux en interaction, afin d'avoir un taux de collisions important ;

- la création d'un milieu fortement ionisé (plasma chaud), qui porte les noyaux dans un état d'agitation thermique élevé. Il permet de communiquer aux particules des vitesses adaptées au franchissement de la barrière électrostatique ;
- une durée importante de confinement du plasma, conservant une forte densité de particules et une température élevée, afin de disposer d'une quantité suffisante de matière qui puisse fusionner.

En fait, il existe un critère établi en 1957 par J.D. Lawson qui veut que les conditions minimales pour obtenir une opération de fusion Deutérium-Tritium (DT) sont représentées par le produit :

$$n_{\text{ions}}\tau \geq 10^{14} \quad (4.38)$$

$n_{\text{ions}}$  est la densité d'ion en  $\text{cm}^{-3}$  et  $\tau$  est le temps de confinement, la température d'ignition devant atteindre  $4,5 \times 10^7$  K. En conséquence, il existe deux approches pour réaliser cette opération. La première consiste à favoriser un temps de confinement élevé d'un plasma chaud et dilué (faible densité d'ions). C'est la voie suivie avec le confinement magnétique (Tokamak et ITER). La seconde nécessite la création d'une forte densité de particules sur un temps bref. Cette condition peut être réalisée avec le confinement inertiel par voie laser. Deux projets sont en cours de réalisation : aux États-Unis, le NIF (National Ignition Facility au Lawrence Livermore National Laboratory) et le LMJ (Laser Méga Joule) au CEA-DAM / CESTA en France près de Bordeaux. Le LMJ est un élément du programme simulation initié après la signature en 1996, et la ratification par la France du « Traité d'interdiction complète des essais nucléaires (Tice) ». Ce programme doit permettre de valider des modèles physiques qui rendent compte des conditions de fonctionnement de l'armement nucléaire.

Le LMJ est un instrument exceptionnel dans ses dimensions, par la complexité du format spatio-temporel des faisceaux à mettre en œuvre et par les précisions de pointage à maîtriser sur la cible. La longueur du bâtiment qui l'accueille est de 300 m (Fig. 4.31). Il est constitué de 240 faisceaux lasers regroupés en 60 « quadruplets » triplés en fréquence. Tous ces faisceaux sont focalisés par les deux extrémités d'une cavité de 2,5 mm de diamètre et de 1cm de long qui fait office de four. Au centre de cette cavité est placée la microbille à partir de laquelle doit être initiée l'opération de fusion. Cette cavité est placée dans une chambre d'expérience et de protection de 10 m de diamètre, elle-même insérée dans un hall de 60 m de diamètre et de 40 m de hauteur. Ce hall est situé au centre du bâtiment avec de part et d'autres les chaînes laser (Fig. 4.32).

L'énergie électrique stockée est de l'ordre de 400 MJ qui se transforme en une énergie laser de 4 MJ à  $\lambda = 1,054 \mu\text{m}$ . Les faisceaux sont ensuite triplés en fréquence pour obtenir un rayonnement UV pouvant pénétrer à travers un plasma chaud (à très haute densité électronique). La masse totale

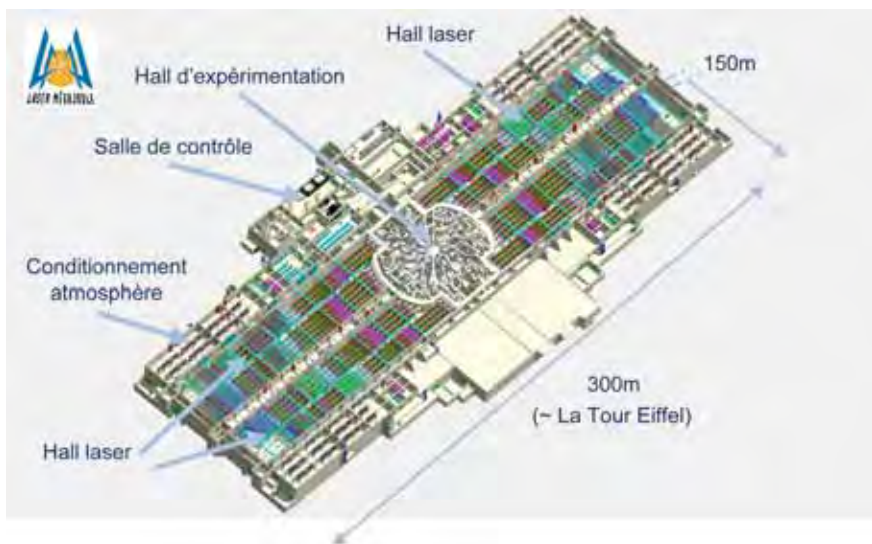


FIG. 4.31 – Configuration architecturale du LMJ. La longueur du bâtiment est équivalente à la hauteur de la Tour Eiffel. (CEA-DAM.)

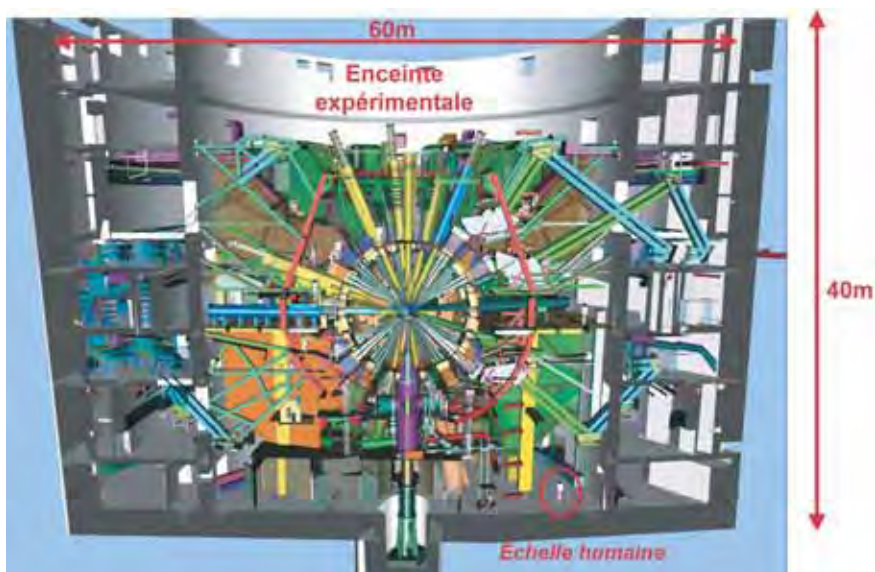


FIG. 4.32 – Vue et dimensions de l'enceinte expérimentale. (CEA-DAM / CESTA.)

des matériaux optiques utilisés dans ce projet (verre dopé  $\text{Nd}^{3+}$ , KDP et DKDP, verres, silice) est de l'ordre de 300 tonnes et les surfaces traitées sont approximativement de  $1750 \text{ m}^2$ .

L'apport en énergie sur la cible est supérieure au mega-joule à la longueur d'onde de 351 nm dans une courte durée d'impulsion (0,3 à 25 ns). L'opération revient à comprimer uniformément et à porter à très haute température une microbille de quelques dixièmes de milligrammes d'un mélange deutérium-tritium jusqu'à atteindre les conditions pour la fusion.

### 4.5.5 Le laser et ses applications industrielles

L'intégration du laser dans le domaine industriel consiste à réaliser les fonctions de découpe, de soudure, de perçage, de traitement de surfaces (trempe superficielle), de prototypage rapide, de micro-fabrication et de micro-usinage.

Quelques-unes de ces fonctions peuvent être réalisées à partir d'un outil laser unique. Mais généralement, une source est dédiée à une fonction spécifique qui gouverne la fluence énergétique ( $\text{J}/\text{m}^2$ ) devant être délivrée par le faisceau.

Les avantages de l'outil laser résultent de la possibilité de déplacement rapide du faisceau d'illumination, et de la précision obtenue, par exemple en micro-usinage. L'usinage à sec sans contact mécanique est également une propriété spécifique à cet instrument. Enfin l'intégration des sources sub-picosecondes ouvre de nouvelles perspectives, comme un traitement interne au matériau (densification ou claquage diélectrique localisé).

Les lasers employés sont divers, ainsi que les domaines spectraux couverts. On utilise essentiellement le laser moléculaire  $\text{CO}_2$ , qui émet dans l'infrarouge à la longueur d'onde de  $10,6 \mu\text{m}$  et le laser Nd:YAG qui, émet dans le proche infra-rouge à  $1,06 \mu\text{m}$ . Pour ce dernier, l'avantage est que le faisceau peut être transporté par une fibre optique à base de silice, transparente à cette longueur d'onde. Les diodes laser de puissance fibrées sont également développées pour des applications industrielles. Selon l'opération à réaliser on choisit une source en continu ou en impulsion.

Enfin, on peut utiliser le laser à excimère, qui est un laser à gaz impulsionnel émettant un rayonnement ultra-violet. Le terme excimère caractérise une molécule constituée d'un atome de gaz halogène et d'un atome de gaz rare excités électroniquement. Il émet un rayonnement entre  $0,15$  et  $0,25 \mu\text{m}$  selon l'halogénure de gaz rare employé et il trouve sa place dans le domaine du traitement des matériaux et plus particulièrement en ablation photochimique.

En régime continu, le paramètre important est la puissance optique fournie par le faisceau. Typiquement, le traitement des matériaux nécessite la mise en oeuvre de sources de la classe 100 W à 20 kW. En mode impulsionnel, les paramètres importants sont l'énergie par impulsion en joules et le nombre d'impulsions par seconde (cadence de répétition). Selon l'application, l'énergie par impulsion est comprise entre 1 mJ et quelques centaines de joules avec des durées impulsionnelles allant de la nanoseconde à la milliseconde et une fréquence de travail de 0,1 Hz à quelques kHz.

Les paramètres qui déterminent la nature de l'interaction sont la fluence ( $\text{J}/\text{m}^2$ ) ou l'éclairement (intensité en  $\text{W}/\text{m}^2$ ) du faisceau et la longueur d'onde

du laser choisie en fonction des propriétés optique et thermique (conductivité et diffusivité) du matériau à traiter [58].

Les modes de transfert d'énergie dans une interaction lumière-matière en régime de forte densité d'énergie s'effectuent par conduction (procédés classiques de soudage) et par vaporisation, qui donne naissance à la formation d'une cavité appelée « capillaire ». Le processus d'échauffement de l'état solide par conduction se traduit par les étapes de fusion du métal en surface, sur une zone définie par la surface d'interaction. Le matériau est en phase liquide et cette propriété est mise à profit pour la réalisation de soudures. Dès que la température de surface dépasse la température d'évaporation, une vapeur métallique se développe. Sa détente induit une pression de recul qui vient déformer la surface du bain liquide en s'opposant à la tension de surface. Celui-ci est repoussé de part et d'autre du faisceau d'illumination et un cratère se forme. Durant cette phase, on observe la formation d'un panache (plume) et d'un plasma auquel est associé une onde de choc. Ensuite, ce trou se comporte comme un guide d'onde (capillaire) pour le rayonnement laser qui pénètre dans le milieu par vaporisation de matière. Ce mécanisme qui décrit le perçage ou la découpe est schématisé sur la figure 4.33.

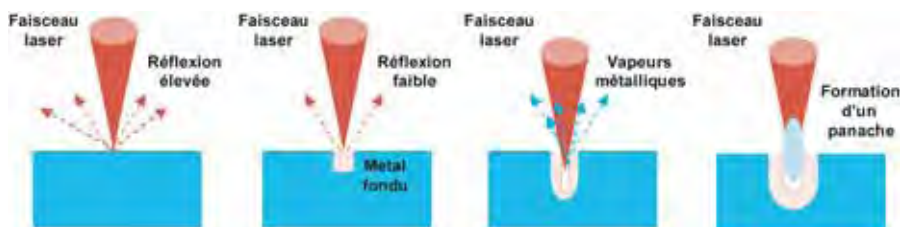


FIG. 4.33 – Différentes étapes qui gouvernent l'interaction laser-matériau sous fort éclairage.

La soudure laser consiste à réaliser une zone fondue entre deux éléments. Elle offre certaines particularités qui rendent cette technique compétitive par rapport aux méthodes conventionnelles (électrique ou brasure). Les avantages de l'outil laser sont :

- la soudure à l'air ambiant ;
- une meilleure qualité (la zone affectée thermiquement est réduite) ;
- la suppression des chanfreins pour soudure bord à bord ;
- l'accès à des zones confinées ;
- la soudure sans contact et sans contrainte appliquée – un accroissement de la vitesse de soudure (jusqu'à 3 m/min.) ;
- une meilleure tenue mécanique ;

- la soudure en un seul passage (jusqu'à 8 mm d'acier avec un laser CO<sub>2</sub> de 6 kW) ...

Un exemple d'application est la soudure des lames de rasoir (multi-lames) réalisée industriellement par laser.

La figure 4.34 illustre les caractéristiques du laser employé, en termes d'éclairement et de durée d'impulsion, selon le procédé d'usinage considéré.

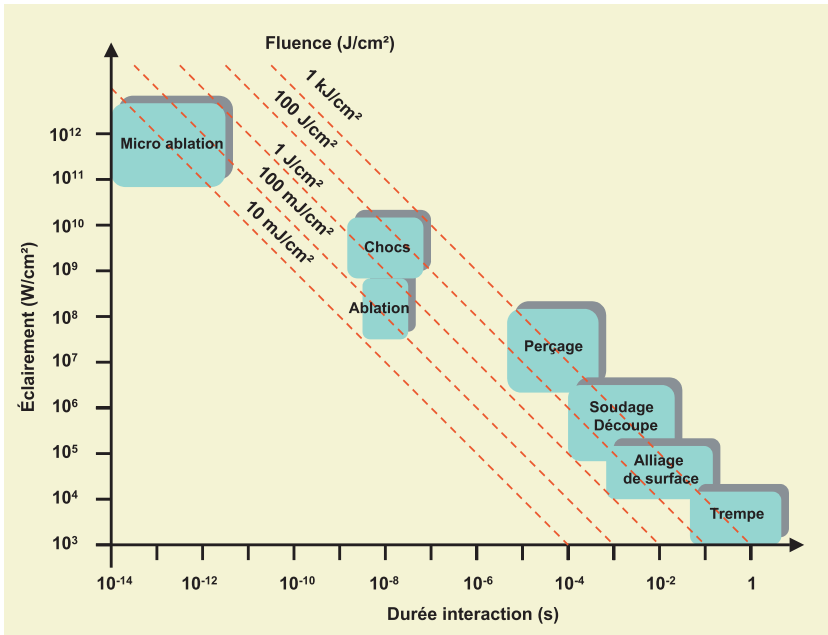


FIG. 4.34 – Diagramme caractérisant l'éclairement laser et la durée d'interaction en fonction du procédé d'usinage.

Aujourd'hui, le laser CO<sub>2</sub> trouve un concurrent avec le laser Nd:YAG pompé par diodes. En multipliant le nombre de modules pompés par diodes dans une cavité linéaire, on augmente la puissance optique délivrée par des laser Nd:YAG industriels. À partir de six modules, la compagnie TOSHIBA a réalisé un laser Nd:YAG délivrant plus de 11 kW en continu avec un rendement à la prise de l'ordre de 20 % (Fig. 4.35).

Il reste cependant à rendre compétitive cette technologie, fondée sur l'utilisation de diodes laser de pompage, en terme de coût à l'achat.

Le micro-usinage a également fait de nouveaux adeptes avec l'emploi de sources sub-picosecondes [59]. La faible durée de dépôt d'énergie fait que la thermalisation du matériau soumis à l'éclairement laser est plus lente que le mécanisme d'ablation lié au développement d'un plasma. Dans ce cas, on obtient une vaporisation du matériau sans passer par la phase de fusion thermique. Il en résulte la réalisation de motifs de micro-perçage très précis avec

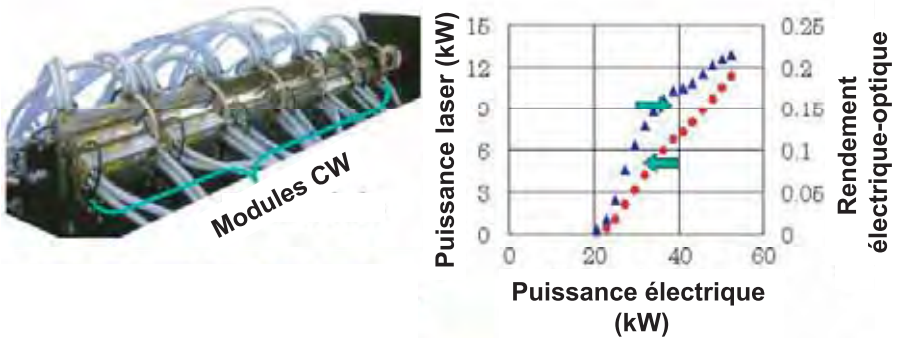


FIG. 4.35 – Illustration d'un laser Nd:YAG pompé par diodes, en continu, pour applications industrielles. Il comporte six modules (têtes de pompage avec refroidissement par eau), en série, insérés dans une cavité optique. (TOSHIBA Corporate Manufacturing Engineering Center.)

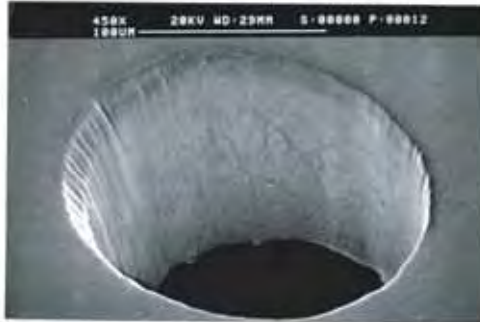


FIG. 4.36 – Photographie d'un résultat de perçage laser de diamètre  $150\ \mu\text{m}$ , d'une micro feuille d'acier, obtenu avec un faisceau laser sub-picoseconde ( $< 100\ \text{fs}$ ).

une excellente géométrie, bien supérieure à ceux obtenus avec des impulsions longues délivrées par des laser Nd:YAG ou à  $\text{CO}_2$  (Fig. 4.36).

La gravure laser est également très répandue. On procède à l'écriture d'un motif par projection à travers un masque (pochoir) ou directement en déplaçant le faisceau par des micro-miroirs galvanométriques. La finesse de résolution et le caractère indélébile de l'inscription attribuent un caractère infalsifiable à ce marquage, ce qui le protège de la contrefaçon. Cette méthode de gravure est réalisée à distance et sans contact, donc sans usure mécanique et de manière silencieuse.





FIG. 4.37 – Illustration des effets de nettoyage de statuaire par ablation laser (THALES Laser).

Le décapage d'objets d'art peut être obtenu avec des lasers Nd:YAG déclenchés par un mécanisme d'ablation de surface lors de l'interaction laser-matière. Cette opération s'accompagne d'un contrôle précis de la profondeur de nettoyage. Ce nettoyage a un côté un peu laborieux en raison de la faible cadence (10 à 100 Hz) du laser employé dans ce type d'intervention. On utilise en effet des sources laser Nd:YAG qui doivent délivrer des impulsions de durée 10 ns, avec une assez grande énergie (quelques centaines de mJ par impulsion) nécessaire au développement du mécanisme d'ablation (Fig. 4.37).

Au niveau des applications industrielles nous pouvons également en citer plusieurs.

Le prototypage rapide de matériaux variés (polymères, métaux, céramiques) qui a bouleversé la phase de conception et de maquettage 3D d'un objet. Différentes techniques sont développées (frittage, dépôt de matière...). La photopolymérisation sous rayonnement laser est une de ces méthodes de prototypage (Fig. 4.38). Elle consiste à réaliser la polymérisation d'un bain de polymère liquide à partir d'un faisceau laser dont la déviation 2D est assurée par un ensemble de miroirs mobiles (M). Un support (S) est animé d'un déplacement vertical afin d'obtenir un motif tri-dimensionnel. Une règle (R) permet d'alimenter en polymère liquide la surface du bain afin de répéter, couche après couche, l'opération de polymérisation. L'épaisseur de chaque couche est typiquement de 50  $\mu\text{m}$  et la vitesse de balayage est comprise entre 2 et 20 m/s. La figure 4.39 illustre quelques réalisations.

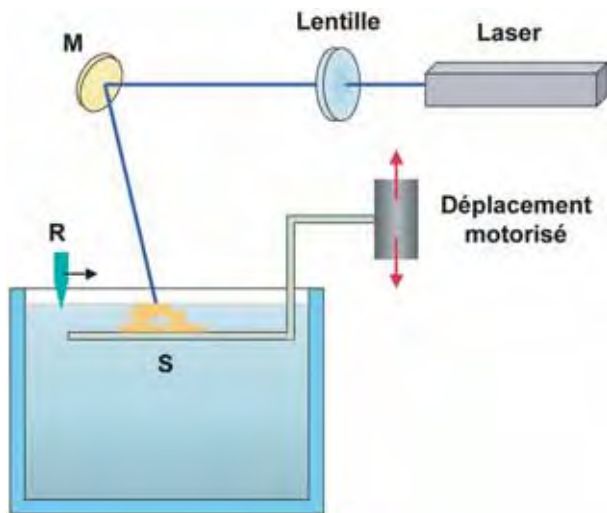
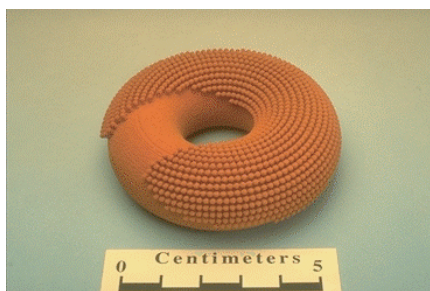


FIG. 4.38 – Schéma synoptique d'un système de prototypage laser rapide.



a)



b)

FIG. 4.39 – Illustration d'objets réalisés par prototypage laser. (a) Surface structurée (<http://web.mit.edu/tdp/www/experimental.html>) ; (b) visualisation de l'assemblage de pièces mobiles d'un système de pompe mécanique, « Rapid Prototyping and Rapid Tooling » Fritz Klocke, Fraunhofer Institute of Production Technology.

Le traitement ou le conditionnement de surfaces par voie thermique, l'activation physique et chimique de surface (ex : collage par adhésion contrôlée, génie de la bio compatibilité-bio activité) utilisent les propriétés de directivité du faisceau laser. Par exemple, on peut localement rendre hydrophile ou hydrophobe une surface par illumination UV, qui facilite l'addition de groupes fonctionnels à l'interface.

Le nettoyage ou la décontamination par voie laser concerne notre environnement (dépollution et stérilisation de l'atmosphère par rayonnement UV, démantèlement de structures et neutralisation de sites nucléaires...)

Le dépôt de couches minces fait également l'objet de travaux. Cette méthode de dépôt assisté par voie laser consiste à soumettre une cible à un rayonnement laser impulsif. Les matériaux éjectés durant cette interaction constituent une plume qui vient se condenser sur un substrat sous la forme d'un film mince.

## 4.6 Conclusions

Les applications du laser couvrent ainsi un vaste champ de domaines extrêmement divers [60] et cette source de lumière est devenue un outil incontournable pour la recherche dans les laboratoires aussi bien que dans notre vie quotidienne. Nous en sommes redevables à toute une lignée de scientifiques, dont Einstein est une figure emblématique, et aussi d'ingénieurs, qui ont apporté leur créativité et leur enthousiasme au développement de cet instrument aux multiples facettes qui, à ce titre, est le résultat d'efforts très multidisciplinaires.

Ainsi, la physique fondamentale exploite les propriétés du laser dans presque tous les domaines. On peut citer quelques exemples :

- l'analyse des propriétés de la matière (moléculaire, fluide, état solide, plasma...);
- la réalisation de condensats de Bose-Einstein (BEC);
- l'initiation de réactions de fusion nucléaire par confinement inertiel;
- la détection d'ondes gravitationnelles et le test de l'électrodynamique quantique (QED)...

La métrologie accroît sa précision en intégrant des lasers ultra-stables et en utilisant des peignes de fréquences raccordant le domaine optique au domaine radiofréquence. Le refroidissement d'atomes par laser est aussi un procédé d'avenir pour augmenter les performances des horloges atomiques ou pour accroître la précision sur la mesure des constantes fondamentales. Les domaines de la navigation, de la géophysique et de l'astrophysique bénéficient de ces avancées.

Le traitement des matériaux (soudure, découpe, perçage et traitement de surface) verra certainement s'étendre son champ d'application pour façonner des systèmes micro et nanostructurés.

Les domaines de la biologie et du médical, qui exploitent déjà l'émission laser dans les domaines du diagnostique, de la thérapie et de la chirurgie, voient s'approfondir leur connaissance du vivant grâce au laser par l'analyse optique de la structure de l'ADN et de la composition cellulaire. Enfin, les technologies de l'information intègrent déjà le laser dans la transmission numérique, dans l'interconnexion optique et pour le stockage optique. Elles verront prochainement s'ouvrir de nouveaux champs d'application dans les domaines de

la projection laser et en informatique par la transmission optique de données dans les processeurs ou entre modules.

La suite de l'histoire du laser et de ses applications devrait mettre à profit diverses avancées dans les domaines des matériaux organiques actifs par injection électrique (laser plastique sous forme de microcavité, de guide planaire ou de fibre), des diodes laser UV de puissance qui ouvriront de nouvelles perspectives pour le pompage de matrices solides dopées, pour la commande optique ou pour d'autres applications industrielles, ou encore des sources laser sub-picoseconde de puissance pompées directement par diodes.

Ces exemples sont loin d'épuiser les perspectives de progrès attendus dans la technologie des lasers. On peut prévoir sans risque que les laser du futur fourniront de multiples prétextes à la commémoration d'Albert Einstein et de ses travaux sur l'émission stimulée dans les décennies qui viennent.

## Appendice A : de la rectification optique au rayonnement THz

Soit une impulsion lumineuse de pulsation  $\omega$ , caractérisée par une fonction enveloppe à distribution gaussienne de largeur à  $1/e^2 = \Delta t$ .

$$E(t) = E_0 e^{-\left(\frac{t}{\Delta t}\right)^2} \cos(\omega t)$$

La polarisation induite en se limitant à une partie des effets du second ordre devient :

$$P_{\text{pol}}(t) = \varepsilon_0 \left[ \left(1 + \chi^{(1)}\right) \cos(\omega t) e^{-\left(\frac{t}{\Delta t}\right)^2} + \left(\frac{\chi^{(2)}}{2} \cos(2\omega t) + \frac{\chi^{(2)}}{2}\right) E_0 e^{-2\left(\frac{t}{\Delta t}\right)^2} \right] E_0 \quad (\text{A.1})$$

Le premier terme est à la même pulsation que celle du champ incident. Dans le second, on voit apparaître un terme à la pulsation double. La création de cette composante harmonique est la source d'une opération de doublement de fréquence. Enfin, on observe l'existence d'une composante continue, proportionnelle à  $E(t)^2$ , qui correspond à la création d'une rectification optique. Cet effet traduit la modification de la constante diélectrique statique du milieu, induite par l'éclairement optique. Il peut être évalué en mesurant, par exemple, une variation de la capacité électrique du milieu traversé par un faisceau optique. Les figures A.1a et A.1b représentent respectivement le champ incident et la polarisation induite avec ses composantes illustrant la formule (A.1). La modification rapide de la polarisation associée à la rectification optique s'accompagne de l'émission rayonnante d'une onde électromagnétique :

$$E_{\text{ray}}(t) = \frac{\partial P_{\text{rect}}(t)}{\partial t} = -2\varepsilon_0 \chi^{(2)} E_0^2 \frac{t}{\Delta t^2} e^{-2\left(\frac{t}{\Delta t}\right)^2} \quad (\text{A.2})$$

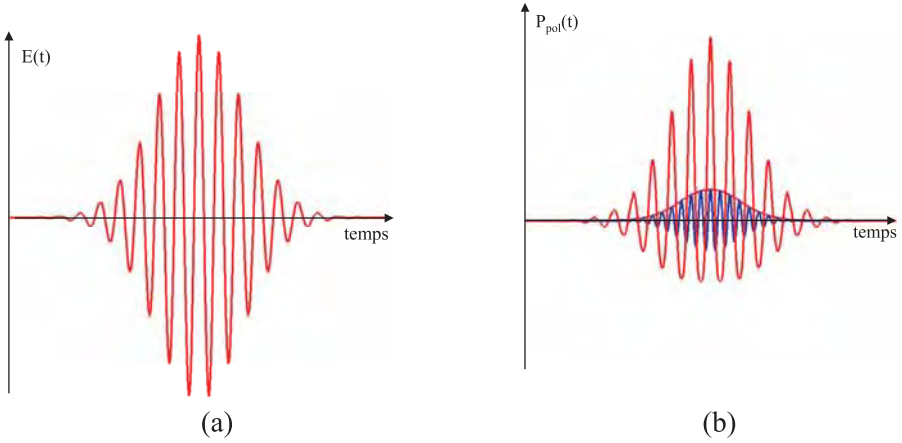


FIG. A.1 – (a) Évolution temporelle du champ optique modulé temporellement par une fonction enveloppe gaussienne. (b) Polarisation induite incluant les composantes linéaire et non linéaire. On observe la partie harmonique (bleu) et la polarisation associée à la rectification optique (rouge).

qui a pour transformée de Fourier :

$$\text{T.F.} \{E_{\text{ray}}(t)\} = i \frac{\varepsilon_0 \chi^{(2)} \sqrt{2\pi} E_0^2 \Delta t}{4} \Omega e^{-\frac{1}{8} \Omega^2 \Delta t^2} \quad (\text{A.3})$$

Les figures A.2a et A.2b représentent les fonctions associées à la polarisation induite, au champ et au spectre rayonné.

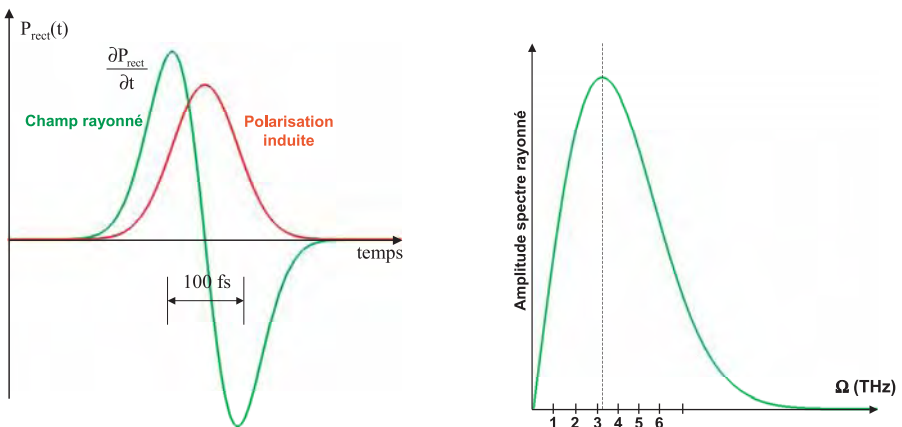


FIG. A.2 – (a) Dépendance temporelle de la polarisation induite par rectification optique et évolution du champ rayonné. (b) Transformée de Fourier du champ rayonné.

La valeur maximale du rayonnement engendré correspond à la pulsation  $\Omega_{\max} = 2/\Delta t$ . Ainsi, avec une impulsion de 100 fs, on peut générer un maximum de rayonnement à 3,18 THz.

## Appendice B : compression et étirement temporel

En présence d'un fort éclairage, l'auto-modulation de phase élargit le spectre d'une impulsion initiale par une modification de la phase instantanée. Cette variation de phase dépend du profil temporel de l'impulsion et de sa largeur. Par ailleurs, une impulsion de durée ultra-courte présente naturellement un large spectre de fréquence. En se propageant dans un milieu dispersif usuel (en régime de dispersion normale), les composantes à haute fréquence sont retardées par rapport aux composantes à basse fréquence. On obtient un étalement de l'impulsion dans le temps. Les techniques pouvant compenser ce retard, entre les composantes de fréquence, peuvent être à base de prismes optiques (Fig. B.1a) où de réseaux de diffraction (Fig. B.1b). Dans la première configuration, on utilise les différences d'épaisseurs de verre traversées et la dispersion chromatique des prismes. En adaptant la distance qui sépare une paire de réseaux de diffraction, nous pouvons utiliser les propriétés de dispersion angulaire afin d'ajuster les chemins optiques parcourus par les différentes composantes spectrales de l'impulsion. Dans ce cas, on va chercher à retarder les basses fréquences par rapport aux hautes fréquences du spectre optique.

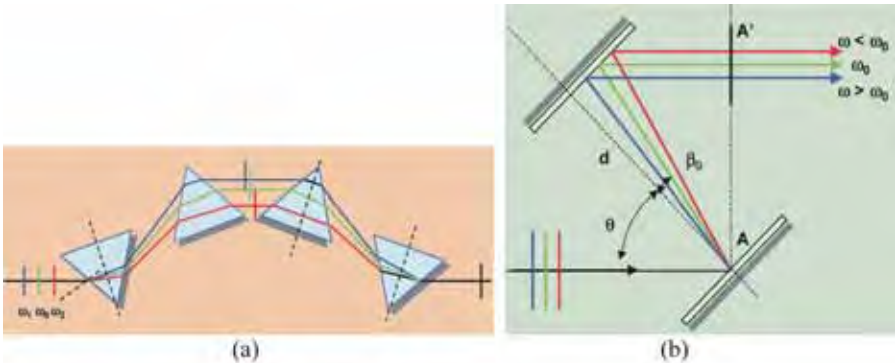


FIG. B.1 – (a) Compresseur à prisme et (b) à deux réseaux. À l'entrée du système, les composantes à basse fréquence arrivent avant les composantes à haute fréquence. On cherche à obtenir en sortie, une concordance des différentes composantes spectrales en jouant sur les différences de chemins optiques.

L'équation du réseau de diffraction s'écrit :  $\sin \beta_0 + \sin \theta = pN\lambda$ , où  $\theta$  est l'angle d'incidence du faisceau contenant l'impulsion étirée spectralement,  $\beta_0$  est l'angle de diffraction,  $N$  est le nombre de traits par unité de longueur

( $N = 1/a$ , où  $a$  est le pas du réseau) et  $p$  correspond à l'ordre de diffraction ( $p = 0, \pm 1, \pm 2 \dots$ ).  $d$  est la distance entre les deux réseaux et  $\omega_0$  représente la pulsation centrale du spectre de l'impulsion. Le chemin optique entre le point A et le point A' (Fig. B.1a) pour la pulsation  $\omega_0$  s'écrit :

$$L = d \left( \frac{1}{\cos \theta} + \frac{1}{\cos \beta_0} \right)$$

La combinaison de cette relation avec celle des réseaux au premier ordre et au voisinage de  $\omega_0$  donne le retard par rapport à la longueur d'onde centrale  $\lambda_0 = 2\pi c/\omega_0$  :

$$\Delta\tau = \tau(\lambda) - \tau(\lambda_0) = - \left( \frac{\lambda - \lambda_0}{2\pi c} \right) \frac{d}{2a} \frac{\text{tg}\beta_0}{\cos^2 \beta_0}$$

Ainsi, pour des réseaux avec 500 traits/mm, et pour un angle d'incidence de  $45^\circ$  avec une distance entre réseaux de 40 cm, nous obtenons le décalage temporel en fonction de la longueur d'onde qui est présenté sur la figure B.2.

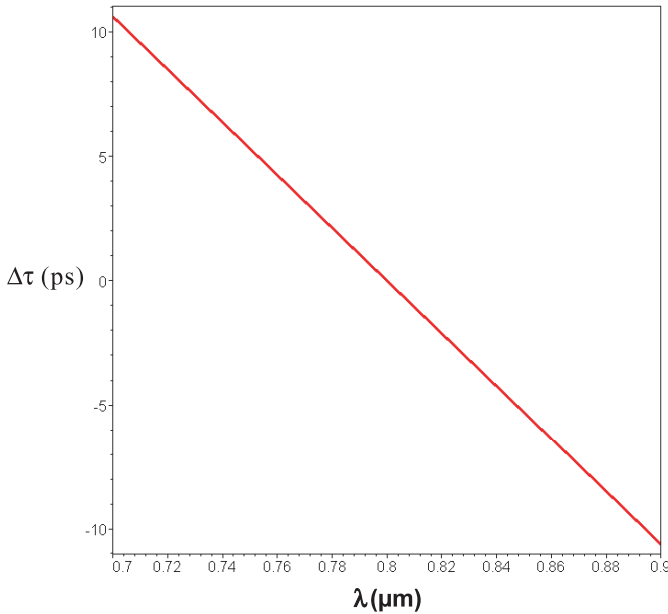


FIG. B.2 – Décalage temporel en fonction de la longueur d'onde. La longueur d'onde centrale est à  $0,8 \mu\text{m}$ . Au-delà de cette longueur d'onde,  $\Delta\tau < 0$ , ce qui correspond à retarder les composantes spectrales par rapport à  $\lambda_0$ . Inversement, pour  $\lambda < \lambda_0$ , les composantes spectrales sont en avance.

En jouant sur la distance, sur l'angle d'incidence et sur les caractéristiques du réseau de diffraction, il est possible d'ajuster la compensation de la dispersion temporelle, sur un domaine plus ou moins important. Enfin, le parcours



FIG. B.3 – Une impulsion à dérive de fréquence, ultra-courte, issue d'un oscillateur est étirée spectro-temporellement afin de réduire l'intensité optique. Cela revient à répartir l'énergie de l'impulsion dans le temps, donc à réduire la puissance crête. Elle est ensuite amplifiée puis elle est comprimée temporellement.

inverse, qui consiste à avoir initialement toutes les composantes spectrales regroupées, se manifeste par la réalisation d'un étirement temporel de l'impulsion. C'est l'intégration de ces deux fonctions (étirement/compression), dans des chaînes d'amplification d'impulsion ultra-courtes (Fig. B.3), qui permet de surmonter les limitations imposés par la tenue à l'éclairement optique des matériaux laser [34].

## 4.7 Références

- [1] M. Planck, *Absorption und Emission elektrischer Wellen durch Resonanz*, Sitzungsber. d.k. preuss. Akad. d. Wissensch. Vom 21, März 1895, p. 289  
M. Planck, *The Theory of Heat Radiation* (traduction de *Waermestrahlung* publié en 1914), Dover Pub. Inc. N.Y. (1959).
- [2] A. Einstein, *Zur Quantentheorie der Strahlung*, Physik. Zeitschr. XVIII, 1917, p. 121.  
A. Einstein, *On the Quantum Theory of Radiation*, The Old Quantum Theory, Elmsford N.Y., Pergamon, 1967, p. 167.
- [3] A. Kastler, *Quelques suggestions concernant la production optique et la détection optique d'une inégalité de population des niveaux de quantification spatiale des atomes. Application à l'expérience de Stern-Gerlach et à la résonance magnétique*, J. Phys. 253 (11 juin 1950).  
J. Brossel, A. Kastler, J. Winter, *Création optique d'une inégalité de population entre les sous-niveaux Zeeman de l'état fondamental des atomes*, J. Phys. 668 (13 décembre 1952).
- [4] A.L. Schawlow and C.H. Townes, *Infrared and Optical Masers*, Phys. Rev. **112** 1940 (1958).
- [5] N.G. Basov, A.M. Prokhorov, *Molecular beams application for radiosopic study of rotational molecular spectra*, JETPh **27**, 431 (1954).



- [6] J.P. Gordon, H.J. Zeiger, C.H. Townes, *Molecular microwave oscillator and new hyperfine structure in the microwave spectrum of NH<sub>3</sub>*, Phys. Rev. **95**, 282 (1954).
- [7] J.P. Gordon, H.J. Zeiger, C.H. Townes, *The Maser-New type of microwave amplifier, frequency standard, and spectrometer*, Phys. Rev. **99**, 1264 (1955).
- [8] T.H. Maiman, *Stimulated optical radiation in ruby Masers*, Nature **187**, 493 (1960).
- [9] A. Javan, W.B. Bennett Jr, D.R. Herriott, *Population inversion and continuous optical maser oscillation in a gas discharge containing He-Ne mixture*, Phys. Rev. Lett. **6**, 106 (1961).
- [10] E. Snitzer, *Optical maser action of Nd<sup>3+</sup> in a barium crown glass*, Phys. Rev. Lett. **7**, 444 (1961).
- [11] R.W. Hellwarth, *Control of fluorescent pulsations*, dans *Advances in Quantum Electronics*, J.R. Singer ed., Columbia Univ. Press, N.Y. (1961), p. 334.
- [12] R. Newman, *Excitation of the Nd<sup>3+</sup> fluorescence in CaWO<sub>4</sub> by recombination radiation in GaAs*, J. Appl. Phys. **34**, 437 (1963).  
R.J. Keyes, T.M. Quist, *Injection luminescent pumping of CaF<sub>2</sub>:U<sup>3+</sup> with GaAs diode lasers*, Appl. Phys. Lett. **4**, 50 (1964).
- [13] C.K. Patel, W.L. Faust, R.A. McFarlane, *CW laser action on rotational transitions of the  $\sum_u^+ = \sum_g^+$  vibrational band of CO<sub>2</sub>*, Bull. Am. Phys. Soc. **9**, 50 (1964).
- [14] J.E. Geusic, H.M. Marcos, L.G. Van Uittert, *Laser oscillations in Nd-doped yttrium aluminium, yttrium gallium and gadolinium garnets*, Appl. Phys. Lett. **4**, 182 (1964).
- [15] P.A. Franken, A.E. Hill, C.W. Peters, G. Weinreich, *Generation of optical harmonics*, Phys. Rev. Lett. **7**, 118 (1961).
- [16] J.A. Giordmaine, R.C. Miller, *Tunable Optical Parametric Oscillation in LiNbO<sub>3</sub> at optical frequencies*, Phys. Rev. Lett. **14**, 973 (1965).
- [17] N. Bloembergen, *Nonlinear Optics*, Benjamin, N.Y. (1965).
- [18] M. Bertolotti, *Masers and Lasers*, Adam Hilger (1983).
- [19] M. Le Bellac, *Physique quantique*, Collection savoirs actuels, EDP Sciences (2004).
- [20] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, Les éditions de Physique (1992).
- [21] R.F. Kazarinov, R.A. Suris, *Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice*, Sov. Phys. Semicond. **5**, 707 (1971).  
R.F. Kazarinov, R.A. Suris, *Electric and electromagnetic properties of semiconductors with a superlattice*, Sov. Phys. Semicond. **6**, 120 (1972).  
R.F. Kazarinov, R.A. Suris, *Theory of electrical properties of semiconductors with superlattices*, Sov. Phys. Semicond. **7**, 347 (1973).
- [22] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, *Quantum Cascade Laser*, Science **264**, 553 (1994).

- [23] J. Faist, F. Capasso, D.L. Sivco, A.L. Hutchinson, C. Sirtori, S.N.G. Chu, A.Y. Cho, *Quantum cascade laser: Temperature dependence of the performance characteristics and high  $T_0$  operation*, Appl. Phys. Lett. **65**, 2901 (1994).  
 J. Faist, F. Capasso, C. Sirtori, D.L. Sivco, A.L. Hutchinson, A.Y. Cho, *Vertical transition quantum cascade laser with Bragg confined excited state*, Appl. Phys. Lett. **66**, 538 (1995).  
 J. Faist, F. Capasso, C. Sirtori, D.L. Sivco, A.L. Hutchinson, A.Y. Cho, *Continuous wave operation of a vertical transition quantum cascade laser above  $T = 80$  K*, Appl. Phys. Lett. **67**, 3057 (1995).  
 J. Faist, F. Capasso, C. Sirtori, D.L. Sivco, J.N. Baillargeon, A.L. Hutchinson, S.-N.G. Chu, A.Y. Cho, *High power mid-infrared ( $\lambda \sim 5 \mu\text{m}$ ) quantum cascade lasers operating above room temperature*, Appl. Phys. Lett. **68**, 3680 (1996).  
 C. Sirtori, J. Faist, F. Capasso, D.L. Sivco, A.L. Hutchinson, A.Y. Cho, *Long wavelength infrared ( $\lambda \sim 11 \mu\text{m}$ ) quantum cascade lasers*, Appl. Phys. Lett. **69**, 2810 (1996).
- [24] C. Sirtori, *GaAs Quantum Cascade Lasers: Fundamentals and Performance*, école de la Société Française d'Optique (S.F.O.) 2002 aux Houches. Les lasers : Applications aux technologies de l'information et au traitement des matériaux. EdP Sciences : [www.bibsciences.org/bibsup/opt-coll/pub/2000/pdf/sirtori.pdf](http://www.bibsciences.org/bibsup/opt-coll/pub/2000/pdf/sirtori.pdf)
- [25] G.A. A'Skaryan, *Cherenkov radiation and transition radiation from electromagnetic waves*, Sov. Phys. JETP **15**, 943 (1962).  
 D.H. Auston, *Suppicosecond electrooptic shockwaves*, Appl. Phys. Lett. **43**, 713 (1983).  
 D.A. Kleinman, D.H. Auston, *Theory of electrooptic shock radiation in nonlinear optical media*, IEEE Journ. Quant. Elect. **QE-20**, 964 (1984).
- [26] R. Köhler, A. Tredicucci, F. Beltram, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, R.C. Iotti, F. Rossi, *Terahertz semiconductor heterostructure laser*, Nature **417**, 156 (2002).
- [27] Y. Jeong, J.K. Sahu, S. Baek, C. Alegria, D.B.S. Soh, C. Codemard, V. Philippov, D.J. Richardson, D.N. Payne, J. Nilsson, *Ytterbium-doped double-clad large-core fiber lasers with kW-level continuous-wave output power*, CLEO/IQEC 2004, San Francisco, May 16–21 2004, papier CMS1. <http://http://www.spioptics.com>
- [28] P.F. Moulton, *Spectroscopic and Laser Characteristics of  $\text{TiAl}_2\text{O}_3$* , J. Opt. Soc. Am. B **3**, 125 (1986).
- [29] A.E. Siegman, *Lasers*, University science Books, Mill Valley, California (1986).
- [30] D.E. Spence, P.N. Kean, W. Sibbett, *60-fs Pulse Generation From a Self-mode-locked Ti-Sapphire Laser*, Opt. Lett. **16**, 42 (1991).
- [31] I.D. Jung, F.X. Kartner, N. Matuschek, D.H. Sutter, F. Morier-Genoud, G. Zhang, U. Keller, V. Scheuer, M. Tilsch, T. Tschudi, *Self-starting 6.5-fs Pulses from a Ti: Sapphire Laser*, Opt. Lett. **22**, 1009 (1997).

- [32] U. Morgner, F.X. Kartner, S.H. Cho, Y. Chen, H.A. Haus, J.G. Fujimoto, E.P. Ippen, V. Scheuer, G. Angelow, T. Tschudi, *Sub-two-cycle Pulses from a Kerr-lens Mode-locked Ti: Sapphire Laser*, Opt. Lett. **24**, 411 (1999).
- [33] D.Strickland, G. Mourou, *Compression of amplified chirped of optical pulses*, Opt. Comm. **56**, 219 (1985).
- [34] G. Mourou, *The ultrahigh-peak-power laser: present and future*, Appl. Phys. B **65**, 205 (1997).
- [35] T. Tajima, G. Mourou, *Zettawatt-exawatt lasers and their applications in ultrastrong-field physics*, Phys. Rev. Special Topics – Accelerators and Beams **5**, 031301 (2002).
- [36] J. Raffy, T. Debuisschert, J.P. Pocholle, M. Papuchon, *Tunable IR laser source with optical parametric oscillators in series*, Appl. Opt. **33**, 985 (1994).  
J. Raffy, T. Debuisschert, J.P. Pocholle, *Widely tunable optical parametric oscillator with electrical wavelength control*, Opt. Lett. **22**, 1589 (1997).  
T. Debuisschert, J. Raffy, J.M. Dupont, J.P. Pocholle, *Nanosecond optical parametric oscillators*, Comptes Rendus de l'Academie des sciences Series IV Physics, July 2000, Vol. **1**, Elsevier Science, 561.
- [37] C. Fabre, E. Giacobino, A. Heidmann, L. Lugiato, S. Reynaud, M. Vadamchino, Wang Kaige, *Squeezing in detuned degenerate optical parametric oscillators*, Quant. Opt. **2**, 159 (1990).  
Ph. Grangier, J.A. Levenson, J.Ph. Poizat, *Quantum non-demolition measurements in optics*, Nature **396**, 537 (1998).
- [38] E. Spitz, *Système optique d'emménagement et de lecture d'information*, Brevet n° 1 589 067 23, mars 1970.  
E. Spitz, J.P. Huignard, C. Puech, *Early experiments on optical disc storage*, IEEE Journ. Selected Topics Quant. Electron. **6**, 1413 (2000).
- [39] M.N. Baibich, J.M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, J. Chazelas, *Giant Magnetoresistance of (001)/Fe/(001)Cr magnetic superlattices*, Phys. Rev. Lett. **61**, 2472 (1998).
- [40] F. Capasso, R. Paiella, R. Martini, R. Colombelli, C. Gmachl, T.L. Myers, M.S. Taubman, R.M. Williams, C.G. Bethea, K. Unterrainer, H.Y. Hwang, D.L. Sivco, A.Y. Cho, A.M. Sergent, H.C. Liu, E.A. Whittaker, *Quantum Cascade Lasers: Ultrahigh-Speed Operation, Optical Wireless Communication, Narrow Linewidth, and Far-Infrared Emission*, IEEE Journ. Quant. Electron. **38**, 511 (2002).  
J.P. Pocholle, *Transmissions optiques en propagation libre dans l'atmosphère : Comparaison des systèmes PIR & MIR / IR*, Note interne THALES Research & Technology, juin 2001.
- [41] D.E. Smith, M.T. Zuber, H.V. Frey, J.B. Garvin, J.W. Head, D.O. Muhleman, G.H. Pettengill, R.J. Phillips, S.C. Solomon, H.J. Zwally,

- W.B. Banerdt, T.C. Duxbury, *Topography of the northern hemisphere of Mars from the Mars Orbiter Laser Altimeter*, Science **279**, 1686 (1998).
- [42] J.B. Abshire, X. Sun, R.S. Afzal, *Mars Orbiter Laser Altimeter: receiver model and performance analysis*, Appl. Opt. **39**, 2449 (2000).
- [43] O. de Joinville, S. Saur, F. Bretar, *Le levé laser aéroporté : techniques, applications et recherche*, Bulletin d'Information de l'IGN n° 74, 2003, p. 52.
- [44] <https://peoiewswebinfo.monmouth.army.mil/JPSD/UrbanRecon>
- [45] G. Grynberg, A. Aspect, C. Fabre, *Introduction aux lasers et à l'optique quantique*, X École Polytechnique, Ellipses (1997).  
A. Cagnac, J.P. Faroux, *Lasers. Interaction lumière-atomes*, EDP Sciences (2002).
- [46] G. Sagnac, *L'éther lumineux démontré par l'effet du vent relatif d'éther dans un interféromètre en rotation uniforme*, Comptes-Rendus de l'Académie des sciences, Vol. 157, 27 oct. 1913, p. 708.
- [47] G. Sagnac, *Sur la preuve de la réalité de l'éther lumineux par l'expérience de l'interféromètre tournant*, Comptes-Rendus de l'Académie des sciences, Vol. 157, 22 déc. 1913, p. 1410.
- [48] W.M. Macek, D.T.M. Davies Jr., *Rotation rate sensing with traveling-wave ring lasers*, Appl. Phys. Lett. **2**, 67 (1963).
- [49] W.M. Macek, D.T.M. Davies Jr., R.W. Olthuis, J.R. Schneider, G.R. White, *Ring laser rotation rate sensor*, Symp. on Optical Masers Polytec. Inst. of Brooklyn, Apr. 16–19 1963, p. 199.
- [50] P.K. Cheo, C.V. Heer, *Beat frequency between two traveling waves in a Fabry-Perot square cavity*, Appl. Opt. **3**, 788 (1964).
- [51] F. Aronowitz, *Theory of a travelling-wave optical maser*, Phys. Rev. **139 A**, 635 (1965).  
F. Aronowitz, *The laser gyro*, dans *Laser Applications*, M. Ross Ed., Academic Press, 1971, pp. 134–197.
- [52] W.W. Chow, J. Gea-Banacloche, L.M. Pedrotti, V.E. Sanders, W. Schleich, M.O. Scully, *The ring laser gyro*, Rev. Mod. Phys. **57**, 61 (1985).
- [53] M. Sargent III, M.O. Scully, W.E. Lamb Jr., *Laser Physics*, Addison-Wesley (1987).
- [54] T.L. Gustavson, P. Bouyer, M.A. Kasevich, *Precision Rotation Measurements with an Atom Interferometer Gyroscope*, Phys. Rev. Lett. **78**, 2046 (1997).
- [55] Ch.J. Bordé, *Applications of Atom Interferometry to Fundamental Physics on Earth and in Space*, Erice (2001).
- [56] A. Ashkin, J.M. Dziedzic, *Optical levitation by radiation pressure*, Appl. Phys. Lett. **19**, 283 (1971).  
A. Ashkin, J.M. Dziedzic, *Stability of optical levitation by radiation pressure*, Appl. Phys. Lett. **24**, 586 (1974).  
G. Roosen, *A theoretical and experimental study of the stable equilibrium positions of spheres levitated by two horizontal laser beams*, Opt. Comm. **21**, 189 (1977).

- [57] J.P. Pocholle, J.P. Raffy, Y. Combemale, M. Papuchon, G. Roosen, M.T. Plantegenest, *Optical levitation using single mode fibers and its application to self-centering of microlenses*, Appl. Phys. Lett. **45**, 350 (1984).
- [58] M. von Allmen, A. Blatter, *Laser-beam interactions with materials: Physical principles and applications*, Springer series in materials science (1995).
- [59] B.N. Chichkov *et al.*, *Femtoseconde, picoseconde and nanoseconde laser ablation of solids*, Appl. Phys. **A 63**, 109 (1996).
- [60] *Les lasers et leurs applications scientifiques et médicales*, Collection de la Société Française d'Optique, C. Fabre, J.P. Pocholle Eds., Éditions de Physique (1996).



# Chapitre 5

## Le mouvement brownien et le théorème de fluctuation-dissipation

Par *Bernard Derrida* et *Éric Brunet*

### Sommaire

<b>5.1</b>	<b>Introduction</b>	<b>207</b>
5.1.1	Le mouvement brownien	208
5.1.2	Le théorème de fluctuation-dissipation	209
<b>5.2</b>	<b>Équation de Langevin</b>	<b>211</b>
5.2.1	Forces $L(t)$ non corrélées	211
5.2.2	Généralisation à des forces $L(t)$ quelconques	212
5.2.3	Validité et limitations de la théorie de Langevin	213
5.2.4	Bruit de Johnson-Nyquist (1928)	214
5.2.5	Exemple : une particule immergée dans un gaz parfait	216
<b>5.3</b>	<b>Équation de Fokker-Planck et équation de la diffusion</b>	<b>217</b>
5.3.1	Force de Langevin en présence d'un potentiel	217
5.3.2	Équation de Fokker-Planck	218
5.3.3	Équation de la diffusion	219
<b>5.4</b>	<b>Irrégularité du mouvement brownien</b>	<b>220</b>
<b>5.5</b>	<b>Somme de variables aléatoires</b>	<b>222</b>
5.5.1	Le théorème de la limite centrale	223
5.5.2	Grandes déviations	224
<b>5.6</b>	<b>Lois de Lévy</b>	<b>225</b>
<b>5.7</b>	<b>Marches aléatoires</b>	<b>228</b>

5.7.1	Marches aléatoires en temps continu sur un réseau régulier . . .	228
5.7.2	Effet de la dimension de l'espace : marche récurrente ou transiente . . . . .	230
5.7.3	Cas d'une marche aléatoire en temps discret . . . . .	231
<b>5.8</b>	<b>Polymères . . . . .</b>	<b>232</b>
5.8.1	L'effet du volume exclu . . . . .	233
5.8.2	Le modèle de Domb Joyce . . . . .	234
<b>5.9</b>	<b>Diffusion en présence de désordre . . . . .</b>	<b>236</b>
5.9.1	Les systèmes désordonnés . . . . .	236
5.9.2	Le modèle de Sinai . . . . .	237
<b>5.10</b>	<b>Théorème de fluctuation-dissipation . . . . .</b>	<b>239</b>
5.10.1	Chaleur spécifique . . . . .	239
5.10.2	Susceptibilité magnétique . . . . .	239
5.10.3	Compressibilité d'un fluide . . . . .	240
<b>5.11</b>	<b>Processus de Markov et notion de bilan détaillé . . . . .</b>	<b>242</b>
5.11.1	Validité d'une description markovienne . . . . .	243
5.11.2	Bilan détaillé et invariance par renversement du temps . . . . .	243
<b>5.12</b>	<b>Système en contact avec un ou plusieurs thermostats . . . . .</b>	<b>244</b>
5.12.1	Condition de bilan détaillé . . . . .	246
5.12.2	Système en contact avec un seul thermostat . . . . .	246
5.12.3	Système en contact avec plusieurs thermostats : équilibre . . . . .	247
5.12.4	Système en contact avec deux thermostats : régime stationnaire hors d'équilibre . . . . .	248
5.12.5	Relation de Gallavotti-Cohen . . . . .	249
5.12.6	Validité du second principe de la thermodynamique . . . . .	250
5.12.7	Théorème de fluctuation . . . . .	250
5.12.8	Exemple : une particule brownienne couplée à deux thermostats	251
<b>5.13</b>	<b>Système en contact avec des réservoirs de particules . . . . .</b>	<b>253</b>
5.13.1	Bruit Schottky . . . . .	254
5.13.2	Relations de réciprocity d'Onsager . . . . .	255
<b>5.14</b>	<b>Théorème de fluctuation-dissipation dynamique . . . . .</b>	<b>256</b>
<b>5.15</b>	<b>Systèmes loin de l'équilibre . . . . .</b>	<b>259</b>
5.15.1	Le vieillissement . . . . .	259
5.15.2	Croissance de domaines . . . . .	261
<b>5.16</b>	<b>Références . . . . .</b>	<b>264</b>





### **Bernard Derrida**

Bernard Derrida est professeur de physique à l'Université Pierre et Marie Curie et à l'École Normale Supérieure. Ses travaux de recherche, effectués jusqu'en 1993 au service de physique théorique de Saclay et depuis au laboratoire de physique statistique de l'École Normale Supérieure, ont surtout été consacrés à la théorie des systèmes désordonnés, aux systèmes hors d'équilibre et à certaines applications de la physique statistique à la biologie.



### **Éric Brunet**

Éric Brunet est enseignant et chercheur en physique à l'Université Pierre et Marie Curie et à l'École Normale Supérieure, dans le laboratoire de physique statistique où il travaille avec Bernard Derrida. Ses recherches portent sur les systèmes hors d'équilibre, en particulier la propagation de fronts et les phénomènes de transport, ainsi que sur les systèmes désordonnés.



# Le mouvement brownien et le théorème de fluctuation-dissipation

## 5.1 Introduction

Une des plus grandes révolutions scientifiques de tous les temps fut la prise de conscience progressive, tout au long du XIX<sup>e</sup> siècle, de la nature atomique de la matière. Utilisés au départ par les physiciens et les chimistes comme un outil commode de raisonnement ou comme un artifice mathématique, les atomes devinrent peu à peu une hypothèse incontournable, en particulier grâce aux succès de la théorie cinétique des gaz développée par Clausius, Maxwell et Boltzmann. Vers la fin du XIX<sup>e</sup> siècle, le problème ouvert était davantage celui d'une détermination précise de la taille des atomes que celui de leur existence. Dans des notes autobiographiques [1], Einstein dit à propos de certaines de ses contributions à la mécanique statistique : « *Mon but principal était de trouver des faits prouvant autant que possible l'existence d'atomes de dimension précise et finie.* »

Si la matière était faite d'atomes et de molécules, il fallait comprendre comment les lois de la physique *macroscopique* que l'on connaissait alors pouvaient être déduites des interactions élémentaires entre atomes à l'échelle *microscopique*. Ce fut la naissance, avec Boltzmann puis Gibbs, d'une nouvelle branche de la physique : la mécanique statistique. À cause du nombre astronomique d'atomes contenus dans chaque parcelle de matière macroscopique, des lois statistiques, comme par exemple la loi des grands nombres, permettent de comprendre pourquoi la physique macroscopique nous apparaît comme déterministe, même quand l'état microscopique du système n'est en fait connu que de manière très approximative.

Il n'en demeure pas moins que les lois déterministes de la physique macroscopique ne décrivent les phénomènes naturels qu'aux fluctuations près. Dans un certain nombre de cas, par exemple si l'on s'intéresse à des systèmes assez petits ou à certaines propriétés plus sensibles, ces fluctuations deviennent

observables. Deux des immenses contributions d'Einstein à la mécanique statistique, la théorie du mouvement brownien et le théorème de fluctuation-dissipation, qui sont l'objet de ce chapitre, furent de donner une prédiction quantitative de ces fluctuations, dont la mesure permettait de déterminer la constante de Boltzmann  $k_B$ , ou, ce qui est équivalent, le nombre d'Avogadro et donc la taille des atomes [2].

### 5.1.1 Le mouvement brownien

Le terme de mouvement brownien provient de l'observation et de l'étude au microscope, par le botaniste Robert Brown, en 1828, du mouvement apparemment erratique de particules légères telles des grains de pollen en suspension dans l'eau (Brown ne fut pas le premier à observer un mouvement brownien : il cite lui-même plusieurs précurseurs et cela dès le XVII<sup>e</sup> siècle). Au cours du XIX<sup>e</sup> siècle, plusieurs explications de ce phénomène furent avancées, à commencer par une origine organique [3]. L'observation de mouvements similaires pour des particules microscopiques inorganiques fit pencher la balance pour une origine mécanique. De nombreuses expériences, en particulier celles de Gouy, établirent que la trajectoire des particules est très irrégulière et n'a pas de tangente, que l'amplitude du mouvement augmente quand la taille des particules diminue, quand la viscosité du fluide diminue, ou quand la température augmente, et que les mouvements de particules placées dans le même fluide sont indépendants. Vers 1905, l'explication la plus satisfaisante était que le mouvement brownien des particules microscopiques était dû à leur bombardement incessant par les molécules du fluide. Un test quantitatif simple de cette théorie semblait être de vérifier que l'énergie cinétique moyenne de la particule est donnée par  $(3/2)k_B T$  comme le prévoit la théorie cinétique des gaz à l'équilibre. Pour cela, il suffisait en principe de mesurer la vitesse de la particule au cours du temps. Mais la difficulté à mesurer la vitesse instantanée (n'oublions pas que la trajectoire est continue mais n'a pas de tangente comme le suggèrent les quelques trajectoires de la figure 5.1) donna des facteurs d'ordre 100 000 entre les valeurs mesurées et les valeurs attendues [3]. La raison de ce désaccord est que la vitesse instantanée change sur des échelles de temps beaucoup trop courtes, de l'ordre de  $10^{-7}$  secondes, pour être mesurée.

Cette difficulté fut évitée par la théorie d'Einstein du mouvement brownien de 1905, qui, au lieu de penser le phénomène en termes de vitesse, essaie de le décrire en termes de la position de la particule. À la suite d'une collision avec une molécule du fluide, une particule (un grain de pollen, par exemple), que l'on suppose initialement au repos, acquiert une vitesse. Cette vitesse est amortie par le fluide environnant (par la force de frottement visqueux, proportionnelle à cette vitesse) et le résultat net de la collision (obtenu en intégrant la vitesse depuis l'instant de la collision jusqu'à son arrêt) est un déplacement de la particule d'une distance finie.

Comme la vitesse et la direction des molécules du fluide sont aléatoires, chaque collision produit ainsi un déplacement aléatoire de la particule.

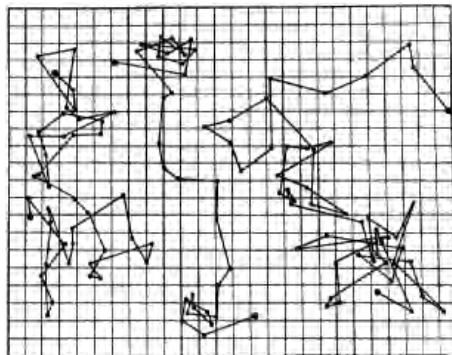


FIG. 5.1 – *Mouvements browniens du livre de Jean Perrin « les Atomes » [4].*

L'addition de ce grand nombre de déplacements élémentaires aléatoires conduit à une équation de diffusion pour la probabilité  $P(\vec{x}, t)$  de trouver la particule au point  $\vec{x}$  à l'instant  $t$ . La contribution majeure d'Einstein à la théorie du mouvement brownien fut d'écrire cette équation de diffusion et de donner une expression de la constante de diffusion de la particule en fonction de sa taille, de la température et de la viscosité du fluide environnant. Une expression très voisine de cette constante de diffusion fut obtenue indépendamment et publiée par Smoluchowski en 1906.

Si l'on place un certain nombre de particules dans un fluide, leur concentration a tendance à s'uniformiser sous l'effet de la diffusion comme le prévoit la théorie du mouvement brownien. Par ailleurs, la gravité tend à faire tomber les particules au fond du récipient qui les contient. Sans la diffusion brownienne, toutes les particules se déposeraient au fond du récipient. Sans la gravité, la densité deviendrait uniforme. L'effet combiné de la diffusion et de la gravité est de produire un profil de concentration (voir Fig. 5.2) qui dépend de la hauteur et dont la mesure permet, à Jean Perrin et à ses élèves, de valider expérimentalement la formule d'Einstein de la constante de diffusion du mouvement brownien [5] et de prédire une valeur du nombre d'Avogadro compatible avec celles connues jusqu'alors à partir de la théorie cinétique des gaz. Cette validation expérimentale fut considérée à l'époque comme une évidence irréfutable de l'existence des atomes [3].

### 5.1.2 Le théorème de fluctuation-dissipation

L'un des buts d'Einstein, dans son travail de thèse achevé en 1905, était de comprendre comment déterminer le nombre d'Avogadro à partir des résultats d'expériences faites à notre échelle. Dans sa thèse, Einstein obtient  $2,1 \times 10^{23}$  à partir de données sur la pression osmotique de solutions aqueuses sucrées, qu'il corrigera, grâce à des données expérimentales plus précises, en  $4,15 \times 10^{23}$  dans la version publiée l'année suivante [1]. La valeur admise de nos jours est

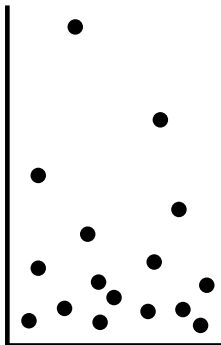


FIG. 5.2 – Représentation de la densité de particules dans un récipient sous l'action combinée de la gravité et de l'agitation thermique.

$6,02 \times 10^{23}$ . Pour parvenir à ce résultat, l'idée d'Einstein est de montrer, à partir de l'expression de Boltzmann de l'entropie, que les fluctuations de quantités macroscopiques sont gaussiennes avec une variance que l'on peut exprimer au moyen de la constante de Boltzmann  $k_B$ . Einstein considère un système isolé, à une certaine énergie  $E$ . La théorie de Boltzmann donne la probabilité  $P(A)$  qu'une certaine quantité physique prenne une valeur  $A$  :

$$P(A) = \frac{\Omega(E, A)}{\sum_{A'} \Omega(E, A')}$$

où  $\Omega(E, A)$  est le nombre de configurations microscopiques d'énergie  $E$  pour lesquelles l'observable prend la valeur  $A$ . En utilisant l'expression de l'entropie de Boltzmann

$$S(E, A) = k_B \log \Omega(E, A)$$

on obtient, en développant autour de la valeur  $A^*$  qui rend cette entropie maximale,

$$P(A) \sim \exp \left[ -\frac{(A - A^*)^2}{2k_B} \frac{\partial^2 S(E, A^*)}{\partial A^2} \right]$$

ce qui donne, si la constante de Boltzmann  $k_B$  est suffisamment petite pour qu'on puisse négliger la suite du développement, des fluctuations gaussiennes de variance  $k_B / \left( \frac{\partial^2 S(E, A^*)}{\partial A^2} \right)$ . La mesure de ces fluctuations permet donc de déterminer la constante de Boltzmann  $k_B$  si l'on connaît  $\frac{\partial^2 S(E, A^*)}{\partial A^2}$ . Mais cette dérivée est une quantité purement macroscopique ; comme nous le verrons sur des exemples concrets, elle mesure la réponse linéaire de l'observable macroscopique  $A^*$  à un faible champ conjugué à  $A$ . Cette relation, entre les fluctuations d'équilibre d'une observable et la réponse de cette observable à un champ, constitue ce qu'on appelle le théorème de fluctuation-dissipation.

## 5.2 Équation de Langevin

L'approche sans doute la plus souvent utilisée [6] pour décrire le mouvement brownien est celle introduite par Langevin dans son article de 1908 qui fait suite aux travaux publiés en 1905 par Einstein et en 1906 par Smoluchowski. Dans la théorie de Langevin, la force exercée par le fluide sur la particule dont on observe le mouvement brownien est décrite par une somme de deux termes ; une force de frottement  $-\gamma V$  (proportionnelle à la vitesse  $V$  de la particule) qui tend à la ralentir et une force aléatoire  $L(t)$  de moyenne nulle qui représente les collisions avec les molécules du fluide :

$$\boxed{m \frac{dV}{dt} = -\gamma V + L(t)} \tag{5.1}$$

Ces deux forces  $-\gamma V$  et  $L(t)$  ont toutes les deux comme même origine les interactions de la particule avec les molécules du fluide : le terme de frottement représente la force moyenne due aux collisions avec les molécules du fluide et  $L(t)$  représente la partie fluctuante de ces forces de collision. Dans (5.1) comme dans ce qui suit, le calcul est fait dans le cas unidimensionnel ; sa généralisation à trois dimensions est totalement évidente.

### 5.2.1 Forces $L(t)$ non corrélées

On suppose que la force  $L(t)$  est un bruit blanc gaussien, c'est-à-dire que les valeurs de  $L(t)$  à deux instants différents sont des variables indépendantes et que les corrélations qui caractérisent cette force gaussienne sont de la forme

$$\boxed{\langle L(t)L(t') \rangle = \Gamma \delta(t - t')} \tag{5.2}$$

(On appelle cela un bruit blanc parce que toutes les composantes de Fourier du bruit ont la même variance et donc toutes les fréquences sont le même poids.)

Parce qu'elle est linéaire, on peut facilement intégrer l'équation de Langevin (5.1) pour obtenir

$$V(t) = V(0)e^{-\gamma t/m} + \frac{1}{m} \int_0^t d\tau e^{-\gamma(t-\tau)/m} L(\tau) \tag{5.3}$$

Comme la vitesse est une fonction linéaire de variables gaussiennes, elle est elle-même une variable gaussienne. Sa valeur moyenne  $\langle V(t) \rangle = V(0) \exp(-\gamma t/m)$  tend rapidement vers zéro et sa variance, obtenue en moyennant  $V(t)^2$  sur les  $L(t)$ , est donnée par

$$\langle V(t)^2 \rangle - \langle V(t) \rangle^2 = \frac{\Gamma}{2m\gamma} \left( 1 - e^{-2\gamma t/m} \right)$$

Dans la limite des temps longs, la distribution gaussienne de la vitesse doit tendre vers la distribution d'équilibre à la température  $T$ . On doit donc avoir, pour  $t \rightarrow \infty$ ,

$$\frac{1}{2}m\langle V(t)^2 \rangle \rightarrow \frac{1}{2}k_B T$$

ce qui impose

$$\boxed{\Gamma = 2\gamma k_B T} \quad (5.4)$$

Cette relation entre la variance  $\Gamma$  du bruit et le coefficient  $\gamma$  de la force de frottement est très générale. Des forces comme celle qui apparaît dans le membre de droite de (5.1), composées d'un terme de frottement et d'une force aléatoire vérifiant (5.2, 5.4) sont très souvent utilisées pour modéliser la dynamique d'un système en contact avec un thermostat, la température  $T$  du thermostat étant en général définie à partir de la relation (5.4).

En intégrant l'équation (5.3), on obtient le déplacement  $x(t) - x(t')$  effectué par la particule entre deux temps  $t > t'$ . Lorsque  $t$  et  $t'$  sont suffisamment grands pour qu'on ait oublié la vitesse initiale, on a

$$x(t) - x(t') = \int_0^t d\tau \frac{1 - e^{-\gamma(t-\tau)/m}}{\gamma} L(\tau) - \int_0^{t'} d\tau \frac{1 - e^{-\gamma(t'-\tau)/m}}{\gamma} L(\tau) \quad (5.5)$$

Si l'on moyenne sur la force aléatoire le carré de cette différence, on obtient, en utilisant (5.4) et en supposant  $t$  et  $t'$  grands

$$\langle [x(t) - x(t')]^2 \rangle = \frac{2k_B T}{\gamma} (t - t') - \frac{2mk_B T}{\gamma^2} \left(1 - e^{-\gamma(t-t')/m}\right) \quad (5.6)$$

Si, de plus, la différence  $t - t'$  est elle aussi grande, l'expression précédente est dominée par le premier terme et l'on obtient un comportement diffusif

$$\boxed{\langle [x(t) - x(t')]^2 \rangle \simeq 2D(t - t')} \quad (5.7)$$

avec une constante de diffusion  $D$  donnée par

$$\boxed{D = \frac{k_B T}{\gamma}} \quad (5.8)$$

### 5.2.2 Généralisation à des forces $L(t)$ quelconques

On peut répéter le calcul qui précède pour une force  $L(t)$  de moyenne nulle, caractérisée par des fonctions de corrélations quelconques dans le temps. Si la fonction de corrélations  $\langle L(t)L(t') \rangle$  ne dépend que de la différence de temps  $t - t'$ , c'est-à-dire si

$$\langle L(t)L(t') \rangle = \langle L(0)L(t' - t) \rangle$$



on retrouve le comportement diffusif (5.7) avec une constante de diffusion donnée par :

$$D = \frac{1}{2\gamma^2} \int_{-\infty}^{\infty} \langle L(0)L(\tau) \rangle d\tau$$

(où l'on a supposé que les fonctions de corrélations de la force décroissent suffisamment vite pour que l'intégrale converge). Dans le régime stationnaire, quand la vitesse initiale  $V(0)$  dans (5.3) a été oubliée, on obtient, en moyennant sur la force aléatoire  $L(t)$ ,

$$\langle V(t)V(t') \rangle = \frac{1}{2m\gamma} \int_{-\infty}^{\infty} du e^{-\gamma|u|/m} \langle L(0)L(u+t-t') \rangle$$

On peut alors vérifier la relation habituelle qui relie la constante de diffusion aux corrélations de la vitesse ; pour  $t$  grand :

$$\boxed{D = \int_0^{\infty} du \langle V(t)V(t+u) \rangle} \tag{5.9}$$

(Pour  $t$  grand, cette expression ne dépend pas de  $t$ .)

### 5.2.3 Validité et limitations de la théorie de Langevin

La description par (5.1) d'une particule de masse  $m$  soumise aux collisions des molécules d'un liquide n'est bien sûr qu'une approximation. Une théorie complète, même en se limitant à des particules classiques, est beaucoup plus difficile : on doit en principe écrire les équations qui gouvernent la dynamique de toutes les molécules du liquide, en prenant à la fois en compte leurs interactions mutuelles et les interactions avec la particule. En remplaçant l'effet des collisions avec les molécules du fluide par une force aléatoire  $L(t)$  non corrélée dans le temps, on a négligé le fait que les molécules qui interagissent avec la particule ont souvent des collisions multiples qui engendrent des corrélations entre la vitesse de la particule et celles des molécules qui l'entourent. La vraie force aléatoire due aux collisions possède donc des corrélations temporelles qui demandent un traitement beaucoup plus détaillé du phénomène.

En plus de ces corrélations temporelles négligées dans (5.1), la forme gaussienne de la force choisie pour décrire l'effet du fluide environnant sur la particule est elle aussi une approximation de la réalité : pour qu'on puisse considérer que la force est gaussienne, il faut que, pendant un temps court par rapport à l'échelle de temps caractéristique  $m/\gamma$  de la particule, il y ait beaucoup de collisions et donc que la particule soit suffisamment massive.

L'expression de la force effective  $-\gamma V + L(t)$  exercée par le fluide sur la particule, qui conduit à des corrélations de vitesse qui décroissent exponentiellement, est donc une approximation. Si l'on part d'une description microscopique des collisions avec les molécules du fluide, on sait, comme cela fut découvert par Alder et Wainwright en 1967 dans des simulations numériques [7] que les corrélations de vitesse ne décroissent pas aux temps longs

exponentiellement (comme le prévoit l'équation de Langevin pour un bruit blanc), mais comme une loi de puissance  $t^{-d/2}$  en fonction du temps, où  $d$  est la dimension de l'espace.

Si l'on adopte néanmoins, comme l'ont fait Einstein et Smoluchowski, un point de vue hydrodynamique, la force  $-\gamma V$  peut se calculer : le calcul de la force de frottement se réduit à celui de la force exercée par un fluide sur une particule se déplaçant à une vitesse uniforme. Elle dépend de la viscosité  $\nu$  du fluide environnant, de la forme et de la taille de la particule. Pour une particule sphérique de petit rayon  $r$  en mouvement uniforme à une faible vitesse  $V$ , cette force est donnée par la loi de Stokes (on peut en trouver une dérivation dans le chapitre 5 du livre de Huang [8]) avec

$$\gamma = 6\pi r\nu$$

qui, avec (5.8), donne pour l'expression d'Einstein de la constante de diffusion du mouvement brownien

$$D = \frac{k_B T}{6\pi r\nu}$$

Une particule de diamètre  $1\ \mu\text{m}$  dans l'eau parcourt une distance typique de l'ordre de  $1\ \mu\text{m}$  en une seconde [2].

## 5.2.4 Bruit de Johnson-Nyquist (1928)

L'étude qui précède d'une particule soumise à une force aléatoire dans le cadre de l'équation de Langevin (5.1) peut se généraliser facilement au cas du bruit Johnson-Nyquist. À l'équilibre thermique, la différence de potentiel aux bornes d'une résistance électrique  $R$  fluctue au cours du temps. Ces fluctuations d'origine thermique, mesurées par Johnson en 1928 et dont la théorie est due à Nyquist en 1928 également [9, 10], donnèrent à l'époque un moyen de mesurer la constante de Boltzmann  $k_B$  avec une précision de l'ordre de 8 %. Si  $U(t)$  est la tension aux bornes de la résistance, les fluctuations de  $U(t)$  sont caractérisées, comme nous allons le voir, par des corrélations temporelles

$$\langle U(t)U(t') \rangle = \int_0^\infty df G(f) \cos [2\pi f \times (t - t')] \quad (5.10)$$

où l'amplitude  $G(f)$  en fréquence du bruit est une constante égale à

$$G(f) = 4Rk_B T \quad (5.11)$$

dans la limite où  $1/f$  est grand devant le temps caractéristique du système.

On peut obtenir ces relations en considérant un circuit électrique, composé d'une résistance et d'un condensateur, comme celui schématisé sur la figure 5.3. On suppose que la charge  $Q$  du condensateur évolue selon une équation de type Langevin (5.1)

$$R \frac{dQ}{dt} = -\frac{Q(t)}{C} + \eta(t) \quad (5.12)$$

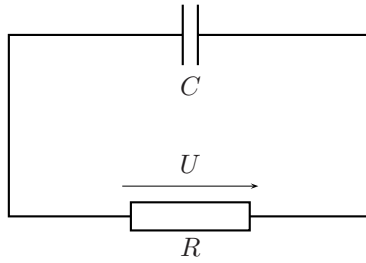


FIG. 5.3 – Bruit Nyquist dans un circuit RC.

où  $\eta(t)$  est un bruit blanc gaussien de moyenne nulle caractérisé par une variance  $\Gamma$  qui représente l'effet des interactions du circuit avec un thermostat à la température  $T$ .

$$\langle \eta(t)\eta(t') \rangle = \Gamma\delta(t - t')$$

L'équation (5.12) s'intègre facilement. Pour un temps suffisamment long pour que le système ait oublié la condition initiale, on a

$$Q(t) = \frac{1}{R} \int_{-\infty}^t d\tau \exp\left(-\frac{t-\tau}{RC}\right) \eta(\tau)$$

ce qui donne  $\langle Q \rangle = 0$  et  $\langle Q^2 \rangle = \Gamma C / (2R)$  pour les deux premiers moments de  $Q$ . Comme  $\eta(t)$  est gaussien et que  $Q$  linéaire en  $\eta$ , la distribution de  $Q$  est également gaussienne

$$P(Q) \sim \exp\left(-\frac{R}{\Gamma C} Q^2\right)$$

L'énergie d'un condensateur étant  $Q^2 / (2C)$ , la distribution de  $Q$ , quand le circuit est à l'équilibre thermique, est

$$P(Q) \sim \exp\left(-\frac{Q^2}{2Ck_B T}\right)$$

Cela fixe l'amplitude du bruit

$$\Gamma = 2Rk_B T$$

On obtient alors

$$\langle Q(t)Q(t') \rangle = k_B T C \exp\left(-\frac{|t-t'|}{RC}\right) \tag{5.13}$$

en utilisant  $Q(t) = CU(t)$  et en prenant la transformée de Fourier de (5.13), cela donne la densité spectrale  $G(f)$  introduite dans (5.10) :

$$G(f) = \frac{4Rk_B T}{1 + (2\pi fRC)^2}$$

qui se réduit à (5.11) pour les faibles fréquences.

### 5.2.5 Exemple : une particule immergée dans un gaz parfait

Dans le cas d'une particule effectuant un mouvement brownien dans un liquide, le libre parcours moyen des molécules de fluide (la distance parcourue entre deux collisions) est très faible (de l'ordre des distances entre les molécules). Les interactions entre molécules ne peuvent donc pas être négligées : dans la description hydrodynamique mentionnée plus haut, la viscosité est le paramètre effectif qui rend compte de ces interactions. Ici nous considérons, à titre d'exemple, le cas du gaz parfait où les molécules du fluide n'interagissent pas entre elles [11].

On considère donc une particule de masse  $M$  de vitesse  $V$  et on suppose qu'elle est immergée dans un gaz parfait de molécules de masse  $m$  à la densité  $\rho$  à l'équilibre. Nous allons voir que le bombardement de la particule par les molécules du gaz est parfaitement décrit par l'équation de Langevin (5.1). On se limite, pour simplifier, au cas unidimensionnel. Lors d'une collision de la particule de vitesse  $V$  avec une molécule de gaz à la vitesse  $v$ , la conservation de l'énergie et de l'impulsion ( $MV^2 + mv^2 = MV'^2 + mv'^2$  et  $MV + mv = MV' + mv'$ ) donne, pour les vitesses  $V'$  et  $v'$  de la particule et de la molécule après la collision

$$V' = \frac{(M - m)V + 2mv}{M + m} \quad \text{et} \quad v' = \frac{(m - M)v + 2MV}{M + m}$$

ce qui se traduit par une variation  $\Delta P$  de l'impulsion de la particule donnée par

$$\Delta P = M(V' - V) = \frac{2mM}{m + M}(v - V)$$

Pour  $M \gg m$ , chaque collision change peu la vitesse de la particule. Si  $\Delta t$  est un temps assez court pour que la vitesse  $V$  de la particule ait peu changé mais assez long pour qu'elle ait subi beaucoup de collisions avec les molécules du gaz parfait, on obtient

$$\Delta P = \frac{2mM}{m + M} \sum_v (v - V) n_{\Delta t}(v)$$

où l'on a découpé les valeurs de la vitesse  $v$  en petits intervalles de taille  $\delta v$  et où  $n_{\Delta t}(v)$  est le nombre de molécules du gaz avec une vitesse comprise entre  $v$  et  $v + \delta v$  ayant eu une collision avec la particule pendant l'intervalle de temps  $\Delta t$ . Ce nombre  $n_{\Delta t}(v)$  fluctue : en moyenne on a

$$\langle n_{\Delta t}(v) \rangle = \underbrace{\sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv^2}{2k_B T}\right)}_{\text{probabilité qu'une molécule ait la vitesse } v} \delta v \times \rho \times \underbrace{|v - V| \Delta t}_{\text{région d'où peuvent venir les molécules}}$$

De plus, si l'on choisit  $\delta v$  suffisamment petit, il n'y a aucun intervalle de vitesse ayant plus d'une molécule (la probabilité d'avoir plus d'une molécule

dans un intervalle de taille  $\delta v$  est d'ordre  $(\delta v)^2$  et les seules valeurs prises par  $n_{\Delta t}(v)$  sont 0 ou 1. On a donc  $\langle n_{\Delta t}(v) \rangle \ll 1$  et, à l'ordre dominant en  $\delta v$ ,

$$\langle n_{\Delta t}(v)^2 \rangle \simeq \langle n_{\Delta t}(v) \rangle \simeq \langle n_{\Delta t}(v)^2 \rangle - \langle n_{\Delta t}(v) \rangle^2$$

ce qui donne, pour  $V$  petit et pour  $m \ll M$ ,

$$\begin{aligned} \langle \Delta P \rangle &= \frac{2mM}{m+M} \sum_v (v-V) \langle n_{\Delta t}(v) \rangle \simeq -4\sqrt{\frac{2mk_B T}{\pi}} V \rho \Delta t \\ \langle (\Delta P)^2 \rangle - \langle \Delta P \rangle^2 &= \left( \frac{2mM}{m+M} \right)^2 \sum_v (v-V)^2 \left[ \langle n_{\Delta t}(v)^2 \rangle - \langle n_{\Delta t}(v) \rangle^2 \right] \\ &\simeq 8k_B T \sqrt{\frac{2mk_B T}{\pi}} \rho \Delta t \end{aligned}$$

Si l'on compare aux deux termes de l'équation de Langevin (5.1), en utilisant que la force moyenne exercée pendant le temps  $\Delta t$  est  $\Delta P/\Delta t$ , on constate que l'évolution de la vitesse de la particule est bien donnée par l'équation de Langevin (5.1) avec

$$\gamma = 4\rho \sqrt{\frac{2mk_B T}{\pi}} \quad \text{et} \quad \Gamma = 8k_B T \rho \sqrt{\frac{2mk_B T}{\pi}}$$

où  $\gamma$  et  $\Gamma$  vérifient (5.4). Les collisions avec les molécules du gaz parfait thermalisent donc bien la particule, comme le prévoit l'équation de Langevin.

## 5.3 Équation de Fokker-Planck et équation de la diffusion

Dès que l'on considère une évolution stochastique, comme celle décrite par l'équation de Langevin (5.1), avec un terme aléatoire, la valeur de la solution à un instant  $t$  est elle aussi une variable aléatoire et il faut considérer la distribution de probabilité de cette solution. C'est précisément ce que fait l'équation de Fokker-Planck dans le cadre général d'équations stochastiques, et l'équation de la diffusion (ou équation de la chaleur) dans le cas particulier du mouvement brownien.

### 5.3.1 Force de Langevin en présence d'un potentiel

On peut introduire des forces aléatoires comme celle qui apparaît dans l'équation de Langevin pour décrire l'effet d'un thermostat ou de plusieurs thermostats sur un grand nombre de systèmes. Par exemple, pour un système de particules classiques de coordonnées  $x_1, x_2, \dots$  interagissant avec un potentiel de paires  $u(x_i - x_j)$  et placées dans un potentiel extérieur  $U(x)$ , les

équations du mouvement prennent la forme

$$m \frac{d^2 x_i}{dt^2} = -U'(x_i) - \sum_{j \neq i} u'(x_i - x_j) - \gamma_i \frac{dx_i}{dt} + L_i(t) \quad (5.14)$$

Chaque particule est donc soumise à une force additionnelle  $-\gamma_i \frac{dx_i}{dt} + L_i(t)$ , et si la force  $L_i(t)$  satisfait

$$\langle L_i(t) L_i(t') \rangle = \Gamma_i \delta(t - t') \quad \text{avec} \quad \Gamma_i = 2\gamma_i k_B T_i$$

cela modélise le couplage de la particule  $i$  avec un thermostat à la température  $T_i$ , le paramètre  $\gamma_i$ , étant une mesure de ce couplage.

On peut par exemple coupler différentes parties d'un système à des thermostats à des températures différentes pour étudier des propriétés de transport : ainsi, si les  $x_i$  représentent l'écart à leurs positions d'équilibre des atomes d'un solide, on peut coupler les deux bords du solide à des thermostats à des températures différentes, couper tout contact avec un thermostat pour les autres atomes et essayer de calculer l'énergie transférée d'un thermostat vers l'autre à travers le solide.

### 5.3.2 Équation de Fokker-Planck

En couplant un système à un seul thermostat (ce qui revient à choisir la même température  $T$  pour tous les degrés de liberté, c'est-à-dire à imposer que  $\Gamma_i = 2\gamma_i k_B T$  pour tout  $i$ ), on peut montrer que l'effet des forces de Langevin  $-\gamma_i \frac{dx_i}{dt} + L_i(t)$  est d'amener le système à l'équilibre à la température  $T$ .

L'exemple sans doute le plus simple est celui d'une particule très légère dans un potentiel unidimensionnel  $U(x)$ . L'équation d'évolution de cette particule s'écrit

$$m \frac{d^2 x}{dt^2} = -\frac{dU(x)}{dx} - \gamma \frac{dx}{dt} + L(t)$$

et si l'on considère la masse suffisamment petite ou la viscosité suffisamment forte, on obtient

$$\gamma \frac{dx}{dt} = -\frac{dU(x)}{dx} + L(t) \quad \text{avec} \quad \langle L(t) L(t') \rangle = 2\gamma k_B T \delta(t - t')$$

Si l'on découpe le temps en petits intervalles  $\Delta t$ , la moyenne temporelle  $F$  de la force  $L(t)$  sur chaque intervalle de temps  $\Delta t$  est une variable gaussienne d'espérance nulle et de variance  $\langle F^2 \rangle = (2\gamma k_B T / \Delta t)$ . La variation de la position au bout d'un intervalle de temps  $\Delta t$  peut s'écrire

$$x(t + \Delta t) = x(t) + \gamma^{-1} [F - U'(x(t))] \Delta t$$

et la probabilité  $P_t(x)$  de trouver la particule au point  $x$  à l'instant  $t$  vérifie

$$P_{t+\Delta t}(x) = \left\langle \int dx' P_t(x') \delta(x' + \gamma^{-1}[F - U'(x')] \Delta t - x) \right\rangle_F$$

$$\simeq \left\langle \frac{P(x - \gamma^{-1}[F - U'(x)] \Delta t)}{1 - \gamma^{-1}U''(x)\Delta t} \right\rangle_F$$

ce qui conduit, en développant au premier ordre en  $\Delta t$  et au second ordre en  $F\Delta t$  (car les valeurs typiques de  $F$  sont d'ordre  $\Delta t^{-1/2}$ ), à l'équation de Fokker-Planck

$$\boxed{\gamma \frac{\partial P_t(x)}{\partial t} = \frac{\partial}{\partial x} \left[ P_t(x) \frac{\partial U(x)}{\partial x} \right] + k_B T \frac{\partial^2}{\partial x^2} P_t(x)} \quad (5.15)$$

Il est facile de vérifier que

$$P_t(x) = C e^{-\frac{U(x)}{k_B T}} \quad (5.16)$$

est bien une solution stationnaire de l'équation de Fokker-Planck et donc, quand le potentiel  $U(x)$  croît suffisamment avec  $x$  pour que l'on puisse normaliser l'expression (5.16), l'état stationnaire est bien l'équilibre de Boltzmann-Gibbs.

Ce calcul peut être généralisé à un nombre quelconque de degrés de liberté, et aussi au cas où les masses des particules sont prises en compte comme en (5.14).

### 5.3.3 Équation de la diffusion

En l'absence de potentiel externe  $U(x)$ , l'équation de Fokker-Planck (5.15) se réduit à l'équation de diffusion

$$\boxed{\gamma \frac{\partial P_t(x)}{\partial t} = k_B T \frac{\partial^2}{\partial x^2} P_t(x)} \quad (5.17)$$

dont la solution, pour une particule initialement localisée au point  $x'$ , à l'instant  $t'$ , est

$$\boxed{P_t(x) = \frac{1}{\sqrt{4\pi D \times (t - t')}} \exp \left[ -\frac{(x - x')^2}{4D \times (t - t')} \right]} \quad (5.18)$$

où la constante de diffusion  $D = k_B T / \gamma$  est exactement la même que celle obtenue (5.8) pour une particule de masse non-nulle.

L'équation de la diffusion et les probabilités gaussiennes se retrouvent un peu partout en physique et ailleurs, dès qu'un effet résulte de l'addition d'un grand nombre de phénomènes élémentaires dont les caractéristiques fluctuent.

Sur le plan mathématique, elles furent introduites, quelques années avant les contributions d'Einstein, par L. Bachelier en 1900 qui modélisa ainsi l'évolution des cours de la Bourse de Paris. Dans le cas du mouvement brownien, chaque collision produit un déplacement de la particule et ce déplacement fluctue à cause des caractéristiques de la molécule incidente qui varient d'une collision à l'autre.

Ce que l'équation de la diffusion et la forme gaussienne ont de remarquable est qu'elles ne dépendent pas de la forme précise des interactions entre la particule et les molécules. Le fait que le déplacement de la particule résulte de l'addition d'un grand nombre de contributions qui fluctuent produit inmanquablement (voir paragraphe 5.5) une distribution gaussienne et le phénomène de diffusion.

## 5.4 Irrégularité du mouvement brownien

Pour une particule de masse  $m$  non négligeable, si l'on se place dans le régime stationnaire de l'équation de Langevin (5.1) c'est-à-dire à des temps suffisamment longs pour que la vitesse initiale ait été oubliée, la différence  $x(t) - x(t')$  est une variable gaussienne complètement caractérisée par sa variance. On a

$$P_t(x(t)) = \frac{1}{\sqrt{4\pi \psi(t-t')}} \exp \left[ -\frac{(x(t) - x(t'))^2}{4 \psi(t-t')} \right]$$

où  $2 \psi(t-t') = \langle [x(t) - x(t')]^2 \rangle$  est donné par (5.6).

Lorsque  $\delta t = t - t'$  est grand, on retrouve une constante de diffusion identique à celle d'une particule de masse nulle (5.18). En revanche, pour  $t - t'$  petit, on a  $\psi(t-t') \sim (t-t')^2$ . On voit donc que sur des échelles de temps grandes par rapport à  $m/\gamma$ , les trajectoires d'une particule massive et d'une particule de masse nulle ont des propriétés identiques. Sur des échelles de temps plus courtes que ce temps caractéristique, la trajectoire de la particule massive est régulière avec une vitesse instantanée  $v = (x - x')/(t - t')$  bien définie. En revanche, la trajectoire de la particule de masse nulle décrite par les équations (5.17, 5.18) reste irrégulière sur des échelles de temps aussi courtes que l'on veut : sur un intervalle de temps  $\delta t$  la distance parcourue est typiquement  $|x - x'| \sim \delta t^{1/2}$ , ce qui fait que la vitesse effective mesurée en comparant la position à deux temps séparés de  $\delta t$  diverge comme  $\delta t^{-1/2}$ . La trajectoire de la particule est partout continue mais nulle part dérivable. De là provient la difficulté à mesurer expérimentalement la vitesse d'une particule brownienne.

Une trajectoire brownienne est un exemple typique d'un objet *fractal* : si la durée  $t$  de la marche est multipliée par un facteur  $\lambda$ , la taille de la région d'espace occupée par la marche est multipliée par  $\sqrt{\lambda}$  [12]. Bien que ce soit une trajectoire, le mouvement brownien est un objet de dimension (fractale) 2 : de la même façon qu'un objet compact en dimension  $d$  a ses



dimensions linéaires qui croissent comme son volume à la puissance  $1/d$ , la dimension linéaire d'un mouvement brownien croît comme la puissance  $1/2$  de sa durée  $t$ . L'irrégularité du mouvement brownien se retrouve à toutes les échelles. Par exemple, si l'on suit un mouvement brownien de durée  $t$  dans le plan et si l'on s'intéresse aux intersections de la trajectoire brownienne avec une droite, on peut montrer que chaque point d'intersection est un point d'accumulation d'une infinité d'autres points d'intersection (voir Fig. 5.4).

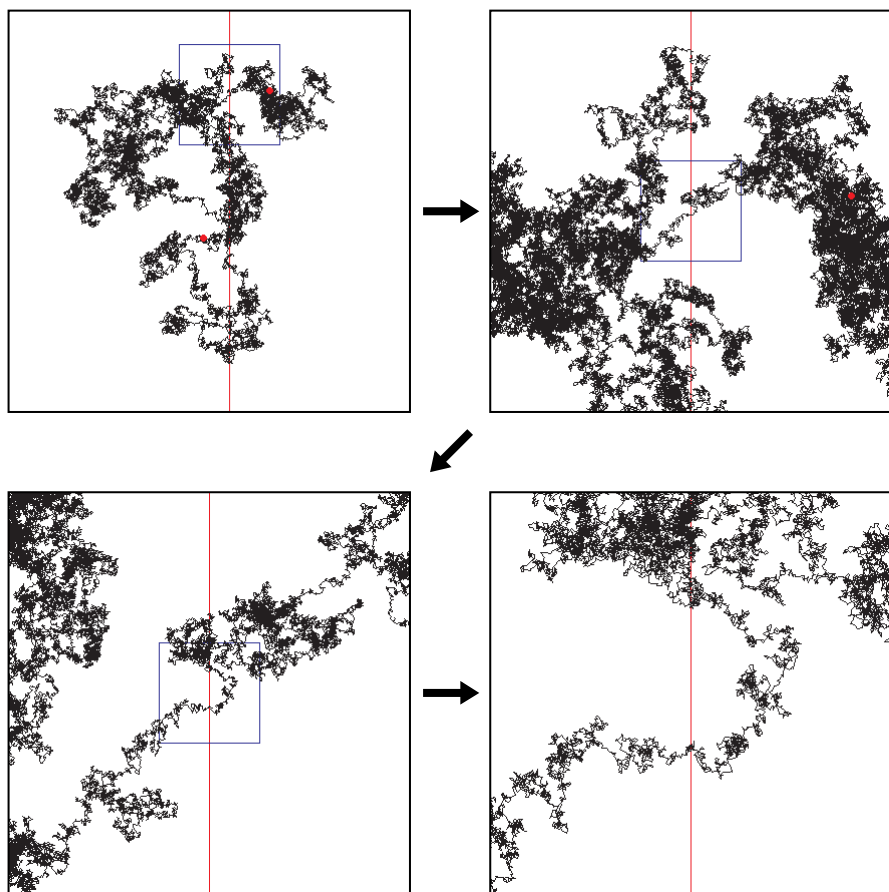


FIG. 5.4 – Les quatre images représentent le même mouvement brownien à plusieurs échelles différentes ; le carré bleu indique la région du zoom et les deux points rouges les extrémités de la ligne. Le mouvement brownien est un fractal dont l'allure générale ne dépend pas de l'échelle. Chaque croisement avec une droite donnée est un point d'accumulation d'une infinité d'autres croisements avec cette droite.

## 5.5 Somme de variables aléatoires

Le calcul des propriétés statistiques des sommes de variables aléatoires est au cœur de la théorie des probabilités et a suscité des travaux de mathématiciens aussi célèbres que De Moivre, Laplace, Bernoulli, Tchebyshev, Markov, Khintchine, Lévy, Kolmogorov. L'équation de diffusion (5.17) et la distribution gaussienne apparaissent dès que l'on s'intéresse à un phénomène qui résulte de l'addition d'un grand nombre de contributions élémentaires aléatoires. Dans le cas du mouvement brownien, le déplacement de la particule est la somme des déplacements élémentaires produits par chaque collision avec les molécules du fluide. L'origine de la loi gaussienne est facile à comprendre si l'on considère la somme  $S$  de  $N$  variables aléatoires  $x_1, x_2, \dots, x_N$  indépendantes et identiquement distribuées selon une loi  $\rho(x)$ .

$$\boxed{S = x_1 + x_2 + \dots + x_N} \quad (5.19)$$

On peut en effet calculer aisément les premiers moments de la somme  $S$

$$\begin{aligned} \langle S \rangle &= N \langle x \rangle \\ \langle S^2 \rangle &= N \langle x^2 \rangle + N(N-1) \langle x \rangle^2 \\ \langle S^3 \rangle &= N \langle x^3 \rangle + 3N(N-1) \langle x^2 \rangle \langle x \rangle + N(N-1)(N-2) \langle x \rangle^3 \\ \langle S^4 \rangle &= N \langle x^4 \rangle + 4N(N-1) \langle x^3 \rangle \langle x \rangle + 3N(N-1) \langle x^2 \rangle^2 \\ &\quad + 6N(N-1)(N-2) \langle x^2 \rangle \langle x \rangle^2 + N(N-1)(N-2)(N-3) \langle x \rangle^4 \end{aligned}$$

et, en définissant une variable réduite  $Y$ ,

$$\boxed{Y = \frac{S - \langle S \rangle}{\sqrt{N}}} \quad (5.20)$$

on obtient, pour  $N$  grand,

$$\begin{aligned} \langle Y \rangle &= 0 \\ \langle Y^2 \rangle &= \langle x^2 \rangle - \langle x \rangle^2 \\ \langle Y^3 \rangle &\rightarrow 0 \\ \langle Y^4 \rangle &\rightarrow 3 \left( \langle x^2 \rangle - \langle x \rangle^2 \right)^2 \end{aligned}$$

qui sont les premiers moments d'une distribution gaussienne de moyenne nulle et de variance

$$\boxed{\Delta = \langle x^2 \rangle - \langle x \rangle^2} \quad (5.21)$$

On pourrait calculer ainsi des moments plus élevés de  $Y$  et vérifier qu'ils tendent tous, pour  $N$  grand, vers ceux de la distribution gaussienne de variance  $\Delta$ .

### 5.5.1 Le théorème de la limite centrale

Une façon simple d'aboutir à cette distribution gaussienne est de calculer directement la distribution de la variable réduite  $Y$  pour  $N$  grand. On a, par définition de  $Y$ ,

$$\text{Prob}(Y) = \int \rho(x_1)dx_1 \cdots \int \rho(x_N)dx_N \delta \left( Y - \frac{1}{\sqrt{N}} \sum_{i=1}^N (x_i - \langle x \rangle) \right)$$

En utilisant la représentation intégrale

$$\delta(y) = \frac{1}{2\pi} \int e^{i\alpha y} d\alpha$$

et le fait que les  $x_i$  sont indépendants, on obtient

$$\text{Prob}(Y) = \frac{1}{2\pi} \int e^{i\alpha Y} \left[ \int \rho(x) e^{-i\alpha \frac{x - \langle x \rangle}{\sqrt{N}}} dx \right]^N d\alpha$$

Pour  $N$  grand

$$\int \rho(x) e^{-i\alpha(x - \langle x \rangle)/\sqrt{N}} dx \simeq 1 - \frac{\alpha^2}{2N} \Delta + O\left(\frac{1}{N^{3/2}}\right) \tag{5.22}$$

ce qui donne

$$\text{Prob}(Y) \rightarrow \frac{1}{2\pi} \int e^{i\alpha Y} \exp\left[-\frac{\alpha^2}{2} \Delta\right] d\alpha$$

c'est-à-dire

$$\boxed{\text{Prob}(Y) \rightarrow \frac{1}{\sqrt{2\pi\Delta}} \exp\left(-\frac{Y^2}{2\Delta}\right)} \tag{5.23}$$

qui est le résultat du théorème de la limite centrale.

En poussant plus loin le développement (5.22), on pourrait calculer les corrections en  $1/N$  de  $\text{Prob}(Y)$  : en particulier, quand  $\langle x^3 \rangle$  est fini, on a génériquement des corrections d'ordre  $N^{-1/2}$  à (5.23).

On peut également noter que, pour arriver à (5.23), on a supposé que les deux premiers moments  $\langle x \rangle$  et  $\langle x^2 \rangle$  de la distribution  $\rho(x)$  des  $x_i$  sont finis. Dans le cas contraire, c'est-à-dire quand la distribution  $\rho(x)$  des variables  $x_i$  décroît trop lentement à l'infini, la distribution de la somme  $S$  n'est plus gaussienne pour  $N$  grand, mais elle est donnée, comme nous le verrons plus loin, par une loi de Lévy.

Quand les variables  $x_i$  sont corrélées, le théorème de la limite centrale (5.19, 5.20, 5.23) reste valable à condition que les corrélations décroissent suffisamment vite. Simplement, l'expression (5.21) de la variance devient

$$\Delta = \sum_{n=-\infty}^{\infty} \langle x_0 x_n \rangle - \langle x \rangle^2$$

qui est une expression du même type que (5.9).

### 5.5.2 Grandes déviations

Le théorème de la limite centrale nous dit que la distribution de la somme  $S$  est gaussienne pour des valeurs de  $S$  proches (d'ordre  $\sqrt{N}$ ) de sa valeur moyenne  $N\langle x \rangle$ . Si l'on s'intéresse aux grandes déviations de la somme, c'est-à-dire si l'on cherche la probabilité que  $S$  prenne une valeur  $Ny$  avec  $y \neq \langle x \rangle$ , la forme de la distribution n'est en général pas gaussienne mais dépend des détails de la distribution  $\rho(x)$ . Elle est de la forme

$$\boxed{\text{Prob} \left( \frac{S}{N} = y \right) \sim e^{Nf(y)}} \quad (5.24)$$

où  $f(y)$  est la fonction de grandes déviations de la somme  $S$ . On peut montrer facilement certaines propriétés de cette fonction de grandes déviations :  $f(y)$  est concave, elle s'annule pour  $y = \langle x \rangle$  et est négative partout ailleurs. La probabilité d'observer une valeur de  $S/N$  différente de  $\langle x \rangle$  décroît donc exponentiellement avec  $N$ . Pour  $y$  proche de  $\langle x \rangle$

$$f(y) \simeq -\frac{(y - \langle x \rangle)^2}{2\Delta}$$

ce qui montre que le théorème de la limite centrale (5.23) n'est que l'approximation parabolique de  $f(y)$  près de son maximum.

On peut déterminer complètement la fonction de grandes déviations  $f(y)$  à partir de  $\rho(x)$ . En effet la probabilité de  $S$  peut s'écrire

$$\text{Prob}(S = yN) = \frac{1}{2\pi} \int d\alpha \exp(N[iy\alpha + g(\alpha)]) \quad (5.25)$$

où  $g(\alpha)$  est

$$g(\alpha) = \log \left[ \int e^{-i\alpha x} \rho(x) dx \right]$$

Pour  $N$  grand, (5.25) peut se calculer par la méthode du col et on obtient la fonction de grandes déviations  $f(y)$  sous forme paramétrique

$$\begin{cases} y = ig'(\alpha), \\ f(y) = g(\alpha) - \alpha g'(\alpha) \end{cases} \quad (5.26)$$

Pour certaines distributions  $\rho(x)$ , on peut avoir une expression plus explicite : par exemple

$$\begin{aligned} \text{pour } \rho(x) &= \frac{1}{2}[\delta(x-1) + \delta(x+1)] & f(y) &= -\frac{1}{2}[(1+y)\log(1+y) \\ & & & + (1-y)\log(1-y)] \\ \text{pour } \rho(x) &= \theta(x) \exp(-x) & f(y) &= -y + 1 + \log y \\ \text{pour } \rho(x) &= \pi^{-1/2} \exp(-x^2) & f(y) &= -y^2 \end{aligned}$$

Nous verrons au paragraphe 5.10 que, pour un fluide, la fonction de grande déviation de la densité est directement reliée à l'énergie libre.

### 5.6 Lois de Lévy

Quand les deux premiers moments de  $x$  sont finis, on a vu que la distribution de la somme (5.19), pour  $N$  grand, est une gaussienne. Quand l'un des deux premiers moments est infini, la distribution de la somme  $S$  devient, pour  $N$  grand, une loi de Lévy [13]. Supposons que la distribution  $\rho(x)$  décroît comme une loi de puissance pour  $x \rightarrow \pm\infty$

$$\begin{aligned} \rho(x) &\simeq \frac{A}{x^{1+\mu}} && \text{pour } x \rightarrow \infty \\ \rho(x) &\simeq \frac{B}{|x|^{1+\mu}} && \text{pour } x \rightarrow -\infty \end{aligned}$$

• Les deux premiers moments de  $x$  et donc de  $S$  sont infinis pour  $0 < \mu \leq 1$ , tandis que pour  $1 < \mu \leq 2$  seul le deuxième moment est infini. Pour obtenir la distribution de  $S$ , on peut répéter la démarche décrite plus haut pour la distribution gaussienne. On écrit

$$\text{Prob}(S) = \frac{1}{2\pi} \int e^{i\alpha S} d\alpha \left[ \int \rho(x) e^{-i\alpha x} dx \right]^N$$

et on utilise le fait que pour  $N$  grand, l'intégrale est dominée par le voisinage de  $\alpha = 0$ .

• Pour  $0 < \mu < 1$ , on peut écrire

$$\int_{-\infty}^{+\infty} e^{-i\alpha x} \rho(x) dx = 1 - \int_{-\infty}^{+\infty} (1 - e^{-i\alpha x}) \rho(x) dx$$

ce qui donne pour  $\alpha$  petit

$$\begin{aligned} \int_{-\infty}^{+\infty} e^{-i\alpha x} \rho(x) dx &\simeq 1 - A|\alpha|^\mu \int_0^\infty \frac{1 - e^{-iy \text{sign}(\alpha)}}{y^{1+\mu}} dy \\ &\quad - B|\alpha|^\mu \int_0^\infty \frac{1 - e^{+iy \text{sign}(\alpha)}}{y^{1+\mu}} dy \end{aligned}$$

En utilisant le fait que  $\int_0^\infty y^{-1-\mu}(1 - e^{-iy}) dy = -\Gamma(-\mu)e^{i\frac{\pi}{2}\mu}$  on obtient ainsi, pour  $N$  grand,

$$\boxed{\text{Prob}(S) = \frac{1}{N^{1/\mu}} L\left(\frac{S}{N^{1/\mu}}\right)} \tag{5.27}$$

où  $L(s)$  est donnée par

$$L(s) = \frac{1}{2\pi} \int_{-\infty}^\infty d\alpha \exp \left[ i\alpha s - |\alpha|^\mu (-\Gamma(-\mu)) (Ae^{i\frac{\pi}{2}\mu \text{sign}(\alpha)} + Be^{-i\frac{\pi}{2}\mu \text{sign}(\alpha)}) \right] \tag{5.28}$$

On voit (5.28) que la distribution de la somme ne dépend que du comportement de  $\rho(x)$  pour  $x$  grand à travers les trois paramètres  $A$ ,  $B$  et  $\mu$ . En fait, on peut éliminer l'un d'entre eux en faisant un changement d'échelle sur  $s$  et l'expression (5.28) constitue donc une famille à deux paramètres de distributions appelées les lois de Lévy.

La forme (5.27) de la distribution de  $S$  montre que pour  $N$  grand, les valeurs typiques de la somme sont de l'ordre de  $N^{1/\mu}$ . On peut calculer facilement, pour  $N$  grand, la distribution  $P(x_{\max})$  de  $x_{\max}$ , le terme le plus grand de la somme (5.19)

$$\begin{aligned} \text{Prob}(x_{\max}) &= N\rho(x_{\max}) \left[ 1 - \int_{x_{\max}}^{\infty} \rho(x) dx \right]^{N-1} \\ &\simeq \frac{NA}{(x_{\max})^{1+\mu}} \exp \left[ - \frac{NA}{\mu(x_{\max})^{\mu}} \right] \end{aligned}$$

On voit que  $P(x_{\max})$  est de la forme  $P(x_{\max}) \sim N^{-1/\mu} H(x_{\max} N^{-1/\mu})$ , ce qui montre que le terme le plus grand de la somme est lui aussi d'ordre  $N^{1/\mu}$  et donc, en comparant à (5.27), du même ordre de grandeur que la somme.

On ne sait pas simplifier l'expression (5.28) sauf dans quelques cas : pour  $\mu = 1/2$  et  $B = 0$ , on peut montrer que

$$L(s) = \frac{A}{s^{3/2}} \exp \left[ - \frac{\pi A^2}{s} \right]$$

tandis que pour  $A = B$  et  $\mu \rightarrow 1$ , on obtient une lorentzienne

$$L(s) = \frac{A}{\pi^2 A^2 + s^2}$$

• Pour  $1 < \mu < 2$  : ce cas peut se traiter de manière identique et cela donne, pour  $N$  grand,

$$\text{Prob}(S) = \frac{1}{N^{1/\mu}} L \left( \frac{S - N\langle x \rangle}{N^{1/\mu}} \right)$$

où  $L(s)$  reste donnée par l'expression (5.28). La valeur typique de la somme  $S$  est proportionnelle à  $N$ , mais les fluctuations, d'ordre  $N^{1/\mu}$  autour de cette valeur moyenne, sont beaucoup plus grandes que celles d'ordre  $N^{1/2}$  prévues lorsque le théorème de la limite centrale s'applique.

## Somme de variables positives

Des distributions qui décroissent en loi de puissance apparaissent fréquemment dans de très nombreux domaines allant de la physique à l'économie [14–16]. C'est le cas, par exemple, de la distribution des énergies en fonction de l'échelle dans les problèmes de turbulence, de celle des amplitudes

des tremblements de terre, des tailles des villes, ou de la répartition des richesses dans une population. Nous avons vu que pour des variables aléatoires dont la distribution décroît, pour  $x$  grand, comme

$$\rho(x) \sim \frac{A}{x^{1+\mu}} \tag{5.29}$$

avec  $0 < \mu < 1$ , le terme le plus grand est du même ordre de grandeur que la somme. On peut montrer en réalité que les quelques termes les plus grands de la somme dominent complètement la somme. Si l'on tire  $N$  variables aléatoires positives avec une distribution  $\rho(x)$  qui décroît comme en (5.29), on peut les ranger par ordre décroissant  $x_1 > x_2 > \dots > x_N$  et appeler  $z_k$  la fraction de la somme que représente la  $k$ -ième plus grande valeur  $x_k$  :

$$z_k = \frac{x_k}{x_1 + x_2 + \dots + x_N}$$

Ces fractions  $z_k$  sont elles aussi des variables aléatoires, puisqu'elles dépendent des  $x_i$  qui varient d'un tirage à l'autre. Pour  $N$  grand, on peut obtenir une expression analytique de  $\langle (z_k)^\nu \rangle$ . En utilisant le fait que  $x^{-\nu}$  peut s'écrire

$$x^{-\nu} = \frac{1}{\Gamma(\nu)} \int_0^\infty t^{\nu-1} e^{-tx} dt$$

on obtient

$$\begin{aligned} \langle (z_k)^\nu \rangle &\simeq \frac{N^k}{(k-1)! \Gamma(\nu)} \int_0^\infty dt t^{\nu-1} \int_0^\infty dx x^\nu \rho(x) e^{-tx} \\ &\quad \times \left[ \int_0^x dy \rho(y) e^{-ty} \right]^{N-k} \left[ \int_x^\infty dy \rho(y) e^{-ty} \right]^{k-1} \end{aligned}$$

ce qui donne pour  $N$  grand, en utilisant (5.29)

$$\begin{aligned} \langle (z_k)^\nu \rangle &\simeq \frac{(NA)^k}{(k-1)! \Gamma(\nu)} \int_0^\infty dt t^{\nu-1} \int_0^\infty x^{\nu-\mu-1} e^{-tx} dx \left[ \int_x^\infty y^{-1-\mu} e^{-ty} dy \right]^{k-1} \\ &\quad \times e^{-NA \left[ \int_0^\infty y^{-\mu-1} (1-e^{-ty}) dy + \int_x^\infty y^{-\mu-1} e^{-ty} dy \right]} \end{aligned}$$

qui peut se simplifier pour donner

$$\langle (z_k)^\nu \rangle \simeq \frac{1}{\mu \Gamma(\nu)} \int_0^\infty dx \frac{x^{\nu-\mu-1} e^{-x} \left[ \int_x^\infty y^{-\mu-1} e^{-y} dy \right]^{k-1}}{\left[ \int_0^\infty y^{-\mu-1} (1-e^{-y}) dy + \int_x^\infty y^{-\mu-1} e^{-y} dy \right]^k}$$

Sur la figure 5.5 sont représentées les valeurs moyennes  $\langle z_1 \rangle$ ,  $\langle z_1 + z_2 \rangle$ ,  $\dots$ ,  $\langle z_1 + \dots + z_5 \rangle$ . On voit que la contribution des quelques termes les plus grands domine complètement la somme. On peut en fait montrer que si l'on approxime la somme par les  $k$  termes les plus grands (avec  $1 \ll k \ll N$ ), l'erreur relative commise décroît comme  $k^{\frac{\mu-1}{\mu}}$ .

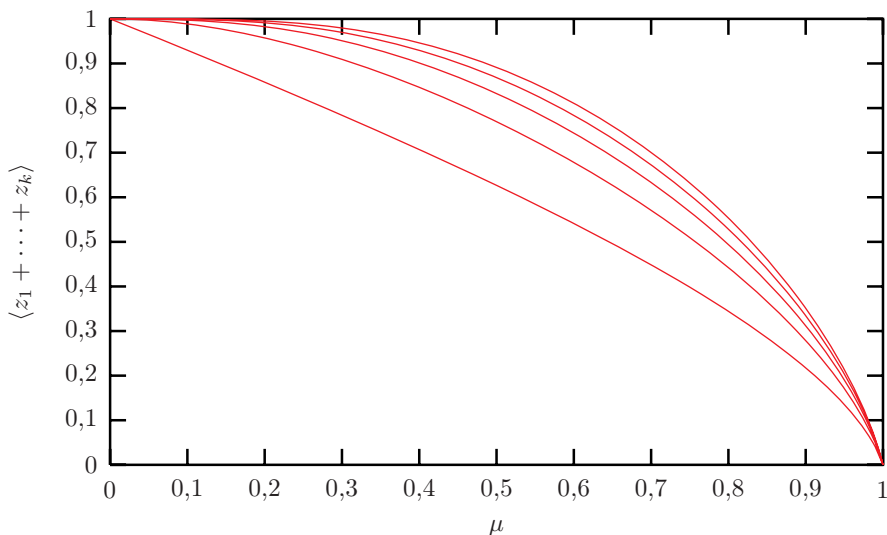


FIG. 5.5 – De bas en haut,  $\langle z_1 \rangle$ ,  $\langle z_1 + z_2 \rangle$ , jusqu'à  $\langle z_1 + z_2 + z_3 + z_4 + z_5 \rangle$  en fonction de  $\mu$ . Pour une somme de  $N$  variables aléatoires dont la distribution décroît comme  $A/x^{\mu+1}$ , c'est la fraction de la somme qui est, en moyenne, due aux  $k$  plus grands éléments. On voit que pour  $\mu < 0,5$ , les cinq termes les plus grands représentent déjà en moyenne 90 % de la somme.

## 5.7 Marches aléatoires

Curieusement, les marches aléatoires sur un réseau, qui constituent des modèles les plus simples du mouvement brownien, furent elles aussi introduites en 1905 par K. Pearson [17]. Elles sont devenues depuis un point de départ obligé pour l'étude des modèles de mécanique statistique ou de théorie des champs sur réseau [18].

### 5.7.1 Marches aléatoires en temps continu sur un réseau régulier

On peut idéaliser le mouvement de la particule brownienne en la plaçant sur un réseau régulier infini (ce qui revient à discrétiser l'espace), et en supposant que chaque collision a pour effet de déplacer la position de la particule d'un nœud du réseau à l'autre. Cela revient à considérer que la particule effectue une marche aléatoire sur le réseau, cette marche étant caractérisée par la probabilité  $\rho(\vec{y})dt$  que la particule saute du site  $\vec{x}$  au site  $\vec{x} + \vec{y}$  pendant l'intervalle de temps infinitésimal  $dt$ . Comme  $dt \ll 1$ , il y a au plus un seul déplacement pendant chaque intervalle de temps  $dt$  et si l'on fait l'hypothèse que les sauts sont non corrélés dans le temps, la probabilité  $P_t(\vec{x})$  de trouver



la particule au point  $\vec{x}$  au temps  $t$  évolue selon l'équation maîtresse

$$\frac{dP_t(\vec{x})}{dt} = \sum_{\vec{y}} \rho(\vec{y}) [P_t(\vec{x} - \vec{y}) - P_t(\vec{x})] \tag{5.30}$$

Il est facile de déduire de cette équation, en multipliant par  $\vec{x}$  et en sommant sur  $\vec{x}$ , l'évolution de la position moyenne  $\langle \vec{x}(t) \rangle$  de la particule :

$$\langle \vec{x}(t) \rangle - \langle \vec{x}(0) \rangle = \vec{V}t$$

où la vitesse  $\vec{V}$  est donnée par

$$\vec{V} = \sum_{\vec{y}} \rho(\vec{y}) \vec{y} \tag{5.31}$$

Si, à l'instant initial, la particule se trouve à l'origine, c'est-à-dire si

$$P_0(\vec{x}) = \delta_{\vec{x}, \vec{0}}$$

on peut résoudre l'équation maîtresse (5.30)

$$P_t(\vec{x}) = \frac{1}{(2\pi)^d} \int_0^{2\pi} \dots \int_0^{2\pi} d\vec{k} e^{-i\vec{k} \cdot \vec{x} - t h(\vec{k})} \tag{5.32}$$

avec

$$h(\vec{k}) = \sum_{\vec{y}} \rho(\vec{y}) (1 - e^{i\vec{k} \cdot \vec{y}})$$

**Remarque :** la marche aléatoire, en temps continu, sur un réseau est un cas particulier de la somme de variables aléatoires. En effet, si l'on découpe le temps en petits intervalles de temps  $\Delta t$ , on peut écrire

$$\vec{x} = \sum_{i=1}^{t/\Delta t} \vec{y}_i$$

où  $\vec{y}_i = \vec{y}$  avec probabilité  $\rho(\vec{y})\Delta t$  et  $\vec{y}_i = \vec{0}$  avec probabilité  $1 - \sum_{\vec{y}} \rho(\vec{y})\Delta t$ . Le comportement de  $P_t(\vec{x})$  aux temps longs (que l'on peut obtenir en évaluant (5.32) par la méthode du col) donne une gaussienne, centrée autour de la position moyenne  $\vec{V}t$  et caractérisée par les deux premiers moments de  $\rho(\vec{y})$ , comme le prévoit le théorème de la limite centrale (voir paragraphe 5.5). La probabilité de trouver le marcheur à une position  $\vec{x} = \vec{W}t$  pour  $\vec{W} \neq \vec{V}$  décroît exponentiellement avec  $t$  comme en (5.24) avec une fonction de grande déviation  $f(\vec{W})$  qui dépend de  $\rho(\vec{y})$ . On obtient en particulier pour la probabilité  $P_t(\vec{0})$  de trouver la particule à l'origine aux temps  $t$

$$\begin{aligned} P_t(\vec{0}) &\sim e^{-t f(\vec{0})} & \text{si } \vec{V} \neq \vec{0} \\ P_t(\vec{0}) &\sim t^{-d/2} & \text{si } \vec{V} = \vec{0} \end{aligned} \tag{5.33}$$

(la puissance  $t^{-d/2}$  provient simplement de la distribution gaussienne de la position : la probabilité qu'une coordonnée s'annule décroît comme  $t^{-1/2}$  aux temps longs et comme pour repasser à l'origine il faut annuler  $d$  coordonnées, on obtient une décroissance en  $t^{-d/2}$ ).

### 5.7.2 Effet de la dimension de l'espace : marche récurrente ou transiente

Certains comportements de la marche aléatoire dépendent qualitativement de la dimension de l'espace dans lequel le réseau se trouve : une marche aléatoire (infiniment longue) revient une infinité de fois à son point départ si sa vitesse moyenne (5.31) est nulle et si la dimension  $d$  de l'espace est inférieure ou égale à 2. En revanche, si  $\vec{V} \neq \vec{0}$  ou si la dimension de l'espace est plus grande que 2, une marche aléatoire a une probabilité non-nulle de ne jamais repasser par l'origine et son nombre moyen de passages est fini.

Quand la particule se trouve en un site  $\vec{x}$  elle a, pendant chaque intervalle de temps  $dt$ , une probabilité  $\sum_{\vec{y}} \rho(\vec{y}) dt$  de sauter vers un autre site. Sa probabilité de ne pas bouger pendant le temps  $t$  est donc  $\exp[-t \sum_{\vec{y}} \rho(\vec{y})]$ , ce qui donne pour le temps moyen d'une visite d'un site quelconque

$$\langle t_{\text{une visite}} \rangle = \left[ \sum_{\vec{y}} \rho(\vec{y}) \right]^{-1}$$

Par ailleurs, la moyenne du temps total passé à l'origine est donnée par

$$\langle t_{\text{total}} \rangle = \int_0^\infty dt P_t(\vec{0}) \quad (5.34)$$

Si on appelle  $N_{\text{retours}}$  le nombre moyen de retours du marcheur à l'origine, on a donc

$$\langle t_{\text{total}} \rangle = (1 + N_{\text{retours}}) \langle t_{\text{une visite}} \rangle$$

(le 1 venant du temps passé avant de quitter l'origine pour la première fois).

À partir de  $N_{\text{retours}}$ , on peut obtenir la probabilité  $R$  que la particule revienne au moins une fois à l'origine. En effet, la probabilité de revenir exactement  $n$  fois à l'origine est  $R^n(1 - R)$ , et donc

$$1 + N_{\text{retours}} = \sum_{n \geq 0} (n + 1) R^n (1 - R) = \frac{1}{1 - R} = \frac{\langle t_{\text{total}} \rangle}{\langle t_{\text{une visite}} \rangle}$$

On voit à partir de (5.33) et (5.34) que :

- si  $\vec{V} = \vec{0}$  et si  $d \leq 2$  : on a  $\langle t_{\text{total}} \rangle = \infty$  et donc  $N_{\text{retours}} = \infty$  et  $R = 1$ . Une marche infiniment longue repasse avec probabilité 1 à l'origine au moins une fois. Si le marcheur repasse une fois, après son premier retour, il repassera avec probabilité 1 encore une deuxième fois, puis une troisième fois et ainsi

de suite. La marche repasse donc une infinité de fois à l'origine. On dit que la marche est *récurrente*.

Comme le marcheur repasse une infinité de fois à l'origine, il a une probabilité 1 de passer au moins une fois par chacun des sites  $\vec{y}$  accessibles en un saut à partir de l'origine. S'il passe au moins une fois sur ces sites  $\vec{y}$  et que la probabilité de retour est 1, il repasse une infinité de fois par chacun de ces sites. En raisonnant ainsi de proche en proche, il est clair que la marche passe une infinité de fois par tous les sites accessibles du réseau ;

- si  $\vec{V} \neq \vec{0}$  ou si  $d > 2$  : on a  $\langle t_{\text{total}} \rangle < \infty$  et donc  $N_{\text{retours}} < \infty$  et  $R < 1$ . Une marche infiniment longue a une probabilité non-nulle  $1 - R$  de ne jamais retourner à l'origine. On dit que la marche est *transiente*.

La dimension 2 est une dimension particulière que l'on appelle *une dimension critique*. La probabilité de retour n'est pas la seule quantité [17, 19] à changer qualitativement quand on traverse la dimension  $d = 2$ . Par exemple le nombre moyen de sites différents visités au bout d'un temps  $t$  augmente linéairement avec  $t$  quand la marche est transiente tandis que ce nombre augmente comme  $\sqrt{t}$  en dimension 1 et comme  $t/\log t$  en dimension 2.

### Exemple

Pour une marche aléatoire à une dimension avec seulement des sauts vers les premiers voisins sur le réseau et  $\rho(1) = p$  et  $\rho(-1) = q$ , la probabilité d'être au point  $x$  au temps  $t$  évolue selon

$$\frac{dP_t(x)}{dt} = pP_t(x - 1) + qP_t(x + 1) - (p + q)P_t(x)$$

On peut alors obtenir la vitesse  $V$ , la probabilité  $R$  de retourner au moins une fois à l'origine et le nombre moyen de retours  $N_{\text{retours}}$  :

$$V = p - q, \quad R = 1 - \frac{|p - q|}{p + q}, \quad N_{\text{retours}} = \frac{p + q}{|p - q|} - 1$$

### 5.7.3 Cas d'une marche aléatoire en temps discret

On considère aussi souvent les marches aléatoires sur un réseau régulier en temps discret. À chaque pas de temps  $\Delta t$ , le marcheur saute d'un site  $\vec{x}$  au site  $\vec{x} + \vec{y}$  avec une probabilité  $p(\vec{y})$  (et reste au même site avec une probabilité  $p(\vec{0})$ ). L'équation maîtresse devient alors

$$P_{t+\Delta t}(\vec{x}) = \sum_{\vec{y}} p(\vec{y})P_t(\vec{x} - \vec{y})$$

On peut répéter tout ce qui précède sur les marches en temps continu. L'expression de  $P_t(\vec{x})$  devient

$$P_t(\vec{x}) = \frac{1}{(2\pi)^d} \int_0^{2\pi} \dots \int_0^{2\pi} d\vec{k} e^{-i \vec{k} \cdot \vec{x}} [h(\vec{k})]^t$$

avec

$$h(\vec{k}) = \sum_{\vec{y}} p(\vec{y}) e^{i \vec{k} \cdot \vec{y}}$$

et toutes les propriétés qualitatives restent les mêmes. En particulier la dimension 2 reste la dimension critique à laquelle la marche passe de récurrente à transiente.

## 5.8 Polymères

L'étude du mouvement brownien et celle des marches aléatoires sont à la base de la théorie des polymères. Les polymères, comme le polystyrène, sont de longues molécules organiques formées, dans les cas les plus simples, d'une succession d'un grand nombre  $N$  (pouvant aller jusqu'à 100 000) de monomères tous identiques. En plus de leur importance industrielle considérable et de tous les travaux de recherche appliquée qu'ils peuvent susciter, les polymères font l'objet de recherches fondamentales à la fois théoriques et expérimentales depuis le milieu des années soixante, en particulier depuis que l'on a découvert le lien qui existe entre la théorie des polymères et celle des transitions de phase du second ordre [20, 21]. Leur étude est également un point de départ obligé pour comprendre les propriétés de macromolécules biologiques, comme l'ADN ou les protéines.

Les chaînes polymériques, en solution, se comportent comme des pelotes dont on peut mesurer la taille par diffusion de la lumière, et l'une des questions les plus simples que l'on peut se poser est de savoir comment cette taille apparente, la taille de la région de l'espace que le polymère occupe, dépend de sa longueur  $N$ . Du point de vue du physicien, la configuration d'un polymère en solution est caractérisée par la position de chacun de ses monomères, c'est-à-dire par  $N$  vecteurs  $\vec{r}_1, \dots, \vec{r}_N$ . Comme la liaison entre monomères successifs est covalente, la distance  $|\vec{r}_{i+1} - \vec{r}_i|$  varie peu. Il existe plusieurs façons de définir la taille apparente de la région occupée par le polymère (qui toutes conduisent aux mêmes comportements qualitatifs). On choisit le plus souvent la distance bout à bout  $|\vec{r}_N - \vec{r}_1|$  entre la position du premier maillon et du  $N$ -ième maillon ou le rayon de giration  $r_{\text{giration}}$  défini par

$$r_{\text{giration}}^2 = \frac{1}{N} \sum_{i=1}^N \left( \vec{r}_i - \frac{1}{N} \sum_{j=1}^N \vec{r}_j \right)^2$$

Le modèle le plus simple de polymère est celui de la chaîne à maillons indépendants. On suppose que les différences  $\vec{r}_{i+1} - \vec{r}_i$  sont des vecteurs de même longueur  $a$  et que leurs directions sont aléatoires et non corrélées. On obtient alors facilement

$$\begin{aligned} \langle (\vec{r}_N - \vec{r}_1)^2 \rangle &= (N-1)a^2 \\ \langle r_{\text{giration}}^2 \rangle &= \frac{N^2-1}{6N} a^2 \end{aligned}$$

ce qui montre que, pour  $N$  grand, la taille typique d'un polymère de  $N$  monomères croît comme  $\sqrt{N}$ . On peut même affirmer que, pour  $N$  grand, la distribution de la distance bout à bout est une gaussienne, puisque c'est la somme de  $N - 1$  vecteurs aléatoires non corrélés. Bien sûr, n'importe quelle modélisation d'un polymère où les directions des éléments successifs le long de la chaîne sont indépendants conduit à

$$|\vec{r}_N - \vec{r}_1| \sim r_{\text{giration}} \sim \sqrt{N}$$

comme le prévoit le théorème de la limite centrale. On s'attend même à ce que des corrélations qui décroissent suffisamment vite ne changent pas cette dépendance en  $\sqrt{N}$  de la taille typique, comme c'était le cas pour le théorème de la limite centrale pour des variables faiblement corrélées.

### 5.8.1 L'effet du volume exclu

Dans tous les modèles simplifiés où les maillons sont indépendants ou peu corrélés, un effet important, l'effet du volume exclu, a été ignoré. Pour un polymère réel en solution, il y a une forte répulsion à courte distance, même entre des monomères très éloignés le long de la chaîne : simplement pour des raisons d'encombrement stérique, la distance  $|\vec{r}_j - \vec{r}_i|$  ne peut pas devenir trop petite même si  $|j - i|$  est grand car la chaîne ne peut pas se recroiser. Le modèle le plus simple qui prenne en compte cette répulsion est *la marche autoévitante* sur un réseau régulier (voir Fig. 5.6). Si  $a$  est le pas du réseau, on modélise un polymère de  $N$  monomères par une marche de  $N - 1$  pas sur ce réseau, avec  $|\vec{r}_{i+1} - \vec{r}_i| = a$ , qui part de l'origine et qui ne repasse jamais deux fois par le même point.

En général, il n'existe pas d'expression fermée du nombre ou de la taille moyenne des marches autoévitantes de  $N$  pas (et donc de  $N + 1$  monomères) sur un réseau donné. On peut néanmoins écrire des programmes, comme cela a été fait dans les années soixante, qui énumèrent toutes les marches autoévitantes jusqu'à une longueur de  $N$  pas (le temps de calcul augmente exponentiellement et les valeurs maximales de  $N$  accessibles n'excèdent pas quelques dizaines). Le résultat de ces énumérations est que

$$|\vec{r}_{N+1} - \vec{r}_1| \sim r_{\text{giration}} \sim N^\nu \quad (5.35)$$

où  $\nu$  un exposant universel dans le sens où il ne dépend que de la dimension  $d$  de l'espace dans lequel le réseau est plongé :  $\nu = 1$  en  $d = 1$ ,  $\nu = 3/4$  en  $d = 2$ ,  $\nu \simeq 0,59$  en  $d = 3$  et  $\nu = 1/2$  en  $d \geq 4$ .

Une façon simple d'estimer l'exposant  $\nu$  est d'énumérer toutes les marches de  $N$  pas, de mesurer le diamètre  $R_N = |\vec{r}_{N+1} - \vec{r}_1|$  des polymères et de calculer  $\langle (R_N)^2 \rangle$  (où la moyenne  $\langle \cdot \rangle$  est réalisée en sommant sur toutes les marches de  $N$  pas). À cause de (5.35), on a

$$\frac{\langle R_{N+1}^2 \rangle}{\langle R_N^2 \rangle} \simeq 1 + \frac{2\nu}{N}$$

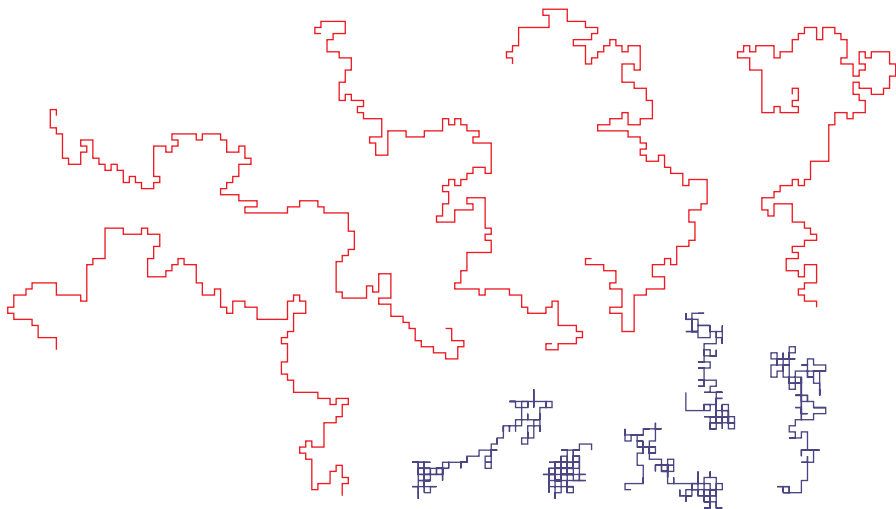


FIG. 5.6 – En rouge, cinq marches autoévitantes ; en bleu cinq marches libres de se croiser. Toutes ces marches font deux cents pas et ont été choisies au hasard de manière uniforme parmi toutes les marches possibles. La distance bout à bout des marches autoévitantes est clairement plus grande que celle des marches sans contrainte.

Cela implique que, pour  $N$  grand,

$$X_N = N \left[ \frac{\langle R_{N+1}^2 \rangle}{\langle R_N^2 \rangle} - 1 \right] \rightarrow 2\nu$$

La figure 5.7 montre cette combinaison en fonction de  $N$ , pour des marches sur un réseau carré, et l'on voit (il y a une convergence alternée) que le résultat est tout à fait compatible avec une valeur asymptotique  $2\nu = 3/2$ .

Déterminer la valeur précise de  $\nu$  en  $d = 3$ , démontrer mathématiquement que  $\nu = 3/4$  en  $d = 2$  ou que  $\nu = 1/2$  au-dessus de la dimension critique 4 font appel aux méthodes les plus puissantes connues à l'heure actuelle en physique théorique ou en théorie des probabilités [19, 22].

### 5.8.2 Le modèle de Domb Joyce

Pour se convaincre que la dimension  $d = 4$  joue un rôle particulier pour les marches autoévitantes, on peut considérer le modèle de Domb-Joyce [23] qui interpole entre la marche aléatoire et la marche autoévitante. Dans ce modèle, les configurations possibles du polymère sont toutes les marches de  $N$  pas, partant de l'origine, sur un réseau régulier de pas  $a$ . Les marches sont libres de se recouper mais on associe une énergie à chaque configuration  $(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N+1})$ , (où  $\vec{r}_1 = \vec{0}$  et où la distance entre monomères successifs est

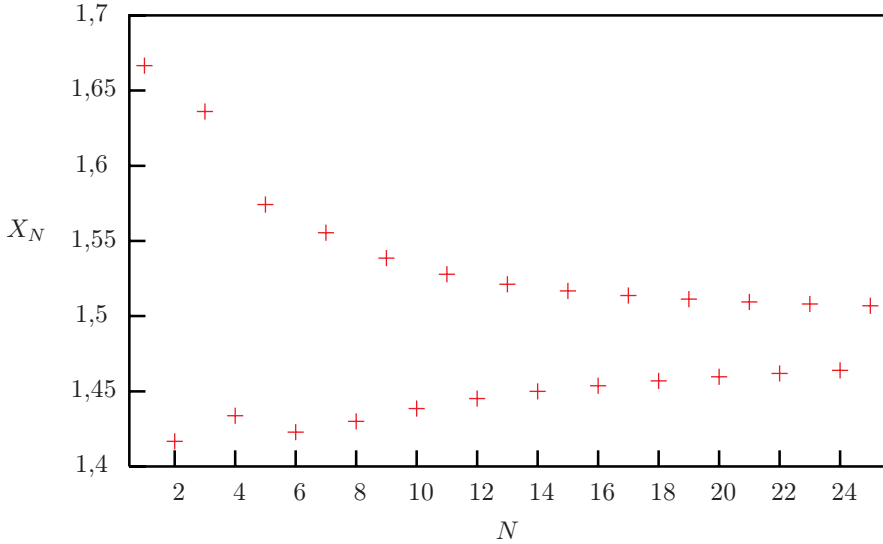


FIG. 5.7 – Le rapport  $X_N$  obtenu en énumérant les marches sans recouvrement sur le réseau carré. (D’après B.D. Hughes [19].)

égale au pas du réseau  $|\vec{r}_i - \vec{r}_{i+1}| = a$ ) :

$$E(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N+1}) = \sum_{1 \leq i < j \leq N+1} \delta_{\vec{r}_i, \vec{r}_j}$$

Chaque fois que la marche se recoupe, il en coûte donc une énergie 1. À la température  $T$ , chaque configuration est occupée avec une probabilité

$$\text{Prob}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N+1}) = Z^{-1} \exp \left[ -\frac{E(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N+1})}{k_B T} \right]$$

où  $Z$  est la constante de normalisation habituelle (la fonction de partition). Dans la limite  $T \rightarrow \infty$ , l’énergie d’interaction ne joue aucun rôle et on a une marche aléatoire normale avec  $\langle (\vec{r}_{N+1} - \vec{r}_1)^2 \rangle = Na^2$ . En revanche, pour  $T \rightarrow 0$ , seules les marches qui ne se recoupent pas ont une probabilité non-nulle.

On peut faire un développement de haute température de ce modèle. En utilisant

$$\langle R_N^2 \rangle = \langle (\vec{r}_{N+1} - \vec{r}_1)^2 \rangle = \frac{\sum (\vec{r}_{N+1} - \vec{r}_1)^2 e^{-E(\vec{r}_1, \dots, \vec{r}_{N+1})/k_B T}}{\sum e^{-E(\vec{r}_1, \dots, \vec{r}_{N+1})/k_B T}}$$

où les sommes portent sur toutes les marches de  $N$  pas, on a, au premier ordre en  $1/T$ ,

$$\langle R_N^2 \rangle = \frac{Na^2 \Omega_N - \frac{1}{k_B T} \sum_{i < j} (N - j + i) a^2 \Omega_i \Omega_{N-j} \Psi_{j-i}}{\Omega_N - \frac{1}{k_B T} \sum_{i < j} \Omega_i \Omega_{N-j} \Psi_{j-i}}$$

où  $\Omega_N$  est le nombre de marches de  $N$  pas sur le réseau et  $\Psi_N$  est le nombre de marches de  $N$  pas sur le réseau qui reviennent à l'origine au bout de ces  $N$  pas. En utilisant le fait que la probabilité  $P_n$  qu'une marche revienne au bout de  $n$  pas à l'origine est donnée par  $P_n = \Psi_n/\Omega_n$ , et que  $\Omega_N = \Omega_i\Omega_{j-i}\Omega_{N-j+i}$ , on obtient

$$\langle R_N^2 \rangle = a^2 \left[ N + \frac{1}{k_B T} \sum_{0 \leq i < j \leq N} (j-i) P_{j-i} + O\left(\frac{1}{(k_B T)^2}\right) \right]$$

ce qui donne

$$\langle R_N^2 \rangle = Na^2 \left[ 1 + \frac{1}{k_B T} \sum_{n=1}^N \frac{n(N+1-n)}{N} P_n + O\left(\frac{1}{(k_B T)^2}\right) \right]$$

Comme on sait (5.33) qu'asymptotiquement  $P_n \simeq Cn^{-d/2}$ , on a deux types de comportements pour  $N$  grand de la correction en  $1/T$  :

- si  $d > 4$ , le terme en  $1/T$  a une limite finie pour  $N$  grand

$$\frac{\langle R_N^2 \rangle_T}{\langle R_N^2 \rangle_\infty} = 1 + \frac{1}{k_B T} \sum_{n=1}^{+\infty} n P_n + O\left(\frac{1}{(k_B T)^2}\right)$$

- si  $d < 4$ , la somme est dominée par les  $n$  grands d'ordre  $N$  et on a

$$\frac{\langle R_N^2 \rangle_T}{\langle R_N^2 \rangle_\infty} = 1 + \frac{C}{k_B T} \frac{4}{(4-d)(6-d)} N^{2-d/2} + O\left(\frac{1}{(k_B T)^2}\right)$$

On voit donc que la théorie de perturbation en  $1/T$  perd son sens pour  $N$  grand au dessous de la dimension critique 4. L'effet a priori faible des interactions (d'ordre  $1/T$ ) donne une correction qui devient de plus en plus grande à mesure que  $N$  augmente. Cela suggère que  $d = 4$  est une dimension critique au dessous de laquelle la théorie de perturbation autour de la chaîne à maillons indépendants est singulière. En fait, le modèle de Domb et Joyce se comporte qualitativement, pour  $N$  grand, à toute température finie comme une marche autoévitante.

## 5.9 Diffusion en présence de désordre

### 5.9.1 Les systèmes désordonnés

Une autre branche importante de la physique statistique qui s'est beaucoup développée depuis les années soixante-dix est la physique des systèmes désordonnés, où l'on s'intéresse à l'effet d'impuretés à l'échelle microscopique sur les propriétés macroscopiques d'un système. Au lieu de supposer un système



constitué d'éléments tous identiques à l'échelle microscopique, on considère une certaine diversité. Par exemple, au lieu d'essayer de décrire les solides comme des alignements d'atomes tous identiques, on s'intéresse à la présence d'impuretés insérées à des positions aléatoires dans ces alignements. On peut également considérer des alliages formés de plusieurs types d'atomes dont les positions sont réparties au hasard à l'intérieur du solide et se demander comment ce désordre de positions à l'échelle microscopique affecte les propriétés macroscopiques de ce matériau.

Parmi les progrès les plus importants réalisés au cours des dernières décennies sur les systèmes désordonnés, on peut citer ceux accomplis sur les verres de spin (des alliages d'atomes magnétiques et non-magnétiques), avec l'émergence de nouvelles idées théoriques [24] qui débordent largement le domaine de la physique pour s'appliquer à des domaines aussi différents que les réseaux de neurones, le repliement des protéines ou l'optimisation. Les expériences sur les verres de spin ont également été à l'origine de l'étude de phénomènes de vieillissement liés à des effets de dynamique lente, qui ont été observés dans bien d'autres contextes [25]. Un autre grand domaine de la théorie des systèmes désordonnés est celui de la localisation qui décrit comment une onde (ou une particule quantique) est diffusée par des impuretés dont les positions sont aléatoires.

Assez souvent, la présence d'une faible concentration d'impuretés ne change pas grand chose aux propriétés d'un matériau. Cela justifie l'étude des systèmes purs (sans impureté). Il y a cependant un certain nombre de cas, bien sûr les plus intéressants, pour lesquels la présence d'un désordre, même très faible, conduit à un comportement complètement différent. Un exemple célèbre est celui de la localisation d'Anderson [26] qui prévoit qu'un potentiel aléatoire, aussi faible soit-il, localise toujours une particule quantique à une dimension. Un autre cas où le désordre change complètement le comportement qualitatif est celui du modèle de Sinai de diffusion d'une particule classique en présence de désordre [27].

### 5.9.2 Le modèle de Sinai

Dans le modèle de Sinai, on considère une marche aléatoire sur un réseau unidimensionnel. À chaque intervalle de temps infinitésimal  $dt$ , un marcheur situé au site  $n$  a une probabilité  $R_n dt$  de sauter au site  $n+1$  et  $L_n dt$  de sauter au site  $n-1$ . L'évolution de la probabilité  $P_n(t)$  de trouver la particule en  $n$  au temps  $t$  est donnée par

$$\frac{dP_n}{dt} = R_{n-1}P_{n-1} + L_{n+1}P_{n+1} - (R_n + L_n)P_n \quad (5.36)$$

Les  $L_n$  et  $R_n$  dépendent de  $n$  et on suppose qu'ils sont choisis au hasard, avec une distribution indépendante de  $n$  non corrélés, et fixés une fois pour toutes. (On peut imaginer, par exemple, que le réseau est un alliage formé de deux types d'atomes  $A, B$  et que le couple  $(L_n, R_n)$  peut prendre deux valeurs

possibles  $(L_A, R_A)$  ou  $(L_B, R_B)$  selon que le site  $n$  représente un atome de type  $A$  ou  $B$ .) Quand

$$\overline{\log R_n} = \overline{\log L_n} \tag{5.37}$$

qui indique que la droite et la gauche sont statistiquement équivalentes (et où la barre représente une moyenne sur les couples  $(L_n, R_n)$ ), le résultat dû à Sinai prévoit une diffusion anormale de la particule aux temps longs

$$\boxed{\langle |n| \rangle \sim \langle n^2 \rangle^{1/2} \sim \log^2 t} \tag{5.38}$$

On peut comprendre (5.38) à partir d'un raisonnement assez simple : pour chaque choix fixé des  $L_n$  et des  $R_n$ , on peut associer à chaque site  $n$  un potentiel  $U_n$  défini par la relation

$$L_{n+1} \exp \left[ -\frac{U_{n+1}}{k_B T} \right] = R_n \exp \left[ -\frac{U_n}{k_B T} \right]$$

Cette relation assure que la dynamique vérifie la propriété de bilan détaillé (voir paragraphe 5.11). Le potentiel  $U_n$  peut alors se calculer par la récurrence

$$U_{n+1} = k_B T (\log L_{n+1} - \log R_n) + U_n$$

Si les  $R_n$  et les  $L_n$  sont aléatoires, indépendants (ou avec des corrélations qui décroissent assez vite) et vérifient la condition (5.37), la différence  $U_n - U_0$  est une somme de variables aléatoires de moyenne nulle et donc

$$|U_n - U_0| \sim k_B T \Delta \sqrt{n}$$

où  $\Delta$  est la constante de diffusion associée à la somme des variables aléatoires  $\log L_{n+1} - \log R_n$ . L'évolution (5.36) décrit donc une particule (classique) qui se déplace dans un potentiel aléatoire. Pour que la particule se trouve à une distance  $n$  de son point de départ 0, il lui faut avoir visité au moins une fois tous les sites entre 0 et  $n$ . Il lui faut donc sauter une barrière de potentiel d'ordre  $\sqrt{n}$ . Si l'on suppose, ce qui donne le bon ordre de grandeur, que la probabilité d'occuper un site déjà visité est donnée par la probabilité d'équilibre  $P_n \sim \exp(-U_n/k_B T)$ , on voit que le temps moyen  $\langle \tau \rangle$  passé sur le site le plus difficile à atteindre entre 0 et  $n$ , celui dont l'énergie potentielle d'ordre  $k_B T \Delta \sqrt{n}$  est la plus haute, est

$$\langle \tau \rangle \sim t e^{-\sqrt{n} \Delta}$$

où  $t$  est le temps total que dure la marche.

Si l'on veut que la particule soit passée réellement au moins une fois par ce sommet le plus haut entre 0 et  $n$  pendant le temps  $t$ , il faut que  $\langle \tau \rangle \sim 1$ , ce qui donne  $\sqrt{n} \sim \log t$  comme le prévoit (5.38).

Ce modèle de Sinai, loin d'être une curiosité mathématique isolée, représente un exemple typique de système possédant une dynamique lente [28, 29]. Plus on attend longtemps, plus le système se trouve face des barrières hautes à franchir. L'évolution se ralentit donc au fur et à mesure que le temps s'écoule, sans jamais trouver une position d'équilibre où s'arrêter.

## 5.10 Théorème de fluctuation-dissipation

Le théorème de fluctuation-dissipation et ses extensions sont au cœur de tous les progrès récents sur les systèmes hors d'équilibre. Au départ, le théorème de fluctuation-dissipation permet de relier, à l'équilibre, la réponse d'un système à une faible force extérieure et les fluctuations de ce système en l'absence de cette force. La relation (5.8) entre la constante de diffusion et la viscosité en est un exemple. La seule propriété utilisée pour le démontrer est que le système est à l'équilibre (et, pour les propriétés dépendant du temps, que la dynamique vérifie la propriété de bilan détaillé).

### 5.10.1 Chaleur spécifique

L'exemple sans doute le plus simple du théorème de fluctuation-dissipation est la relation entre la chaleur spécifique  $C$  d'un système en contact avec un thermostat à une température  $T$  et la variance de l'énergie  $E$  de ce système.

$$\langle E^2 \rangle - \langle E \rangle^2 = k_B T^2 C$$

La chaleur spécifique est en effet un coefficient de réponse qui décrit la variation d'énergie  $\Delta \langle E \rangle = C \Delta T$  due à une petite variation de température  $\Delta T$ .

En utilisant le fait que, pour un système à l'équilibre à la température  $T$ , la probabilité  $P_{\text{eq}}(\mathcal{C}, T)$  de trouver le système dans une configuration microscopique  $\mathcal{C}$  est

$$P_{\text{eq}}(\mathcal{C}, T) = \frac{e^{-\frac{E(\mathcal{C})}{k_B T}}}{\sum_{\mathcal{C}'} e^{-\frac{E(\mathcal{C}')}{k_B T}}} \quad (5.39)$$

on voit facilement que la chaleur spécifique est donnée par

$$C = \frac{d\langle E \rangle}{dT} = \frac{d}{dT} \left[ \frac{\sum_{\mathcal{C}} E(\mathcal{C}) e^{-\frac{E(\mathcal{C})}{k_B T}}}{\sum_{\mathcal{C}} e^{-\frac{E(\mathcal{C})}{k_B T}}} \right] = \frac{1}{k_B T^2} [\langle E^2 \rangle - \langle E \rangle^2]$$

### 5.10.2 Susceptibilité magnétique

On peut considérer également un système magnétique uniaxe comme le modèle d'Ising (dont l'aimantation  $M(\mathcal{C})$  de chaque configuration  $\mathcal{C}$  est un scalaire), pour lequel l'énergie  $E(\mathcal{C})$  d'une configuration est donnée par

$$E(\mathcal{C}) = E_0(\mathcal{C}) - HM(\mathcal{C})$$

où  $E_0(\mathcal{C})$  ne dépend pas du champ magnétique  $H$ . On peut alors montrer que la susceptibilité magnétique  $\chi$  qui décrit la réponse de l'aimantation  $\langle \Delta M \rangle = \chi \Delta H$  à une variation  $\Delta H$  du champ magnétique est reliée à la variance de l'aimantation par

$$\langle M^2 \rangle - \langle M \rangle^2 = k_B T \chi$$

en notant simplement que

$$\chi = \frac{d\langle M \rangle}{dH} = \frac{d}{dH} \left[ \frac{\sum_{\mathcal{C}} M(\mathcal{C}) e^{\frac{HM(\mathcal{C}) - E_0(\mathcal{C})}{k_B T}}}{\sum_{\mathcal{C}} e^{\frac{HM(\mathcal{C}) - E_0(\mathcal{C})}{k_B T}}} \right] = \frac{1}{k_B T} [\langle M^2 \rangle - \langle M \rangle^2]$$

### 5.10.3 Compressibilité d'un fluide

Un troisième exemple est la relation entre les fluctuations de densité d'un fluide et la compressibilité. Si l'on considère un fluide de densité moyenne  $\rho$  de particules dans un volume constant  $V$ , les fluctuations de la densité  $r$  dans une petite région de volume  $v$  sont données par

$$\langle (r - \rho)^2 \rangle = \frac{1}{v} k_B T \rho^2 \kappa(\rho)$$

où la compressibilité  $\kappa(\rho)$  donne la variation de densité  $\Delta\rho$  produite par une petite variation de pression  $\Delta p$

$$\kappa(\rho) = -\frac{1}{V} \frac{dV}{dp} = \frac{1}{\rho} \frac{d\rho}{dp} \quad (5.40)$$

qui devient, en termes du nombre  $n$  de particules présentes dans le volume  $v$ ,

$$\boxed{\langle n^2 \rangle - \langle n \rangle^2 = v k_B T \rho^2 \kappa(\rho)} \quad (5.41)$$

Comme l'ont montré Smoluchowski et Einstein, cette expression permet d'expliquer le phénomène d'opalescence critique qui s'observe quand un fluide est au voisinage d'un point critique. Au voisinage d'un tel point, la compressibilité  $\kappa(\rho)$  devient très grande (elle diverge au point critique). À cause de la relation (5.41), cette grande compressibilité est accompagnée de grandes fluctuations de densité, qui entraînent de grandes fluctuations de l'indice optique et ces grandes variations d'indice augmentent la diffusion du milieu au point de le rendre opaque [1].

Une façon simple d'obtenir la relation (5.41) est de considérer  $N$  particules dans un volume formé de deux régions, l'une de volume  $V - v$  et l'autre de volume  $v$  avec  $v \ll V$ . La probabilité  $\text{Prob}_v(n)$  de trouver  $n$  particules dans le petit volume  $v$  est donnée par

$$\text{Prob}_v(n) = \frac{Z_v(n) Z_{V-v}(N-n)}{Z_V(N)}$$

où  $Z_V(N)$  est la fonction de partition d'un système de  $N$  particules dans un volume  $V$ . Si les volumes  $V$  et  $v$  et les nombres de particules  $N$  et  $n$  sont assez grands (pour qu'on puisse négliger les effets de surface), on a

$$Z_V(N) \sim \exp \left[ -\frac{V}{k_B T} f \left( \frac{N}{V} \right) \right]$$

où  $f(\rho)$  est l'énergie libre par unité de volume à la densité  $\rho$  (dans la limite thermodynamique). Si l'on remplace dans l'expression de  $\text{Prob}_v(n)$ , on voit que

$$\text{Prob}_v(vr) \sim \exp \left[ -v \frac{f(r) - f(\rho) - (r - \rho)f'(\rho)}{k_B T} \right] \quad (5.42)$$

et, pour de faibles variations de densité, on a, au premier ordre, des fluctuations gaussiennes :

$$\text{Prob}_v(vr) \sim \exp \left[ -v \frac{f''(\rho)(r - \rho)^2}{2k_B T} \right] \quad (5.43)$$

En utilisant l'expression usuelle de la pression  $p$

$$p = k_B T \left. \frac{\log Z_V(N)}{dV} \right|_{N \text{ et } T \text{ fixes}} = -f(\rho) + \rho f'(\rho)$$

on voit que

$$\kappa(\rho) = \frac{1}{\rho^2 f''(\rho)}$$

et, avec (5.43), on obtient le résultat annoncé (5.41).

L'approximation gaussienne (5.43) est suffisante pour obtenir (5.41), mais l'expression plus générale (5.42) montre que la fonction de grandes déviations (définie comme en (5.24)) est intimement liée à l'énergie libre. Après tout, si on associe à chaque particule  $i$  une variable binaire  $y_i$ , qui indique si cette particule se trouve dans le volume  $v$  (alors  $y_i = 1$ ) ou pas ( $y_i = 0$ ), le nombre de particules dans le volume  $v$  est la somme de variables aléatoires  $y_i$  corrélées par les interactions.

À l'équilibre, les fluctuations ne sont gaussiennes que si la compressibilité est finie. À une transition de phase, comme par exemple à une transition du 2<sup>e</sup> ordre où  $\kappa(\rho)$  est infinie, les fluctuations de densité ne sont plus gaussiennes. Les variables  $y_i$  deviennent alors trop corrélées pour que le théorème de la limite centrale s'applique. Un autre exemple récent est celui du régime stationnaire de système hors d'équilibre où la présence de corrélations à longue portée induit des fluctuations non gaussiennes de densité [30].

### Exemple du gaz parfait

L'expression (5.41), valable comme nous l'avons vu pour un fluide a priori quelconque (tant que les corrélations entre particules décroissent assez vite avec la distance), peut se vérifier facilement dans le cas d'un gaz parfait. Pour un gaz parfait, la probabilité pour chaque particule de se trouver dans le volume  $v$  est  $v/V$  et comme il n'y a aucune corrélation, on a  $\langle n^2 \rangle - \langle n \rangle^2 = N \frac{v}{V} (1 - \frac{v}{V})$ . Par ailleurs, il est facile de voir à partir de l'équation d'état  $pV = Nk_B T$  et de la définition de la compressibilité (5.40) que  $\kappa(\rho) = \frac{1}{k_B T \rho}$  et donc que (5.41) est vérifiée (pour  $v \ll V$ ).

## 5.11 Processus de Markov et notion de bilan détaillé

Une façon assez simple de décrire un système en contact avec un thermostat est de supposer que le système saute de configuration en configuration en parcourant ainsi tout l'espace des phases et de se donner la probabilité  $W(\mathcal{C}', \mathcal{C})dt$  de sauter pendant chaque intervalle de temps infinitésimal  $dt$  d'une configuration  $\mathcal{C}$  à une configuration  $\mathcal{C}'$ . On dit dans ce cas que l'évolution est un *processus de Markov*, ce qui signifie que la probabilité d'aller de  $\mathcal{C}$  à  $\mathcal{C}'$  entre les instants  $t$  et  $t + dt$  ne dépend que des configurations  $\mathcal{C}$  et  $\mathcal{C}'$  et pas des configurations visitées avant l'instant  $t$ .

La probabilité  $P_t(\mathcal{C})$  de trouver le système dans la configuration  $\mathcal{C}$  au temps  $t$  évolue selon l'équation maîtresse

$$\boxed{\frac{dP_t(\mathcal{C})}{dt} = \sum_{\mathcal{C}' \neq \mathcal{C}} W(\mathcal{C}, \mathcal{C}')P_t(\mathcal{C}') - \sum_{\mathcal{C}' \neq \mathcal{C}} W(\mathcal{C}', \mathcal{C})P_t(\mathcal{C})} \quad (5.44)$$

et, sous une condition assez simple (mathématiquement il faut qu'il existe un entier  $n$  tel que la matrice  $W^n$  ait tous ses coefficients non nuls, ce qui signifie que pendant un temps  $t$  donné, il y a une probabilité non nulle d'atteindre n'importe quelle configuration  $\mathcal{C}'$  à partir d'une configuration  $\mathcal{C}$  quelconque), le système atteint dans la limite des temps longs un régime stationnaire avec une distribution stationnaire  $P_{\text{sta}}(\mathcal{C})$  indépendante de la distribution initiale  $P_0(\mathcal{C})$  et qui vérifie

$$0 = \sum_{\mathcal{C}' \neq \mathcal{C}} W(\mathcal{C}, \mathcal{C}')P_{\text{sta}}(\mathcal{C}') - \sum_{\mathcal{C}' \neq \mathcal{C}} W(\mathcal{C}', \mathcal{C})P_{\text{sta}}(\mathcal{C}) \quad (5.45)$$

Dans ce régime stationnaire, la probabilité d'observer une trajectoire  $(\mathcal{C}_1, \dots, \mathcal{C}_k; t_1, \dots, t_k)$  où le système passe consécutivement un temps  $t_1$  dans la configuration  $\mathcal{C}_1$  puis un temps  $t_2$  dans la configuration  $\mathcal{C}_2$ , etc., est donnée par

$$\text{Prob}(\mathcal{C}_1, \dots, \mathcal{C}_k; t_1, \dots, t_k) = e^{W(\mathcal{C}_k, \mathcal{C}_k)t_k + \dots + W(\mathcal{C}_1, \mathcal{C}_1)t_1} W(\mathcal{C}_k, \mathcal{C}_{k-1}) \dots W(\mathcal{C}_2, \mathcal{C}_1) P_{\text{sta}}(\mathcal{C}_1) \quad (5.46)$$

où l'élément diagonal  $W(\mathcal{C}, \mathcal{C})$  est défini, par convention, par

$$W(\mathcal{C}, \mathcal{C}) = - \sum_{\mathcal{C}' \neq \mathcal{C}} W(\mathcal{C}', \mathcal{C}) \quad (5.47)$$

Les marches aléatoires (paragraphe 5.7) sont des exemples simples de processus de Markov.

### 5.11.1 Validité d’une description markovienne

La description de l’évolution d’un système réel par une dynamique de Markov n’est bien sûr qu’une approximation. Comme l’a écrit N. Van Kampen, la principale difficulté de la mécanique statistique est d’établir (5.44). Pour décrire un système réel et ses interactions avec un thermostat, on doit en principe écrire les équations d’évolution du système et celles du thermostat, en prenant en compte leur couplage. À cause de ce couplage, les trajectoires du système et du thermostat deviennent corrélées au cours du temps. Si la dynamique du système total (composé du système et du thermostat) est suffisamment chaotique (ce qu’on doit forcément supposer pour que le système puisse atteindre l’équilibre en oubliant sa condition initiale) et que l’état initial du système total n’est pas connu avec une précision infinie, on est forcé d’adopter une description probabiliste des trajectoires [31–33]. Une description markovienne revient essentiellement à négliger les corrélations entre l’état du système et celui du thermostat et à supposer que la dynamique du thermostat est suffisamment chaotique pour que l’on puisse négliger ses corrélations temporelles sur les échelles de temps caractéristiques du système que l’on étudie.

Un exemple de dynamique markovienne est celui d’un système de particules (classiques) soumises à des forces aléatoires (comme dans l’équation de Langevin (5.14)) quand les forces aléatoires n’ont pas de corrélations temporelles. L’avantage d’une dynamique markovienne est de permettre d’établir la plupart des résultats avec des outils mathématiques plus simples que ceux que l’on doit mettre en œuvre si l’on veut faire une description plus réaliste du thermostat et de ses interactions avec le système [32, 33].

### 5.11.2 Bilan détaillé et invariance par renversement du temps

Pour qu’un processus de Markov puisse décrire un système à l’équilibre avec une distribution  $P_{\text{eq}}(\mathcal{C})$  comme (5.39), il faut bien sûr que la matrice de Markov  $W(\mathcal{C}, \mathcal{C}')$  soit telle que la solution  $P_{\text{sta}}(\mathcal{C})$  de (5.45) vérifie

$$P_{\text{sta}}(\mathcal{C}) = P_{\text{eq}}(\mathcal{C}) \tag{5.48}$$

On impose en général une condition plus restrictive sur la matrice de Markov, la *condition de bilan détaillé* : pour toute paire de configurations  $\mathcal{C}, \mathcal{C}'$

$$\boxed{W(\mathcal{C}', \mathcal{C})P_{\text{eq}}(\mathcal{C}) = W(\mathcal{C}, \mathcal{C}')P_{\text{eq}}(\mathcal{C}')} \tag{5.49}$$

(ce qui n’empêche pas certains éléments de matrice  $W(\mathcal{C}', \mathcal{C})$  d’être nuls pourvu qu’ils vérifient (5.49)). Cette condition permet de vérifier facilement que  $P_{\text{eq}}(\mathcal{C})$  est bien la distribution stationnaire solution de (5.45). Mais elle donne aussi à la dynamique la propriété très importante d’invariance par renversement du temps (on dit que la dynamique est réversible). Elle exprime le fait qu’à l’équilibre, la probabilité d’observer une trajectoire dans l’espace

des phases est égale à celle d’observer la trajectoire obtenue en renversant le sens du temps : on peut vérifier facilement à partir de (5.46) puis de (5.48) et (5.49) que la trajectoire pour laquelle le système passe consécutivement des périodes de temps  $t_1, t_2, \dots, t_k$  dans les configurations  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k$  et celle où le système passe consécutivement des périodes  $t_k, t_{k-1}, \dots, t_1$  dans les configurations  $\mathcal{C}_k, \mathcal{C}_{k-1}, \dots, \mathcal{C}_1$  sont équiprobables

$$\frac{\text{Prob}(\mathcal{C}_1, \dots, \mathcal{C}_k; t_1, \dots, t_k)}{\text{Prob}(\mathcal{C}_k, \dots, \mathcal{C}_1; t_k, \dots, t_1)} = \frac{W(\mathcal{C}_k, \mathcal{C}_{k-1}) \cdots W(\mathcal{C}_2, \mathcal{C}_1) P_{\text{éq}}(\mathcal{C}_1)}{W(\mathcal{C}_1, \mathcal{C}_2) \cdots W(\mathcal{C}_{k-1}, \mathcal{C}_k) P_{\text{éq}}(\mathcal{C}_k)} = 1$$

Cela signifie que si l’on filme l’évolution temporelle d’un système à l’équilibre, et si l’on passe le film une fois à l’endroit et une fois à l’envers, il n’y a aucun moyen, en regardant les deux versions, de décider celle qui a été réellement filmée.

## 5.12 Système en contact avec un ou plusieurs thermostats

Lorsqu’un système est en contact avec un ou plusieurs thermostats, on peut s’intéresser à ses échanges d’énergie avec l’un de ces derniers. Pour cela, on peut modéliser l’interaction du système avec le thermostat par un processus de Markov caractérisé par une probabilité  $W_q(\mathcal{C}', \mathcal{C})dt$  pour que, pendant un intervalle de temps  $dt$ , un saut de la configuration interne  $\mathcal{C}$  du système vers une configuration  $\mathcal{C}'$  soit accompagné par un transfert  $q$  d’énergie du thermostat vers le système (pour simplifier les notations on suppose que la configuration interne  $\mathcal{C}$  change chaque fois qu’une énergie  $q$  est transférée; on peut également noter que  $q$  peut être aussi bien positif que négatif). La probabilité  $P_t(\mathcal{C}, Q)$  de trouver le système dans la configuration  $\mathcal{C}$ , sachant que l’énergie  $Q$  a été transférée du thermostat vers le système entre le temps 0 et le temps  $t$ , évolue alors selon

$$\frac{dP_t(\mathcal{C}, Q)}{dt} = \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}, \mathcal{C}') P_t(\mathcal{C}', Q - q) - \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}', \mathcal{C}) P_t(\mathcal{C}, Q) \tag{5.50}$$

Cette description, plus fine que celle donnée par (5.44), se ramène à (5.44) si l’on ne s’intéresse qu’aux configurations internes du système et si l’on choisit  $W(\mathcal{C}, \mathcal{C}')$  tel que

$$W(\mathcal{C}, \mathcal{C}') = \sum_q W_q(\mathcal{C}, \mathcal{C}')$$

L’équation d’évolution (5.50) permet a priori de décrire complètement la statistique des échanges d’énergie entre le thermostat et le système. Dans la limite des temps longs, le calcul de la distribution de  $Q$  peut se ramener à chercher la plus grande valeur propre  $\mu(\lambda)$  d’une matrice. En effet, si on définit



la fonction génératrice  $\mathcal{P}_t(\mathcal{C}, \lambda)$  par

$$\mathcal{P}_t(\mathcal{C}, \lambda) = \sum_Q P_t(\mathcal{C}, Q) e^{\lambda Q}$$

on voit, à partir de (5.50), que

$$\frac{d\mathcal{P}_t(\mathcal{C}, \lambda)}{dt} = \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} e^{\lambda q} W_q(\mathcal{C}, \mathcal{C}') \mathcal{P}_t(\mathcal{C}', \lambda) - \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}', \mathcal{C}) \mathcal{P}_t(\mathcal{C}, \lambda) \tag{5.51}$$

et comme

$$\langle e^{\lambda Q} \rangle = \sum_{\mathcal{C}} \sum_Q P_t(\mathcal{C}, Q) e^{\lambda Q} = \sum_{\mathcal{C}} \mathcal{P}_t(\mathcal{C}, \lambda)$$

on obtient, pour les temps longs,

$$\boxed{\langle e^{\lambda Q} \rangle \sim e^{t \mu(\lambda)}} \tag{5.52}$$

où  $\mu(\lambda)$  est la plus grande valeur propre de la matrice  $\mathcal{W}_\lambda$  qui apparaît dans le membre de droite de (5.51) et dont les éléments de matrice sont donnés par

$$\mathcal{W}_\lambda(\mathcal{C}, \mathcal{C}') = \sum_q e^{\lambda q} W_q(\mathcal{C}, \mathcal{C}') - \delta_{\mathcal{C}, \mathcal{C}'} \sum_q \sum_{\mathcal{C}'' \neq \mathcal{C}} W_q(\mathcal{C}'', \mathcal{C}) \tag{5.53}$$

Pour trouver  $\mu(\lambda)$ , il faut donc chercher le vecteur propre  $p(\mathcal{C}, \lambda)$  qui vérifie

$$\mu(\lambda) p(\mathcal{C}, \lambda) = \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} e^{\lambda q} W_q(\mathcal{C}, \mathcal{C}') p(\mathcal{C}', \lambda) - \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}', \mathcal{C}) p(\mathcal{C}, \lambda) \tag{5.54}$$

La connaissance de cette valeur propre  $\mu(\lambda)$  permet alors de calculer dans la limite des temps longs toute la distribution de  $Q$ . En particulier on voit facilement à partir de (5.52) que les cumulants de  $Q$  sont donnés par les dérivées successives de  $\mu(\lambda)$

$$\begin{aligned} \frac{\langle Q \rangle}{t} &= \left. \frac{d\mu}{d\lambda} \right|_{\lambda=0} \\ \frac{\langle Q^2 \rangle_c}{t} &= \frac{\langle Q^2 \rangle - \langle Q \rangle^2}{t} = \left. \frac{d^2\mu}{d\lambda^2} \right|_{\lambda=0} \\ \frac{\langle Q^3 \rangle_c}{t} &= \frac{\langle Q^3 \rangle - 3\langle Q^2 \rangle \langle Q \rangle + 2\langle Q \rangle^3}{t} = \left. \frac{d^3\mu}{d\lambda^3} \right|_{\lambda=0} \end{aligned} \tag{5.55}$$

et ainsi de suite. On peut noter que l'on a toujours

$$\mu(0) = 0 \tag{5.56}$$

car pour  $\lambda = 0$ , on a  $p(\mathcal{C}, \lambda) = P_{\text{sta}}(\mathcal{C})$  et (5.54) se ramène à (5.45).

### 5.12.1 Condition de bilan détaillé

Pour un système en contact avec un ou plusieurs thermostats une trajectoire est spécifiée par la suite des configurations microscopiques  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k$  visitées par le système, par les temps  $t_1, t_2, \dots, t_k$  passés dans chacune de ces configurations et par les transferts d'énergie  $q_1, \dots, q_{k-1}$  du thermostat vers le système qui accompagnent les sauts d'une configuration à la configuration suivante. La probabilité d'une telle trajectoire est donnée, dans le régime stationnaire, par

$$\text{Prob}(\mathcal{C}_1, \dots, \mathcal{C}_k; t_1, \dots, t_k; q_1, \dots, q_{k-1}) = e^{W(\mathcal{C}_k, \mathcal{C}_k)t_k + \dots + W(\mathcal{C}_1, \mathcal{C}_1)t_1} \\ \times W_{q_{k-1}}(\mathcal{C}_k, \mathcal{C}_{k-1}) \cdots W_{q_1}(\mathcal{C}_2, \mathcal{C}_1) P_{\text{sta}}(\mathcal{C}_1)$$

où l'élément diagonal est défini comme en (5.47)

$$W(\mathcal{C}, \mathcal{C}) = - \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}', \mathcal{C})$$

Pour un système à l'équilibre, c'est-à-dire si chacune des configurations microscopiques  $\mathcal{C}$  est occupée avec la probabilité  $P_{\text{éq}}(\mathcal{C})$  donnée par (5.39), la condition de bilan détaillé (5.49) devient, quand on prend en compte les échanges avec le thermostat

$$\boxed{W_q(\mathcal{C}', \mathcal{C}) P_{\text{éq}}(\mathcal{C}) = W_{-q}(\mathcal{C}, \mathcal{C}') P_{\text{éq}}(\mathcal{C}')} \quad (5.57)$$

Cette condition assure en effet que la probabilité d'une trajectoire est égale à celle de la trajectoire obtenue en renversant le sens du temps,

$$\text{Prob}(\mathcal{C}_1, \dots, \mathcal{C}_k; t_1, \dots, t_k; q_1, \dots, q_{k-1}) = \\ \text{Prob}(\mathcal{C}_k, \dots, \mathcal{C}_1; t_k, \dots, t_1; -q_{k-1}, \dots, -q_1)$$

(quand on remonte le temps, il faut bien sûr changer le signe des échanges d'énergie avec le thermostat).

### 5.12.2 Système en contact avec un seul thermostat

Le cas le plus simple est celui d'un système en contact avec un seul thermostat à la température  $T$ . La conservation de l'énergie impose que les seules transitions  $W_q(\mathcal{C}', \mathcal{C})$  possibles sont celles pour lesquelles

$$E(\mathcal{C}') = E(\mathcal{C}) + q \quad (5.58)$$

Il est alors facile de vérifier que

$$p(\mathcal{C}, \lambda) = e^{\lambda E(\mathcal{C})} P_{\text{éq}}(\mathcal{C})$$

est un vecteur propre de (5.54) (on peut même montrer qu'il s'agit bien du vecteur correspondant à la plus grande valeur propre en utilisant le théorème de Perron-Frobenius) et que la valeur propre associée est

$$\mu(\lambda) = 0$$

ce qui montre (5.55) que tous les cumulants (une fois divisés par  $t$ ) de l'énergie totale  $Q$  transférée du thermostat vers le système s'annulent dans la limite des temps longs. Cela est bien sûr une évidence : l'énergie  $Q$  transférée du thermostat vers le système ne peut pas augmenter indéfiniment.

### 5.12.3 Système en contact avec plusieurs thermostats : équilibre

On considère un système à l'équilibre avec deux (ou plusieurs) thermostats à la même température, et l'on s'intéresse aux échanges avec l'un de ces thermostats.  $W_q(C', C)dt$  représente alors la probabilité que le saut d'une configuration  $C$  vers une configuration  $C'$  soit accompagné d'un transfert d'énergie  $q$  du thermostat qui nous intéresse vers le système, mais la relation (5.58) n'est en général plus vérifiée car un changement de l'énergie interne  $E(C)$  peut être dû à un échange d'énergie avec l'un quelconque des autres thermostats.

Si la relation de bilan détaillé (5.57) est satisfaite, on peut voir facilement que si  $p(C, \lambda)$  est un vecteur propre à droite (5.54) de la matrice  $\mathcal{W}_\lambda$  donnée en (5.53), il vérifie également

$$\mu(\lambda)p(C, \lambda) = \sum_q \sum_{C' \neq C} e^{\lambda q} W_{-q}(C', C) \frac{P_{\text{eq}}(C)}{P_{\text{eq}}(C')} p(C', \lambda) - \sum_q \sum_{C' \neq C} W_q(C', C) p(C, \lambda)$$

ce qui montre que le vecteur  $p(C, \lambda)/P_{\text{eq}}(C)$  est un vecteur propre à gauche de  $\mathcal{W}_{-\lambda}$  avec la valeur propre  $\mu(\lambda)$ . On voit donc que  $\mu(\lambda)$  est à la fois valeur propre de  $\mathcal{W}_\lambda$  et de  $\mathcal{W}_{-\lambda}$  ce qui entraîne que

$$\mu(\lambda) = \mu(-\lambda) \tag{5.59}$$

Cette relation exprime évidemment la réversibilité du système : sur une longue période de temps il est aussi probable d'observer un transfert  $Q$  qu'un transfert  $-Q$  du thermostat vers le système, ce qui fait que

$$e^{\mu(\lambda)t} \sim \langle e^{\lambda Q} \rangle = \langle e^{-\lambda Q} \rangle \sim e^{\mu(-\lambda)t}$$

D'après (5.59) et (5.55),  $\mu'(0) = 0$  et l'énergie  $Q$  transférée par le thermostat vers le système est bien sûr nulle en moyenne (puisque le système est en équilibre avec le thermostat), mais cette énergie fluctue et ses fluctuations sont caractérisées par  $\mu(\lambda)$  défini en (5.52).

**Remarque :** quand la condition de bilan détaillé (5.57) n'est pas vérifiée par la distribution stationnaire  $P_{\text{sta}}(C)$ , le flux d'énergie moyen  $\langle Q \rangle/t$  n'est plus

forcément nul : on peut avoir un flux d'énergie du thermostat qui nous intéresse vers le système et cette énergie est évacuée grâce aux contacts avec les autres thermostats. Cette situation se produit pour les systèmes dissipatifs, comme par exemple des systèmes mécaniques soumis à des forces de frottement. Les forces de frottement ou la viscosité jouent alors le rôle du contact avec les autres thermostats qui évacuent l'énergie que l'on injecte en appliquant des forces en général à grande échelle.

### 5.12.4 Système en contact avec deux thermostats : régime stationnaire hors d'équilibre

Considérons, pour simplifier, une situation où le thermostat qui nous intéresse est à la température  $T_1$ , tous les autres thermostats étant à la température  $T$ . Si  $T_1 > T$ , un régime stationnaire s'établit avec un flux moyen d'énergie non nul du thermostat vers le système.

On suppose que la matrice  $W_q(\mathcal{C}', \mathcal{C})$  qui caractérise les échanges entre le thermostat qui nous intéresse et le système dépend des températures  $T_1$  et  $T$  et vérifie la condition de bilan détaillé modifiée :

$$\boxed{e^{q\left(\frac{1}{k_B T_1} - \frac{1}{k_B T}\right)} W_q(\mathcal{C}, \mathcal{C}') P_{\text{éq}}(\mathcal{C}') = W_{-q}(\mathcal{C}', \mathcal{C}) P_{\text{éq}}(\mathcal{C})} \quad (5.60)$$

où  $P_{\text{éq}}(\mathcal{C})$  est la distribution d'équilibre (5.39) à la température  $T$ . Ce choix est naturel car si les contacts avec tous les autres thermostats étaient coupés, c'est-à-dire si  $q = E(\mathcal{C}) - E(\mathcal{C}')$ , la relation (5.60) serait bien la condition de bilan détaillé (5.57) pour une distribution d'équilibre à la température  $T_1$ .

Pour une telle dynamique, si l'on cherche  $\mu(\lambda)$  en essayant de résoudre le problème aux valeurs propres (5.54), comme on l'a fait dans le cas précédent, on constate que  $p(\mathcal{C}, \lambda)$ , solution de (5.54), vérifie également

$$\begin{aligned} \mu(\lambda) p(\mathcal{C}, \lambda) &= \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} e^{q\left(\lambda + \frac{1}{k_B T} - \frac{1}{k_B T_1}\right)} W_{-q}(\mathcal{C}', \mathcal{C}) \frac{P_{\text{éq}}(\mathcal{C})}{P_{\text{éq}}(\mathcal{C}')} p(\mathcal{C}', \lambda) \\ &\quad - \sum_q \sum_{\mathcal{C}' \neq \mathcal{C}} W_q(\mathcal{C}', \mathcal{C}) p(\mathcal{C}, \lambda) \end{aligned}$$

ce qui (en faisant le changement de variable  $q \rightarrow -q$  dans le premier terme) montre que  $p(\mathcal{C}, \lambda) / P_{\text{éq}}(\mathcal{C})$  est vecteur propre à gauche de la matrice  $W_{\frac{1}{k_B T} - \frac{1}{k_B T_1} - \lambda}$  avec la valeur propre  $\mu(\lambda)$ . On a donc

$$\boxed{\mu(\lambda) = \mu\left(\frac{1}{k_B T_1} - \frac{1}{k_B T} - \lambda\right)} \quad (5.61)$$

Cette relation est une symétrie des systèmes hors d'équilibre, découverte numériquement en 1993 [34] et établie par Gallavotti et Cohen en 1995 [35].

### 5.12.5 Relation de Gallavotti-Cohen

Si l'on s'intéresse aux échanges d'énergie entre le thermostat (à la température  $T_1$ ) et le système pendant une longue période  $t$ , on s'attend à observer un flux moyen d'énergie  $\bar{q}$

$$\frac{\langle Q \rangle}{t} = \bar{q}$$

Dans ce régime stationnaire, l'énergie moyenne du système reste constante et toute l'énergie transférée du thermostat à la température  $T_1$  finit par être évacuée par les autres thermostats. Ce flux d'énergie peut s'interpréter comme un taux  $\partial_t S$  de création d'entropie (c'est la somme des entropies des thermostats qui augmente)

$$\partial_t S = \bar{q} \left( \frac{1}{T} - \frac{1}{T_1} \right)$$

La probabilité d'observer, pendant une longue période  $t$ , un transfert d'énergie  $Q = qt$  avec  $q \neq \bar{q}$  décroît exponentiellement avec le temps  $t$  :

$$\text{Prob} \left( \frac{Q}{t} = q \right) \sim e^{tF(q)} \tag{5.62}$$

où la fonction de grandes déviations  $F(q)$  peut être déterminée à partir de la connaissance de  $\mu(\lambda)$ . En effet,

$$e^{\mu(\lambda)t} \sim \langle e^{\lambda Q} \rangle = \int dq e^{\lambda qt} \text{Prob} \left( \frac{Q}{t} = q \right) \sim e^{t \max_q (F(q) + \lambda q)}$$

Cela donne  $F(q)$  à partir de la connaissance de  $\mu(\lambda)$  sous forme paramétrique :

$$q = \mu'(\lambda) \quad \text{et} \quad F(q) = \mu(\lambda) - \lambda \mu'(\lambda)$$

La relation de Gallavotti-Cohen (5.61) s'écrit alors

$$F(q) - F(-q) = q \left( \frac{1}{k_B T} - \frac{1}{k_B T_1} \right) \tag{5.63}$$

On voit donc, comme l'ont établi Gallavotti et Cohen dans le cadre de la théorie des systèmes dynamiques, que la différence  $F(q) - F(-q)$  est linéaire en  $q$ , ce qui n'était certainement pas une évidence puisque  $F(q)$  est en principe une fonction compliquée de  $q$ .

La relation de Gallavotti-Cohen (5.63) est a priori difficile à observer car elle relie la fonction de grande déviation  $F(q)$  pour deux valeurs  $q$  et  $-q$ , dont au moins l'une correspond à des événements rares : la plupart des événements que l'on peut observer dans une expérience ou dans une simulation numérique donnent des valeurs de  $Q/t$  proches de  $\bar{q}$ , la valeur pour laquelle  $F(\bar{q}) = 0$ . Mesurer  $F(q)$  pour d'autres valeurs de  $q \neq \bar{q}$  nécessite d'observer le système pendant un temps  $t$  suffisamment long pour que (5.62) soit valable et de répéter la mesure un nombre de fois supérieur à  $e^{-tF(q)}$ .

## 5.12.6 Validité du second principe de la thermodynamique

Le fait que  $F(-q)$  est en général finie (et non  $-\infty$ ) montre que le second principe de la thermodynamique (qui postule que l'énergie est cédée par la source chaude à la source froide) n'est lui aussi qu'une loi macroscopique valide aux fluctuations près.

Si  $T_1 > T$ , la relation de Gallavotti-Cohen nous dit que  $F(q) > F(-q)$  pour  $q > 0$ , ce qui signifie que le plus probable est de voir le thermostat à la température  $T_1$  céder de l'énergie au système (qui lui-même cède cette énergie aux autres thermostats). Il y a néanmoins une probabilité non nulle que l'échange se fasse en sens inverse, c'est-à-dire que  $q$  soit négatif.

Sur de courtes périodes de temps et pour des systèmes suffisamment petits (pour lesquels les fluctuations sont les plus faciles à observer), il n'est donc pas impossible de voir l'énergie aller dans le sens opposé à celui prédit par le second principe de la thermodynamique. La probabilité d'observer un événement qui viole le second principe pendant un temps d'observation  $t$  décroît néanmoins exponentiellement avec  $t$ . Elle est donc d'autant plus faible que ce temps est long.

## 5.12.7 Théorème de fluctuation

La relation de Gallavotti-Cohen généralise des relations bien connues pour les systèmes proches de l'équilibre. Si l'on considère à la fois que  $\lambda$  et  $T_1 - T$  sont petits, le développement de  $\mu(\lambda)$  au 2<sup>e</sup> ordre se réduit à

$$\mu(\lambda) \simeq A(T)(T_1 - T)\lambda + B(T)\lambda^2$$

(On utilise le fait (5.56, 5.59) que l'on a toujours  $\mu(0) = 0$  et que  $\mu(\lambda)$  est paire pour  $T_1 = T$ .) La symétrie de Gallavotti-Cohen (5.61) donne une relation entre les deux coefficients  $A(T)$  et  $B(T)$  qui s'écrit

$$B(T) = k_B T^2 A(T)$$

En utilisant (5.55), on voit que cette égalité est encore une fois une relation entre un coefficient de réponse, la conductivité  $A(T)$  qui relie le flux thermique à un petit gradient de température

$$\frac{\langle Q \rangle}{t} = A(T)(T_1 - T) + O[(T_1 - T)^2]$$

et  $B(T)$ , qui représente les fluctuations de l'énergie échangée avec le thermostat quand  $T_1 = T$  :

$$\frac{\langle Q^2 \rangle_c}{t} = 2B(T) = 2k_B T^2 A(T)$$

### 5.12.8 Exemple : une particule brownienne couplée à deux thermostats

Dans certains cas simples [36], comme dans l'exemple qui suit, on peut arriver à calculer complètement la fonction  $\mu(\lambda)$ . Considérons le cas d'une particule décrite par l'équation de Langevin (5.1) et couplée à deux thermostats, l'un à la température  $T$  et l'autre à la température  $T_1$ . L'évolution de la vitesse est donnée par

$$m \frac{dv}{dt} = -(\gamma + \gamma_1)v + L(t) + L_1(t)$$

où  $\gamma$  et  $L(t)$  représentent l'effet sur la particule du thermostat à la température  $T$ , et  $\gamma_1$  et  $L_1(t)$  celui du thermostat à la température  $T_1$ . Les forces  $L(t)$  et  $L_1(t)$  sont deux forces gaussiennes, indépendantes et de moyennes nulles, caractérisées par leurs variances :

$$\langle L(t)L(t') \rangle = \Gamma \delta(t - t'), \quad \langle L_1(t)L_1(t') \rangle = \Gamma_1 \delta(t - t') \quad \text{et} \quad \langle L(t)L_1(t') \rangle = 0$$

avec

$$\Gamma = 2\gamma k_B T \quad \text{et} \quad \Gamma_1 = 2\gamma_1 k_B T_1$$

On s'intéresse à l'énergie  $Q$  transférée du thermostat à la température  $T_1$  vers le système pendant le temps  $t$ . Cette énergie est un travail : c'est l'intégrale sur le temps du produit de la vitesse  $v(t)$  par la force  $L_1(t) - \gamma_1 v(t)$  (exercée par le thermostat sur la particule)

$$Q = \int_0^t d\tau v(\tau)[L_1(\tau) - \gamma_1 v(\tau)]$$

Pour éviter les difficultés techniques de calcul liées aux corrélations de type  $\delta(t - t')$  des forces aléatoires, il est commode de découper le temps en petits intervalles  $\Delta t$  au cours desquels les forces aléatoires sont constantes : ces forces sont définies comme la moyenne sur l'intervalle de temps  $\Delta t$  des forces  $L(t)$  et  $L_1(t)$

$$F = \frac{1}{\Delta t} \int_t^{t+\Delta t} L(\tau) d\tau \quad \text{et} \quad F_1 = \frac{1}{\Delta t} \int_t^{t+\Delta t} L_1(\tau) d\tau$$

Ces deux forces ne sont pas corrélées entre elles, elles sont indépendantes d'un intervalle de temps à l'autre, gaussiennes et de moyenne nulle. Elles sont donc complètement caractérisées par leurs variances

$$\langle F^2 \rangle = \frac{\Gamma}{\Delta t}, \quad \langle F_1^2 \rangle = \frac{\Gamma_1}{\Delta t} \quad \text{et} \quad \langle FF_1 \rangle = 0 \tag{5.64}$$

Si l'intervalle de temps  $\Delta t$  est suffisamment court, la variation de vitesse et la variation d'énergie transférée du thermostat à la température  $T_1$  vers le

système peuvent s'écrire

$$v(t + \Delta t) - v(t) = \frac{F + F_1}{m} \Delta t - \frac{\gamma + \gamma_1}{m} v(t) \Delta t$$

$$Q(t + \Delta t) - Q(t) = \int_0^{\Delta t} d\tau v(t + \tau) (-\gamma_1 v(t + \tau) + F_1)$$

Pour  $\Delta t$  petit, cela donne

$$Q(t + \Delta t) - Q(t) = F_1 v(t) \Delta t - \gamma_1 v(t)^2 \Delta t + \frac{F F_1 + F_1^2}{2m} (\Delta t)^2$$

Il peut paraître surprenant de conserver pour  $\Delta t$  petit des termes d'ordre  $F^2(\Delta t)^2$ . C'est cependant ce qu'il faut faire, comme c'était déjà le cas lors de la dérivation de l'équation de Fokker-Planck. Si l'on regarde l'ordre de grandeur des forces  $F$  et  $F_1$  (5.64), on voit que  $F$  et  $F_1$  sont d'ordre  $(\Delta t)^{-1/2}$  et donc le terme en  $F^2(\Delta t)^2$  est du même ordre de grandeur que le premier terme. On peut par contre négliger les termes en  $\gamma^2(\Delta t)^2$ .

Si on introduit une fonction génératrice

$$G_t(\alpha, \lambda) = \langle e^{\alpha v^2(t) + \lambda Q(t)} \rangle$$

on voit que

$$G_{t+\Delta t}(\alpha, \lambda) = \left\langle e^{\alpha v^2(t) + \lambda Q(t)} \left[ 1 + v \Delta t \left( 2\alpha \frac{F + F_1}{m} + \lambda F_1 \right) + v^2 \Delta t \left( -2\alpha \frac{\gamma + \gamma_1}{m} - \lambda \gamma_1 \right) + (\Delta t)^2 \left( \alpha \frac{(F + F_1)^2}{m^2} + \lambda \frac{(F + F_1) F_1}{2m} \right) + v^2 (\Delta t)^2 \left( 2\alpha^2 \frac{(F + F_1)^2}{m^2} + \frac{\lambda^2}{2} F_1^2 + 2\lambda \alpha \frac{F_1 (F + F_1)}{m} \right) \right] \right\rangle$$

Après moyennage sur  $F_1$  et  $F$ , on obtient

$$G_{t+\Delta t}(\alpha, \lambda) = \left\langle e^{\alpha v^2(t) + \lambda Q(t)} \left[ 1 + \Delta t \left( \alpha \frac{\Gamma + \Gamma_1}{m^2} + \lambda \frac{\Gamma_1}{2m} \right) + v^2 \Delta t \left( -2\alpha \frac{\gamma + \gamma_1}{m} - \lambda \gamma_1 + 2\alpha^2 \frac{\Gamma + \Gamma_1}{m^2} + \frac{\lambda^2}{2} \Gamma_1 + 2\lambda \alpha \frac{\Gamma_1}{m} \right) \right] \right\rangle \quad (5.65)$$

Si l'on choisit pour  $\alpha$  la plus petite solution de

$$-2\alpha \frac{\gamma + \gamma_1}{m} - \lambda \gamma_1 + 2\alpha^2 \frac{\Gamma + \Gamma_1}{m^2} + \frac{\lambda^2}{2} \Gamma_1 + 2\lambda \alpha \frac{\Gamma_1}{m} = 0 \quad (5.66)$$



on annule le terme en  $v^2$  dans l'équation (5.65) et on obtient

$$G_{t+\Delta t}(\alpha, \lambda) = G_t(\alpha, \lambda) \left[ 1 + \Delta t \left( \alpha \frac{\Gamma + \Gamma_1}{m^2} + \lambda \frac{\Gamma_1}{2m} \right) \right]$$

Pour des temps longs,

$$G_t(\alpha, \lambda) \sim \langle e^{\lambda Q(t)} \rangle \sim e^{\mu(\lambda)t} \quad \text{pour } t \text{ grand ;} \tag{5.67}$$

ce qui donne

$$\begin{aligned} \mu(\lambda) &= \alpha \frac{\Gamma + \Gamma_1}{m^2} + \lambda \frac{\Gamma_1}{2m} \\ &= \frac{\gamma + \gamma_1 - \sqrt{\gamma^2 + \gamma_1^2 + 2\gamma\gamma_1(1 - 2\lambda k_B T_1 + 2\lambda k_B T - 2\lambda^2 k_B^2 T T_1)}}{2m} \end{aligned}$$

(Si l'on choisit pour  $\alpha$  l'autre solution de (5.66), on obtient une expression différente de  $\mu(\lambda)$  qui ne peut pas être correcte parce qu'elle ne vérifie pas  $\mu(0) = 0$ . En fait, pour cette autre solution,  $\alpha$  est suffisamment grand pour que  $\langle \exp(\alpha v^2) \rangle$  et  $G_t(\alpha, \lambda)$  soient infinis, et on ne peut alors pas utiliser (5.67).)

L'expression de  $\mu(\lambda)$  obtenue vérifie la relation de Gallavotti-Cohen (5.61). En utilisant (5.55), on obtient

$$\frac{\langle Q \rangle}{t} = \frac{\gamma\gamma_1 k_B (T_1 - T)}{m(\gamma + \gamma_1)} \quad \text{et} \quad \frac{\langle Q^2 \rangle_c}{t} = \frac{2k_B^2 T T_1 \gamma\gamma_1}{m(\gamma + \gamma_1)} + \frac{2k_B^2 (T - T_1)^2 \gamma^2 \gamma_1^2}{m(\gamma + \gamma_1)^3}$$

Quand  $T_1 \neq T$ , il y a un flux moyen d'énergie de la source chaude vers la source froide. Ce flux d'énergie fluctue et sa distribution peut être obtenue à partir de  $\mu(\lambda)$ . En revanche, quand les deux thermostats sont à la même température ( $T = T_1$ ), ce flux moyen d'énergie est nul, mais la variance des fluctuations de l'énergie échangée augmente linéairement avec le temps. On a alors, comme prévu,  $\mu(\lambda) = \mu(-\lambda)$ . Enfin, si l'on coupe le couplage avec l'un des deux thermostats, c'est-à-dire si  $\gamma = 0$  ou si  $\gamma_1 = 0$ , le système se retrouve en contact avec un seul thermostat et l'on a bien  $\mu(\lambda) = 0$ .

### 5.13 Système en contact avec des réservoirs de particules

Toute la discussion qui précède peut se répéter, à l'identique, si l'on remplace les thermostats (qui sont des réservoirs d'énergie) par des réservoirs de particules. Pour un système en contact avec un seul réservoir ayant un potentiel chimique  $\nu$ , la distribution d'équilibre vérifie

$$P_{\text{eq}}(C) = e^{-\frac{E(C) - \nu N(C)}{k_B T}} \bigg/ \sum_{C'} e^{-\frac{E(C') - \nu N(C')}{k_B T}}$$

où  $N(\mathcal{C})$  est le nombre de particules de la configuration  $\mathcal{C}$ . Si un système est en contact avec plusieurs réservoirs, l'un au potentiel chimique  $\nu_1$  et les autres au potentiel  $\nu$ , on peut s'intéresser au nombre de particules  $Q(t)$  transférées du réservoir au potentiel  $\nu_1$  vers le système. On suppose que la dynamique satisfait les conditions de bilan détaillé (voir (5.49), (5.57) et (5.60)) avec (5.60) remplacé par

$$\boxed{e^{\frac{q(\nu-\nu_1)}{k_B T}} W_q(\mathcal{C}, \mathcal{C}') P_{\text{éq}}(\mathcal{C}') = W_{-q}(\mathcal{C}', \mathcal{C}) P_{\text{éq}}(\mathcal{C})} \quad (5.68)$$

où  $q$  est le nombre de particules transférées du réservoir vers le système quand le système saute de la configuration  $\mathcal{C}'$  à la configuration  $\mathcal{C}$ .

Toutes les relations établies dans le cadre d'échanges d'énergie avec des thermostats restent valables dans le cas présent : si l'on définit  $\mu(\lambda)$  comme précédemment à partir de la fonction génératrice du nombre de particules transférées du réservoir vers le système

$$\langle e^{\lambda Q(t)} \rangle \sim e^{t \mu(\lambda)} \quad \text{pour } t \text{ grand}, \quad (5.69)$$

la relation de Gallavotti-Cohen devient

$$\mu(\lambda) = \mu \left( \frac{\nu - \nu_1}{k_B T} - \lambda \right) \quad (5.70)$$

Un cas particulier important est celui de particules chargées : si  $W_q(\mathcal{C}, \mathcal{C}')$  représente une évolution du système accompagnée du transfert d'une charge  $q$  du réservoir vers le système, les potentiels chimiques s'identifient aux potentiels électrostatiques des deux réservoirs. (On peut remarquer que, par sa définition (5.69) même, si dans (5.70) le paramètre  $\lambda$  n'a pas de dimension, il a la dimension de l'inverse d'une énergie dans (5.61).)

### 5.13.1 Bruit Schottky

L'exemple sans doute le plus simple pour lequel on sait calculer les fluctuations de courant dans un système hors d'équilibre est celui du bruit Schottky, le bruit du courant des électrons émis par la cathode d'un tube à vide. On peut modéliser la charge  $Q(t)$  transférée pendant le temps  $t$  en supposant que, pendant chaque intervalle de temps infinitésimal  $dt$ , une particule de charge  $q$  a une probabilité  $\alpha dt$  de quitter la cathode pour rejoindre l'anode. La charge  $Q(t)$  reçue par l'anode pendant le temps  $t$  évolue donc de la manière suivante :

$$Q(t + dt) = \begin{cases} Q(t) & \text{avec probabilité } 1 - \alpha dt \\ Q(t) + q & \text{avec probabilité } \alpha dt \end{cases}$$

ce qui donne, pour les moments de  $Q(t)$ ,

$$\langle Q^k(t + dt) \rangle = (1 - \alpha dt) \langle Q^k(t) \rangle + \alpha dt \langle (Q(t) + q)^k \rangle \quad (5.71)$$

En particulier,

$$\begin{aligned} \langle Q(t) \rangle &= \alpha q t \\ \langle Q^2(t) \rangle - \langle Q(t) \rangle^2 &= \alpha q^2 t \end{aligned}$$

et on obtient que

$$\boxed{\frac{\langle Q^2(t) \rangle_c}{\langle Q(t) \rangle} = q} \tag{5.72}$$

qui est indépendant de  $t$  et  $\alpha$ . Une mesure du courant moyen et de ses fluctuations (bruit de grenaille) donne donc la charge élémentaire des particules qui participent au transport du courant.

À partir de (5.71), on voit que la fonction génératrice de  $Q$  vérifie

$$\frac{d\langle e^{\lambda Q(t)} \rangle}{dt} = \alpha (e^{\lambda q} - 1) \langle e^{\lambda Q(t)} \rangle$$

qui, aux temps longs, a la forme attendue

$$\langle e^{\lambda Q(t)} \rangle \sim e^{t\mu(\lambda)} \quad \text{avec} \quad \mu(\lambda) = \alpha (e^{\lambda q} - 1)$$

Lorsque l'on s'intéresse aux fluctuations de courant à travers un matériau et que l'on prend en compte les interactions entre électrons, leur caractère fermionique et les collisions avec les impuretés, on observe des écarts à la relation (5.72). Un résultat remarquable sur les conducteurs mésoscopiques à température nulle est que la relation (5.72) devient

$$\frac{\langle Q^2 \rangle_c}{\langle Q \rangle} = \frac{q}{3}$$

indépendamment du détail des interactions des électrons avec les impuretés [37,38]. D'une manière surprenante, ce même facteur 1/3 se retrouve pour des particules classiques qui diffusent sur un réseau entre un réservoir plein et un réservoir vide, avec des interactions de cœur dur [39].

### 5.13.2 Relations de réciprocité d'Onsager

Quand un système est en contact avec plusieurs réservoirs, on peut s'intéresser aux corrélations d'énergies ou de particules échangées avec chacun de ces réservoirs. Par exemple, on peut essayer de calculer les corrélations entre le courant d'énergie et celui de particules qui traversent un système.

Pour se limiter à un cas précis, imaginons un système en contact avec plusieurs réservoirs de particules muni d'une dynamique markovienne. La probabilité de sauter, pendant l'intervalle de temps  $dt$ , d'une configuration  $\mathcal{C}$  vers une configuration  $\mathcal{C}'$ , avec un transfert de  $q_1$  particules du réservoir 1 vers le système et de  $q_2$  particules du réservoir 2 vers le système est donnée par

$W_{q_1, q_2}(\mathcal{C}', \mathcal{C})dt$ . Si on appelle  $Q_1$  et  $Q_2$  les charges totales transférées des réservoirs 1 et 2 vers le système pendant le temps  $t$ , on peut montrer que la fonction génératrice de  $Q_1$  et  $Q_2$  vérifie, aux temps longs,

$$\langle e^{\lambda_1 Q_1 + \lambda_2 Q_2} \rangle \sim e^{\mu(\lambda_1, \lambda_2)t}$$

Si on introduit des potentiels chimiques  $\nu_1$  associé au premier réservoir,  $\nu_2$  au second et  $\nu$  à tous les autres, la relation (5.68) doit être remplacée par

$$e^{\frac{q_1(\nu - \nu_1) + q_2(\nu - \nu_2)}{k_B T}} W_{q_1, q_2}(\mathcal{C}, \mathcal{C}') P_{\text{éq}}(\mathcal{C}') = W_{-q_1, -q_2}(\mathcal{C}', \mathcal{C}) P_{\text{éq}}(\mathcal{C})$$

où  $q_1$  et  $q_2$  sont les incréments de  $Q_1$  et  $Q_2$  quand le système saute de la configuration  $\mathcal{C}'$  à la configuration  $\mathcal{C}$ . En repassant exactement par les mêmes étapes que précédemment, on peut montrer que

$$\mu(\lambda_1, \lambda_2) = \mu \left( \frac{\nu - \nu_1}{k_B T} - \lambda_1, \frac{\nu - \nu_2}{k_B T} - \lambda_2 \right) \tag{5.73}$$

qui généralise la relation de Gallavotti-Cohen à plusieurs courants. Cette symétrie impose des relations entre les corrélations de  $Q_1$  et  $Q_2$ . En particulier, en développant (5.73) au second ordre dans les  $\nu_i - \nu$  et les  $\lambda_i$ , on trouve que

$$\begin{aligned} \frac{\langle Q_1 \rangle}{t} &= M_{11}(\nu_1 - \nu) + M_{12}(\nu_2 - \nu) \\ \frac{\langle Q_2 \rangle}{t} &= M_{21}(\nu_1 - \nu) + M_{22}(\nu_2 - \nu) \end{aligned}$$

où la matrice  $M_{ij} = \frac{\partial^2 \mu}{\partial \lambda_i \partial \nu_j}$  est une matrice symétrique

$$\boxed{M_{ij} = M_{ji}}$$

Cette symétrie constitue la relation de réciprocité d’Onsager [40]. Elle exprime que la réponse (linéaire) de la variable  $Q_i$  à la différence de potentiel chimique  $\nu_j - \nu$  est la même, à l’équilibre, que la réponse de  $Q_j$  à la différence  $\nu_i - \nu$ . Les relations d’Onsager sont valables que  $Q_1$  et  $Q_2$  soient des quantités de même nature ou pas. Par exemple,  $Q_1$  peut être un flux d’énergie et  $Q_2$  un courant de particules.

## 5.14 Théorème de fluctuation-dissipation dynamique

Il existe une version dynamique du théorème de fluctuation dont l’origine est une autre très grande contribution d’Onsager [41].

Considérons un système proche de l’équilibre que l’on soumet à un champ extérieur  $h_t$  couplé à une observable  $B(\mathcal{C})$  qui dépend de la configuration  $\mathcal{C}$

du système (par exemple, le champ  $h_t$  pourrait être un champ magnétique dépendant du temps, et l'observable  $B(\mathcal{C})$  serait alors l'aimantation). Dans le cadre d'une dynamique de Markov, cela signifie que la probabilité  $w_t(\mathcal{C}', \mathcal{C})dt$  de passer de  $\mathcal{C}$  à  $\mathcal{C}'$  pendant l'intervalle de temps  $t, t + dt$  dépend maintenant du temps à travers le champ  $h_t$ . La probabilité  $p_t(\mathcal{C})$  de trouver le système dans une configuration  $\mathcal{C}$  au temps  $t$  évolue alors selon l'équation maîtresse

$$\frac{dp_t(\mathcal{C})}{dt} = \sum_{\mathcal{C}' \neq \mathcal{C}} w_t(\mathcal{C}, \mathcal{C}')p_t(\mathcal{C}') - \sum_{\mathcal{C}' \neq \mathcal{C}} w_t(\mathcal{C}', \mathcal{C})p_t(\mathcal{C}) \quad (5.74)$$

et l'on suppose que les  $w_t(\mathcal{C}', \mathcal{C})$  vérifient une condition de bilan détaillé modifiée :

$$\boxed{e^{\frac{h_t}{k_B T} [B(\mathcal{C}) - B(\mathcal{C}')]} w_t(\mathcal{C}', \mathcal{C}) P_{\text{éq}}(\mathcal{C}) = w_t(\mathcal{C}, \mathcal{C}') P_{\text{éq}}(\mathcal{C}')} \quad (5.75)$$

où  $P_{\text{éq}}(\mathcal{C})$  est la distribution d'équilibre (5.39) à la température  $T$ , en l'absence de champ.

Toute la discussion qui suit est fondée sur cette relation (5.75) qui peut se justifier en considérant le cas particulier d'un champ  $h_t$  constant. En effet, si  $h_t$  est constant, la distribution stationnaire de (5.74) est donnée par  $P_{\text{éq}}(\mathcal{C}) \exp [h_t B(\mathcal{C}) / (k_B T)]$  et (5.75) n'est rien d'autre que la condition de bilan détaillé (5.49) pour cette nouvelle distribution d'équilibre.

Si le champ  $h_t$  est petit, on suppose que  $w_t(\mathcal{C}', \mathcal{C})$  est proche du taux de transition  $W(\mathcal{C}', \mathcal{C})$  qu'on avait pour  $h_t = 0$

$$w_t(\mathcal{C}', \mathcal{C}) = W(\mathcal{C}', \mathcal{C}) + O(h_t)$$

On obtient, en utilisant les relations de bilan détaillé (5.49, 5.75), l'évolution de la différence

$$Q_t(\mathcal{C}) = p_t(\mathcal{C}) - P_{\text{éq}}(\mathcal{C})$$

au premier ordre dans les  $h_t$  :

$$\begin{aligned} \frac{dQ_t(\mathcal{C})}{dt} = \sum_{\mathcal{C}' \neq \mathcal{C}} \left[ W(\mathcal{C}, \mathcal{C}') Q_t(\mathcal{C}') - W(\mathcal{C}', \mathcal{C}) Q_t(\mathcal{C}) \right. \\ \left. + \frac{h_t}{k_B T} W(\mathcal{C}, \mathcal{C}') [B(\mathcal{C}) - B(\mathcal{C}')] P_{\text{éq}}(\mathcal{C}') \right] \end{aligned}$$

On peut alors intégrer, au moins formellement, et l'on obtient, en utilisant (5.47) et (5.49),

$$Q_t(\mathcal{C}) = -\frac{1}{k_B T} \sum_{\mathcal{C}', \mathcal{C}''} \int_{-\infty}^t d\tau h_\tau M_{t-\tau}(\mathcal{C}, \mathcal{C}') W(\mathcal{C}', \mathcal{C}'') B(\mathcal{C}'') P_{\text{éq}}(\mathcal{C}'')$$

où

$$M_t = \exp(tW)$$

Si l'on mesure une grandeur  $A$ , on obtient donc, au temps  $t$ ,

$$\langle A(\mathcal{C}_t) \rangle = \sum_{\mathcal{C}} A(\mathcal{C}) [P_{\text{éq}}(\mathcal{C}) + Q_t(\mathcal{C})] = \langle A \rangle_{\text{éq}} + \int_{-\infty}^t d\tau R_{A,B}(t, \tau) h_{\tau}$$

où la fonction  $R_{A,B}(t, t')$  est la réponse de l'observable  $A$  au temps  $t$  à un champ  $h_{t'}$  couplé à l'observable  $B$  à un temps antérieur  $t'$  :

$$R_{A,B}(t, t') = -\frac{1}{k_B T} \sum_{\mathcal{C}, \mathcal{C}', \mathcal{C}''} A(\mathcal{C}) M_{t-t'}(\mathcal{C}, \mathcal{C}') W(\mathcal{C}', \mathcal{C}'') B(\mathcal{C}'') P_{\text{éq}}(\mathcal{C}'')$$

Si l'on compare cette expression à la fonction de corrélation à l'équilibre

$$C_{A,B}(t, t') = \langle A(\mathcal{C}_t) B(\mathcal{C}_{t'}) \rangle = \sum_{\mathcal{C}, \mathcal{C}'} A(\mathcal{C}) M_{t-t'}(\mathcal{C}, \mathcal{C}') B(\mathcal{C}') P_{\text{éq}}(\mathcal{C}')$$

on voit que

$$\boxed{R_{A,B}(t, t') = \frac{1}{k_B T} \frac{d}{dt'} C_{A,B}(t, t')} \quad (5.76)$$

Ceci est la relation de fluctuation-dissipation dynamique. Pour établir cette relation, on a seulement utilisé que la distribution  $p_t(\mathcal{C})$  reste toujours proche de la distribution d'équilibre  $P_{\text{éq}}(\mathcal{C})$  en faisant un développement du premier ordre en  $h_t$ . C'est ce qu'on appelle *une théorie de la réponse linéaire*. On verra que pour des systèmes loin de l'équilibre, cette relation cesse d'être vraie.

Une situation typique, dans des expériences est de brancher (ou de couper) le champ  $h_t$  à partir d'un temps  $t'$  ; si l'on suppose  $h_t = 0$  pour  $t < t'$  et  $h_t$  constant non nul pour  $t > t'$ , la réponse  $\chi_{A,B}$  à ce champ constant est alors

$$\chi_{A,B}(t, t') = \int_{t'}^t R_{A,B}(t, \tau) d\tau \quad (5.77)$$

ce qui donne

$$\boxed{\chi_{A,B}(t, t') = \frac{1}{k_B T} (C_{A,B}(t, t) - C_{A,B}(t, t'))} \quad (5.78)$$

On peut noter que, dans le cadre de la théorie de la réponse linéaire, les fonctions  $C_{A,B}(t, t')$ ,  $R_{A,B}(t, t')$  et  $\chi_{A,B}(t, t')$  ne dépendent que de la différence  $t - t'$ .

## Exemple de l'équation de Langevin

On peut illustrer la relation (5.76) par l'exemple simple de l'équation de Langevin (5.1). Si l'on choisit pour  $A$  la vitesse  $V(t)$  de la particule et pour  $B$

sa position  $x(t)$ , on voit (5.3, 5.5) que, dans le régime stationnaire, (c'est-à-dire pour  $t$  et  $t'$  grands,  $t > t'$  avec  $t - t'$  d'ordre 1)

$$\langle V(t) x(t') \rangle = \frac{1}{m\gamma} \int_{-\infty}^t d\tau \int_{-\infty}^{t'} d\tau' e^{-\gamma(t-\tau)/m} \left(1 - e^{-\gamma(t'-\tau')/m}\right) \langle L(\tau)L(\tau') \rangle$$

ce qui donne, pour une force aléatoire décorrélée,

$$C_{A,B}(t, t') = \langle V(t) x(t') \rangle = \frac{k_B T}{\gamma} e^{-\gamma(t-t')/m}$$

Le déplacement  $x(t)$  est la grandeur conjuguée à une force extérieure  $F$  car le produit de la force par le déplacement est bien une énergie, comme l'était le produit  $h_t B(\mathcal{C})$  dans (5.75). Si l'on soumet la particule à une force extérieure  $F(t)$  qui dépend du temps,

$$m \frac{dV}{dt} = -\gamma V + L(t) + F(t)$$

la réponse de la vitesse moyenne  $\langle V(t) \rangle$  au temps  $t$  à cette force est de la forme

$$\langle V(t) \rangle = \frac{1}{m} \int_{-\infty}^t d\tau e^{-\gamma(t-\tau)/m} F(\tau)$$

ce qui signifie que

$$R_{A,B}(t, t') = \frac{\delta V(t)}{\delta L(t')} = \frac{1}{m} e^{-\gamma(t-t')/m}$$

et la relation de fluctuation-dissipation (5.76) est bien vérifiée.

## 5.15 Systèmes loin de l'équilibre

Quand un système est maintenu dans un régime stationnaire hors de l'équilibre (par exemple par des contacts avec des thermostats à des températures différentes) ou quand la condition initiale fait que, sur l'échelle de temps d'observation, le système reste loin de l'équilibre, les relations de fluctuation-dissipation (5.76, 5.78) cessent d'être vérifiées.

### 5.15.1 Le vieillissement

Loin de l'équilibre, les fonctions de corrélation  $C_{A,B}(t, t')$  comme les fonctions de réponse  $R_{A,B}(t, t')$  ou  $\chi_{A,B}(t, t')$  ne dépendent plus seulement de la différence  $t - t'$

$$C_{A,B}(t, t') \neq C_{A,B}(t - t', 0)$$

mais en général des deux temps  $t$  et  $t'$ . Comme la réponse  $R_{A,B}(t, t')$  au temps  $t$  à un champ exercé au temps  $t'$  dépend à la fois de la différence  $t - t'$

et de l'âge  $t'$  (temps depuis lequel on a fixé les paramètres externes comme la température), on dit que le système a vieilli pendant le temps  $t'$ .

Un exemple où ce phénomène de vieillissement est bien visible est celui de l'aimantation thermorémanente d'un verre de spin (comme sur la figure 5.8 dans le cas d'un alliage AgMn). L'expérience réalisée est (conceptuellement) très simple : on refroidit l'échantillon au temps  $t = 0$  en présence d'un faible champ magnétique, on attend un certain temps  $t'$  (ce temps est en général appelé un temps d'attente) et l'on coupe le champ magnétique. L'aimantation se met à décroître et l'on mesure l'aimantation en fonction de la différence  $t - t'$ . On voit que la courbe d'aimantation thermorémanente dépend du temps d'attente, la relaxation étant d'autant plus lente que  $t'$  est long.

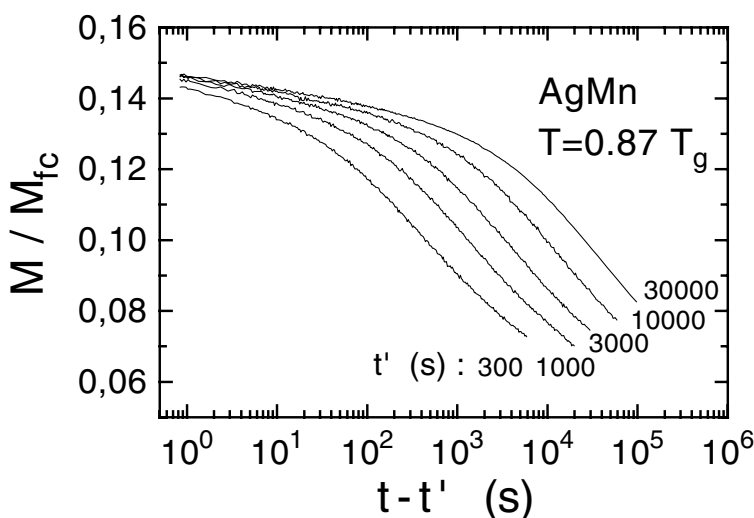


FIG. 5.8 – Aimantation thermorémanente d'un alliage AgMn. À  $t = 0$ , le système est plongé dans un champ magnétique et il est refroidi à une température  $T = 0,87 T_g \approx 9 \text{ K}$  inférieure à la température de transition vitreuse  $T_g$ . À cette température et en présence de ce champ, son aimantation d'équilibre est  $M_{fc}$ . À l'instant  $t'$ , le champ magnétique est coupé ; le système perd 85 % de son aimantation pendant la première seconde, puis il entre dans un régime de relaxation lente où son aimantation dépend de  $t'$ . (Figure empruntée à Eric Vincent et al. [42].)

Cela peut se comprendre qualitativement en pensant à la dynamique comme à une évolution dans un espace de phases dont le paysage d'énergie comporte de nombreuses vallées. Plus le temps d'attente  $t'$  est long, plus le système a de chances de tomber dans une vallée d'énergie (ou plutôt d'énergie libre) profonde, et plus il est difficile de sortir de cette vallée pour relaxer vers le nouvel équilibre. L'évolution de l'aimantation est donc d'autant plus lente que le temps d'attente est long.



De gros progrès [25, 41] ont été faits au cours de la dernière décennie pour comprendre quantitativement comment ces effets de vieillissement modifient les relations de fluctuation-dissipation (5.76, 5.78). Pour un certain nombre de modèles théoriques, il est possible de définir une température effective hors d'équilibre. D'un point de vue expérimental, la susceptibilité  $\chi_{A,B}(t, t')$  et la fonction de corrélation  $C_{A,B}(t, t')$  pour un système de verre de spins ont pu être mesurées et l'on peut voir sur les mesures de Hérisson et Ocio (Fig. 5.9) le point où (5.78) cesse d'être valide.

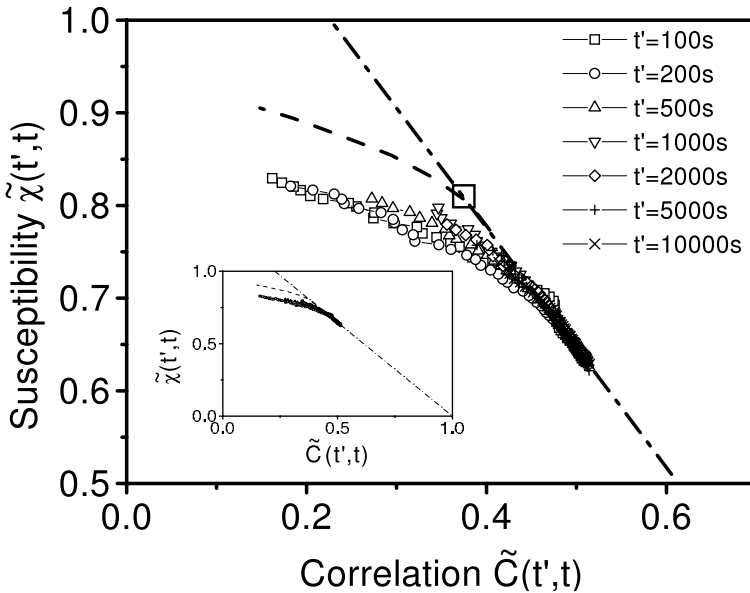


FIG. 5.9 – Susceptibilité magnétique en fonction de l'auto-corrélation de l'aimantation dans un verre de spin. La ligne points-tirets correspond au domaine de validité du théorème de fluctuation-dissipation dynamique (5.78). Hors de ce domaine de validité, la relation entre  $\tilde{C}$  et  $\tilde{\chi}$  dépend explicitement de l'âge  $t'$  du système. (Figure empruntée à D. Hérisson et M. Ocio [43].)

### 5.15.2 Croissance de domaines

Un autre type de systèmes pour lesquels on peut observer des effets de vieillissement sont les problèmes de croissance de domaines, comme dans les mousses ou dans la décomposition spinodale.

Si on laisse évoluer une mousse, on observe un grossissement des domaines au cours du temps donnant lieu à un régime auto similaire : la taille moyenne des domaines augmente au cours du temps, mais la figure composée par les domaines reste statistiquement semblable à elle-même.

Le même phénomène se produit lors de la décomposition spinodale d'un fluide en deux phases, quand on baisse sa température au dessous de la température de coexistence : le fluide a alors le choix entre la phase de basse densité (le gaz) et celle de haute densité (le liquide) et, localement, des domaines de chacune des phases se créent, qui grossissent au cours du temps pour donner un régime auto similaire comme celui de la figure 5.10, qui montre cette croissance dans le cas du modèle d'Ising à température nulle.

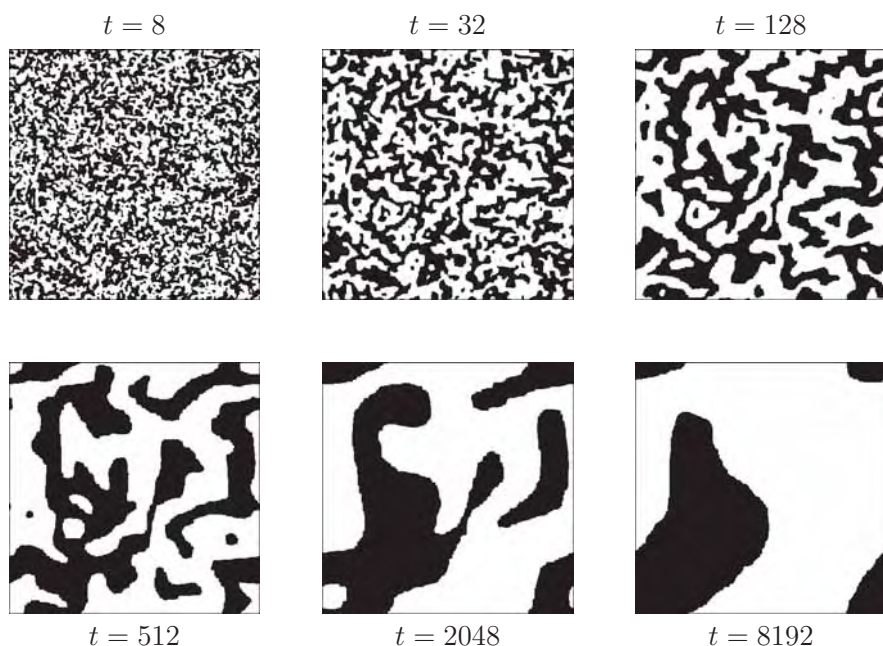


FIG. 5.10 – Évolution d'un modèle d'Ising de taille  $400 \times 400$  sur un réseau carré à température nulle, à partir d'une condition initiale totalement désordonnée.

Le modèle d'Ising est l'un des modèles les plus simples (et les plus étudiés en mécanique statistique) du ferromagnétisme. Sur chaque site d'un réseau régulier se trouve un spin  $S_i$  qui peut prendre deux valeurs possibles :  $S_i = \pm 1$ . On peut penser à ce spin comme à un petit aimant élémentaire qui peut pointer dans deux directions possibles, et l'énergie d'une configuration  $\mathcal{C} = \{S_1, S_2, \dots\}$  des spins sur le réseau est de la forme

$$E(\mathcal{C}) = -J \sum_{i,j \text{ voisins}} S_i S_j$$

(Le modèle d'Ising permet également de modéliser un fluide : un spin  $+1$  représente alors un site occupé par une particule et un spin  $-1$  un site vide.)

L'interaction  $J > 0$  tend à aligner des spins voisins. À mesure que l'on baisse la température, la tendance pour des spins voisins sur le réseau à être

alignés augmente et cette corrélation se propage de sites voisins en sites voisins. Au-dessous d'une certaine température  $T_c$ , la température de Curie, ces corrélations sont devenues suffisantes pour assurer un ordre à longue portée : deux spins, même très éloignés sur le réseau, ont, à l'équilibre, une probabilité plus grande ou égale à  $(1 + m_{\text{spont}}^2)/2$  d'être dans le même état (l'égalité n'étant réalisée que quand la distance entre les deux spins devient infiniment grande) où  $m_{\text{spont}}$  est l'aimantation spontanée. La température de Curie est non-nulle quand le réseau sur lequel se trouvent les spins est infiniment grand et que la dimension de l'espace est supérieure ou égale à 2. (Par exemple, en dimension 2 sur un réseau carré, la température de Curie est  $k_B T_c = 2J/\ln(1 + \sqrt{2}) \simeq 2,269J$ .) En dimension 1, les fluctuations détruisent l'ordre ferromagnétique à toute température non-nulle, mais le système s'ordonne à  $T = 0$ .

Au-dessous de  $T_c$ , et en particulier à  $T = 0$ , si l'on part d'une condition initiale aléatoire, de petits domaines se forment qui croissent au cours du temps, avec une taille caractéristique qui augmente comme  $t^{1/2}$ . Pour un système infiniment grand, la dynamique se ralentit avec le temps (la vitesse des frontières entre les deux phases décroît comme  $t^{-1/2}$ ) sans jamais s'arrêter. Les mesures des fonctions de corrélations montrent que le système reste perpétuellement hors d'équilibre.

Pour implémenter la dynamique à température nulle, il suffit d'actualiser chaque spin avec une probabilité  $dt$  pendant chaque intervalle de temps  $dt$ . Chaque fois qu'on actualise un spin, sa nouvelle valeur devient celle de la majorité de ses voisins, s'il y a une majorité. Si le nombre de voisins  $+$  est égal au nombre de voisins  $-$ , on tire à pile ou face le signe du spin à actualiser. Cette dynamique conduit à des domaines qui grossissent avec le temps comme sur la figure 5.10. Chaque image ressemble à un grossissement de l'image précédente. Pour un système infini, l'évolution ne s'arrête donc jamais.

En général, on ne sait calculer qu'approximativement les fonctions de corrélations à temps différents qui caractérisent ces phénomènes de croissance [44]. Néanmoins, à une dimension, un calcul élémentaire permet d'obtenir l'expression exacte de la fonction de corrélation (aux temps longs) :

$$\langle S_i(t) S_i(t') \rangle \simeq 1 - \frac{2}{\pi} \tan^{-1} \sqrt{\frac{t-t'}{2t'}} \quad (5.79)$$

Cette expression montre bien le phénomène de vieillissement : quand  $t$  et  $t'$  deviennent très grands, la fonction de corrélation ne devient jamais une fonction de la seule différence  $t - t'$  comme pour un système à l'équilibre. La dynamique devient de plus en plus lente sans jamais s'arrêter !

On peut également calculer la fonction de réponse [45] :

$$R_{i,i}(t, t') = \frac{\delta \langle S_i(t) \rangle}{\delta h_i(t')} = \frac{1}{k_B T} \frac{1}{\pi \sqrt{2t'(t-t')}}$$

et l'on voit que la relation de fluctuation-dissipation (5.76) cesse d'être vérifiée dès que  $t - t'$  et  $t$  sont du même ordre de grandeur.

À partir de cette expression de  $R_{i,i}(t, t')$ , on a tracé, sur la figure 5.11, la fonction de réponse  $\chi_{i,i}(t, t')$  définie comme dans (5.77) en fonction de l'autocorrélation  $C_{i,i}(t, t') = \langle S_i(t)S_i(t') \rangle$  donnée par (5.79). Comme sur la figure 5.9, l'écart à la droite en pointillé montre que la relation de fluctuation-dissipation (5.78) n'est pas vérifiée pour ce système hors d'équilibre.

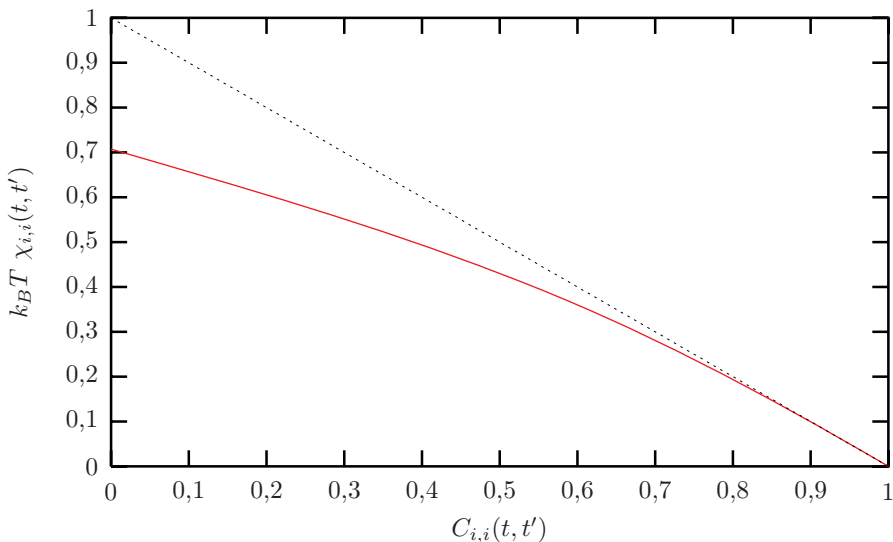


FIG. 5.11 – Susceptibilité en fonction de l'auto-corrélation d'une chaîne d'Ising à température nulle. L'écart à la droite en pointillé montre que la relation de fluctuation-dissipation n'est pas valable pour ce système hors d'équilibre.

Nous remercions Vincent Hakim, Yves Pomeau et Marc Yor de nous avoir permis de clarifier certains aspects exposés dans ce chapitre lors de discussions.

## 5.16 Références

- [1] A. Pais, *Albert Einstein, la vie et l'œuvre*, Paris, InterEditions (1993).
- [2] P. Nakroshis, M. Amoroso, J. Legere, C. Smith, *Measuring Boltzmann's constant using video microscopy of Brownian motion*, Am. J. Phys. **71**, 568 (2003).
- [3] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press (1967).
- [4] J. Perrin, *Les atomes*, Paris, Felix Alcan (1912).
- [5] M.D. Haw, *Colloidal suspensions, Brownian motion, molecular reality: a short history*, J. Phys. C **14**, 7769 (2002).

- [6] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier (2001).
- [7] B.J. Alder, T.E. Wainwright, *Decay of velocity autocorrelation function*, Phys. Rev. A **1**, 18 (1970).
- [8] K. Huang, *Statistical Mechanics*, John Wiley & Sons (1987).
- [9] J. Johnson, *Thermal agitation of electricity in conductors*, Phys. Rev. **32**, 97 (1928).
- [10] H. Nyquist, *Thermal agitation of electric charge in conductors*, Phys. Rev. **32**, 110 (1928).
- [11] D. Durr, S. Goldstein, J.L. Lebowitz, *A mechanical model of Brownian Motion*, Comm. Math. Phys. **78**, 507 (1981).
- [12] B.B. Mandelbrot, *Fractals: form, chance & dimension*, San Francisco, W.H. Freeman ed. (1977).
- [13] P. Lévy, *Théorie de l'addition de variables aléatoires*, Paris, Gauthier-Villard (1954).
- [14] O. Malcai, O. Biham, S. Solomon, *Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements*, Phys. Rev. E **60**, 1299 (1999).
- [15] R.N. Mantegna, H.E. Stanley, *An introduction to econophysics*, Cambridge University Press (2000).
- [16] P. Bak, *How Nature works*, New York, Springer-Verlag (1996).
- [17] E.W. Montroll, M.F. Shlesinger, *On the Wonderful World of Random Walks*, in Nonequilibrium Phenomena II, J.L. Lebowitz et E.W. Montroll eds., Elsevier (1984).
- [18] C. Itzykson, J.M. Drouffe, *Statistical field theory, from Brownian motion to renormalization and lattice gauge theory*, Cambridge University Press (1991).
- [19] B.D. Hughes, *Random walks and random environments, Volume I, Random walks*, Oxford University Press (1995).
- [20] J. des Cloizeaux, G. Jannink, *Les polymères en solution*, Les Ulis, Éditions de Physique (1987).
- [21] P.-G. de Gennes, *Scaling concepts in polymer physics*, Cornell University Press (1979, 1<sup>re</sup> édn. ; 1985, 2<sup>e</sup> édn.).
- [22] N. Madras, G. Slade, *The self avoiding walk*, Boston, Birkhäuser (1993).
- [23] C. Domb, G.S. Joyce, *Cluster expansion for a polymer chain*, J. Phys. C **5**, 956 (1972).
- [24] M. Mézard, G. Parisi, M. Virasoro, *Spin Glass Theory and Beyond*, Lecture Notes Phys. **9**, World Scientific Publishing (1987).
- [25] J.L. Barrat, M. Feigelman, J. Kurchan, J. Dalibard (eds.), *Relaxations lentes et dynamiques hors d'équilibre en physique de la matière condensée*, Springer-Verlag (2003).
- [26] P.W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev. **109**, 1492 (1958).
- [27] Y.A.G. Sinai, *The limiting behavior of a one-dimensional random walk in a random medium*, Theor. Prob. Appl. **27**, 256 (1982).

- [28] J.P. Bouchaud, A. Georges, *Anomalous diffusion in disordered media. Statistical mechanisms, models and physical applications*, Phys. Rep. **195**, 123 (1990).
- [29] C. Monthus, P. Le Doussal, *Localization of thermal packets and metastable states in the Sinai model*, Phys. Rev. E **65**, 66129 (2002).
- [30] B. Derrida, C. Enaud, J.L. Lebowitz, *The asymmetric exclusion process and Brownian excursions*, J. Stat. Phys. **115**, 365 (2004).
- [31] C. Maes, K. Notecny, *Time-reversal and entropy*, J. Stat. Phys. **110**, 269 (2003).
- [32] J.R. Dorfman, *An introduction to chaos in nonequilibrium statistical mechanics*, Cambridge University Press (1999).
- [33] D. Ruelle, *Smooth Dynamics and New Theoretical Ideas in Nonequilibrium Statistical Mechanics*, J. Stat. Phys. **95**, 393 (1999).
- [34] D.J. Evans, E.G.D. Cohen, G.P. Morris, *Probability of second law violations in shearing steady states*, Phys. Rev. Lett. **71**, 2401 (1993).
- [35] G. Gallavotti, E.D.G. Cohen, *Dynamical ensembles in stationary states*, J. Stat. Phys. **80**, 931 (1995).
- [36] J. Farago, *Injected power fluctuations in Langevin equation*, J. Stat. Phys. **107**, 781 (2002).
- [37] C.W.J. Beenakker, M. Büttiker, *Suppression of shot noise in metallic diffusive conductors*, Phys. Rev. B **46**, 1889 (1992).
- [38] Y.N. Blanter, M. Büttiker, *Shot noise in mesoscopic conductors*, Phys. Rep. **336**, 1 (2000).
- [39] B. Derrida, B. Douçot, P.-E. Roche, *Current fluctuations in the one dimensional Symmetric Exclusion Process with open boundaries*, J. Stat. Phys. **115**, 717 (2004).
- [40] L. Onsager, *Reciprocal Relations in Irreversible Processes*, Phys. Rev. **37**, 405 et Phys. Rev. **38**, 2265 (1931).
- [41] A. Crisanti, F. Ritort, *Violation of the fluctuation-dissipation theorem in glassy systems: basic notions and the numerical evidence*, J. Phys. A **36**, R181 (2003).
- [42] E. Vincent, J. Hammann, M. Ocio, J.-P. Bouchaud, L.F. Cugliandolo, *Slow Dynamics and Aging in Spin Glasses*, Proceedings of the Sitges conference, E. Rubi ed., Springer-Verlag (1997).
- [43] D. Hérisson, M. Ocio, *Fluctuation-dissipation ratio of a spin glass in the aging regime*, Phys. Rev. Lett. **88**, 257202 (2002).
- [44] A. Bray, *Theory of phase ordering kinetics*, Adv. Phys. **43**, 357 (1994).
- [45] C. Godrèche, J.M. Luck, *Non equilibrium dynamics of ferromagnetic spin systems*, J. Phys. C. **14**, 1589 (2002).

# Chapitre 6

## Relativité générale

Par *Thibault Damour*

### Sommaire

6.1	Introduction . . . . .	269
6.2	Relativité restreinte . . . . .	270
6.3	Principe d'équivalence . . . . .	272
6.4	Gravitation et chrono-géométrie de l'espace-temps . . . . .	273
6.5	Équations d'Einstein : l'espace-temps élastique . . . . .	276
6.6	Limite des champs faibles et limite newtonienne . . . . .	279
6.7	Approximation post-newtonienne et confirmations expérimentales dans le régime des champs gravitationnels faibles et quasi-stationnaires . . . . .	283
6.8	Champs gravitationnels forts. Trous noirs . . . . .	286
6.9	Pulsars binaires et confirmations expérimentales du régime des champs gravitationnels forts et radiatifs . . . . .	290
6.10	Ondes gravitationnelles : propagation, génération et détection	296
6.11	Relativité générale et théorie quantique : de la supergravité à la théorie des cordes . . . . .	301
6.12	Conclusion . . . . .	316
6.13	Références . . . . .	316



### **Thibault Damour**

Thibault Damour est professeur à l'Institut des Hautes Études Scientifiques, membre de l'Académie des sciences, récipiendaire de nombreux prix dont la médaille Einstein. Ses travaux portent sur la physique de la gravitation et tout particulièrement les rapports entre les théories et les expériences. Il a introduit de nouveaux concepts pour la physique des trous noirs et des pulsars binaires dans le cadre de la relativité générale. Il travaille actuellement sur les relations entre la théorie des cordes et la gravitation, la cosmologie primordiale et les tests expérimentaux possibles de la théorie des cordes.



# Relativité générale

## 6.1 Introduction

Le vocable « relativité » recouvre, en physique, deux concepts très différents. Celui de *relativité restreinte* désigne la théorie physique de la structure de l'espace-temps de Poincaré-Minkowski, issue de la nouvelle cinématique des mesures d'espace et de temps proposée par Einstein en juin 1905 [1, 2] (après d'importants travaux pionniers de Lorentz et de Poincaré). Cette théorie, initialement suggérée par l'étude de l'électromagnétisme, est devenue l'une des « théorie-cadres » de la physique du vingtième siècle, c'est-à-dire qu'elle a servi de cadre fondamental pour la formulation de toutes les lois fondamentales de la nature, à l'exception cependant de la *gravitation*. En effet, Einstein, dans un travail étendu sur les années 1907–1915, s'est rendu compte que l'interaction gravitationnelle possédait des caractères particuliers qui lui ont suggéré la nécessité de *généraliser* sa théorie de la relativité de 1905. Cette théorie de la *relativité généralisée* (ou « relativité générale ») a apporté une modification profonde au concept d'espace-temps. De simple arène neutre, donnée a priori, indépendamment de tout contenu matériel, et servant de cadre à l'existence et à l'évolution de la matière, l'espace-temps est devenu, en 1915, un « champ » physique (identifié au champ gravitationnel), c'est-à-dire une entité dynamique influencée par et influençant la distribution de matière-énergie qu'il contient.

Cette conception radicalement nouvelle de la structure de l'espace-temps est longtemps restée en marge du développement de la physique. La théorie de la relativité générale apparaissait à la fois comme une théorie mal confirmée par l'expérience, et sans contact avec les progrès extraordinaires issus de l'application de la théorie quantique (et de la relativité restreinte) à la physique des hautes énergies. Cette situation de marginalisation de la relativité générale n'est plus du tout d'actualité. Aujourd'hui la relativité générale est devenue l'un des acteurs essentiels de la science de pointe. De nombreux tests expérimentaux de haute précision ont confirmé, dans le détail, la pertinence de cette théorie. La relativité générale est devenue un outil privilégié de description de l'univers macroscopique, depuis le big bang jusqu'aux trous noirs, en passant par le système solaire, les étoiles à neutrons, les pulsars et les ondes gravitationnelles. De plus, la recherche d'une cohérence de toute

la physique fondamentale a suggéré l'exploration de théories unifiant, dans un cadre quantique général, la description de la matière et de toutes ses interactions (y compris l'interaction gravitationnelle). Ces théories, encore en construction, et provisoirement appelées théories des cordes, contiennent de façon centrale la relativité générale, mais suggèrent que la structure fondamentale de l'espace-temps-matière est encore plus riche que ce que suggèrent séparément la théorie quantique et la relativité générale.

## 6.2 Relativité restreinte

Rappelons que la structure de l'espace-temps de Poincaré-Minkowski est donnée par une généralisation de la structure géométrique euclidienne de l'espace ordinaire. Cette dernière est résumée par la formule  $L^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$  (conséquence du théorème de Pythagore), exprimant le carré de la distance  $L$  entre deux points de l'espace en fonction de la somme des carrés des différences des coordonnées (orthonormales)  $x, y, z$  repérant les deux points. Le groupe de symétrie de la géométrie euclidienne est le groupe des transformations des coordonnées  $(x, y, z) \rightarrow (x', y', z')$  laissant invariante la forme  $L^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ . (Ce groupe est engendré par les translations, les rotations et les « retournements », comme par exemple la transformation dans un miroir :  $x' = -x, y' = y, z' = z$ .)

L'espace-temps de Poincaré-Minkowski est défini comme l'ensemble des *événements* (idéalisations de ce qui se passe en un point particulier de l'espace, à un moment particulier du temps), muni de la notion d'*intervalle* (carré)  $S^2$  entre deux événements. Un événement est repéré par quatre coordonnées,  $x, y, z, t$ , où  $(x, y, z)$  sont les coordonnées spatiales du point de l'espace « où se passe » l'événement en question, et où  $t$  repère l'instant « quand se passe » cet événement. Un autre événement sera repéré (dans le même référentiel) par quatre coordonnées différentes, disons  $x + \Delta x, y + \Delta y, z + \Delta z, t + \Delta t$ . Les points de l'espace où se passent ces deux événements sont séparés d'une distance  $L$  donnée par la formule ci-dessus,  $L^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ . Les moments du temps « quand se passent » ces deux événements sont séparés d'une durée  $T$  donnée par  $T = \Delta t$ . En fonction de ces quantités, l'intervalle carré  $S^2$  entre les deux événements est donné, par définition, par la généralisation suivante du théorème de Pythagore :

$$S^2 = L^2 - c^2 T^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2 (\Delta t)^2 \quad (6.1)$$

où  $c$  désigne la vitesse de la lumière (ou plus exactement la vitesse maximale de propagation des signaux).

La formule (6.1) définit la *chrono-géométrie* de l'espace-temps de Poincaré-Minkowski. Le groupe de symétrie de cette chrono-géométrie est le groupe des transformations des coordonnées  $(x, y, z, t) \rightarrow (x', y', z', t')$  laissant invariante la forme (6.1) de l'intervalle  $S$ . On démontre que ce groupe est constitué de transformations linéaires et qu'il est engendré par les translations d'espace et

de temps, les rotations spatiales, les « boutées »<sup>1</sup> (c'est-à-dire les transformations de Lorentz spéciales), et les retournements d'espace et de temps.

Il est commode de remplacer la coordonnée temporelle  $t$  par le « temps-lumière »  $x^0 \equiv ct$ , et de noter collectivement  $x^\mu \equiv (x^0, x^i)$  où les indices grecs  $\mu, \nu, \dots = 0, 1, 2, 3$ , et les indices latins  $i, j, \dots = 1, 2, 3$  (avec  $x^1 = x$ ,  $x^2 = y$ ,  $x^3 = z$ ). La formule (6.1) s'écrit alors

$$S^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu \quad (6.2)$$

où l'on a utilisé la convention de sommation d'Einstein<sup>2</sup>, et où  $\eta_{\mu\nu}$  est une matrice diagonale dont les seuls éléments non nuls sont  $\eta_{00} = -1$  et  $\eta_{11} = \eta_{22} = \eta_{33} = +1$ . Le groupe de symétrie de l'espace-temps de Poincaré-Minkowski est alors l'ensemble des transformations de Lorentz-Poincaré,

$$x'^\mu = \Lambda_\nu^\mu x^\nu + a^\mu \quad (6.3)$$

où  $\eta_{\alpha\beta} \Lambda_\mu^\alpha \Lambda_\nu^\beta = \eta_{\mu\nu}$ .

La chrono-géométrie de l'espace-temps de Poincaré-Minkowski peut être visualisée en représentant, autour de chaque point  $x$  de l'espace-temps, le lieu des points qui sont séparés du point  $x$  par un intervalle (carré) unité, c'est-à-dire l'ensemble des points  $x'$  tels que  $S_{xx'}^2 = \eta_{\mu\nu}(x'^\mu - x^\mu)(x'^\nu - x^\nu) = +1$ . Ce lieu est un hyperboloïde (unité) à une nappe.

Si l'on était dans un espace euclidien habituel, l'ensemble des points  $x'$  tracerait une sphère (unité) centrée sur  $x$ , et le « champ » de ces sphères centrées en chaque point  $x$  permettrait de caractériser complètement la géométrie euclidienne de l'espace. De même, dans le cas de l'espace-temps de Poincaré-Minkowski, le « champ » des hyperboloïdes unités centrés en chaque point  $x$  est une caractérisation visuelle de la géométrie de cet espace-temps (Fig. 6.1). Cette figure donne une idée du groupe de symétrie de l'espace-temps de Poincaré-Minkowski, et rend en particulier clair le caractère rigide et homogène de sa géométrie.

L'idée essentielle de l'article d'Einstein de juin 1905 était d'imposer le groupe de transformations (6.3) comme groupe de symétrie des lois fondamentales de la physique (« principe de relativité »). Ce point de vue s'est révélé être extraordinairement fructueux, car il a conduit à la découverte de nouvelles lois et à la prédiction de nouveaux phénomènes. Citons seulement pour mémoire : la dynamique relativiste des particules classiques, la dilatation du temps de vie des particules relativistes, la relation  $E = mc^2$  entre énergie et masse inertielle, la théorie relativiste des particules quantiques de spin  $\frac{1}{2}$  de Dirac, la prédiction de l'antimatière, la classification des particules par la masse au repos et le spin, la relation entre spin et statistique, le théorème CPT.

Cependant, Einstein remarqua rapidement (1907) que la gravitation posait un problème particulier, et ne pouvait pas être traitée, comme les autres

<sup>1</sup>Vieux mot français, du verbe « bouter », et correspondant à l'anglais « boosts ».

<sup>2</sup>Tout indice répété est supposé être sommé sur toutes ses valeurs possibles.

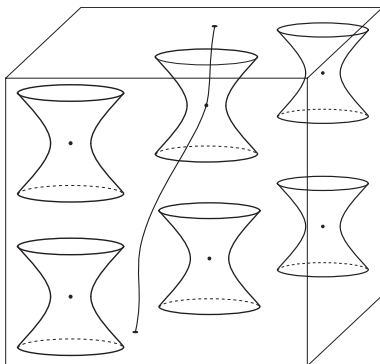


FIG. 6.1 – Géométrie de l'espace-temps « rigide » de la théorie de la relativité restreinte. Cette géométrie est visualisée en représentant, autour de chaque point  $x$  de l'espace-temps, le lieu des points séparés du point  $x$  par un intervalle (carré) unité. L'espace-temps représenté ici n'a que trois dimensions : une dimension temporelle (représentée verticalement),  $x^0 = ct$ , et deux dimensions spatiales (représentées horizontalement),  $x, y$ . On a aussi visualisé la ligne d'espace-temps (allant de bas en haut du « bloc espace-temps », c'est à dire du passé vers le futur) représentant l'histoire du mouvement d'une particule.

interactions, dans le cadre de la théorie de la relativité restreinte. Ceci conduisit Einstein à généraliser la théorie de la relativité.

### 6.3 Principe d'équivalence

Le point de départ d'Einstein est un fait expérimental remarquable : tous les corps tombent, dans un champ gravitationnel extérieur, avec la même accélération. Ce fait avait été indiqué par Galilée en 1638. Par une combinaison remarquable de raisonnements logiques, d'expériences de pensée et d'expériences réelles faites sur des plans inclinés<sup>3</sup>, Galilée fut en effet le premier à concevoir ce qu'on appelle aujourd'hui « universalité de la chute libre », ou « principe d'équivalence faible ». Citons la conclusion que Galilée tire d'un raisonnement où il fait varier le rapport entre la densité des corps en chute libre considérés et la résistance du milieu dans lequel ils tombent : « *c'est alors, considérant ces faits qu'il me vint à l'esprit que si l'on supprimait totalement la résistance du milieu, tous les corps descendraient avec la même vitesse* » [3]. Cette universalité de la chute libre fut vérifiée avec plus de précision par des expériences de Newton sur des pendules, et fut incorporée par lui dans sa théorie de la gravitation (1687) sous la forme d'une identification entre masse inertielle  $m_i$  (apparaissant dans la loi fondamentale de la

<sup>3</sup>L'expérience de chute de corps supposée faite depuis le haut de la tour de Pise est un mythe, même s'il résume bien l'essentiel de l'innovation galiléenne.

dynamique  $\mathbf{F} = m_i \mathbf{a}$ ) et masse gravitationnelle  $m_g$  (apparaissant dans la force gravitationnelle,  $F_g = G m_g m'_g / r^2$ ) :

$$m_i = m_g \quad (6.4)$$

À la fin du 19<sup>e</sup> siècle, le baron Roland von Eötvös vérifia l'équivalence (6.4) entre  $m_i$  et  $m_g$  avec une précision de l'ordre de  $10^{-9}$ , et Einstein était au courant de cette vérification de haute précision. (À l'heure actuelle, l'équivalence entre  $m_i$  et  $m_g$  est vérifiée au niveau  $10^{-12}$  [4].) Le point qui frappa Einstein fut que, vu la précision avec laquelle  $m_i = m_g$  était vérifiée, et vu l'équivalence entre masse inertielle et énergie découverte par Einstein en septembre 1905 [2] ( $E = m_i c^2$ ), il fallait conclure que toutes les formes très diverses d'énergie contribuant à la masse d'un corps (masse au repos des constituants élémentaires, énergies de liaison variées, énergie cinétique interne, ...) contribuaient de façon strictement identique à la masse gravitationnelle de ce corps, c'est-à-dire à la fois à sa capacité à réagir à un champ gravitationnel extérieur, et à sa capacité à créer un champ gravitationnel.

En 1907, Einstein réalisa que l'équivalence entre  $m_i$  et  $m_g$  contenait implicitement une équivalence plus profonde entre inertie et gravitation ayant des conséquences importantes pour la notion de référentiel inertiel (lequel était un concept fondamental de la théorie de la relativité restreinte). Dans une expérience de pensée novatrice, Einstein imagina le comportement de solides et d'horloges de référence dans un ascenseur en chute libre. À cause de l'universalité de la chute libre, dans un tel « référentiel local en chute libre » tous les corps présents dans ce référentiel apparaîtront comme non accélérés par rapport à lui. Donc, par rapport à un tel repère, le champ gravitationnel extérieur est « effacé ». Einstein postula alors ce qu'il appela le « principe d'équivalence » entre gravitation et inertie. Ce principe a deux volets, qu'Einstein utilisa tour à tour. Le premier volet dit que, pour n'importe quel champ gravitationnel extérieur, il est possible d'« effacer » localement le champ de gravitation en utilisant un référentiel local en chute libre adéquat, et que, du coup, les lois de la physique non gravitationnelle s'appliquent dans ce référentiel local, comme elles le faisaient, dans un référentiel inertiel (libre de gravitation) en relativité restreinte. Le deuxième volet du principe d'équivalence d'Einstein dit que, en partant d'un référentiel inertiel en relativité restreinte (et en l'absence de tout « vrai » champ de gravitation), on peut créer un champ gravitationnel apparent dans un référentiel local, si ce référentiel est accéléré (soit en ligne droite, soit par une rotation).

## 6.4 Gravitation et chrono-géométrie de l'espace-temps

Einstein sut (par un extraordinaire chemin intellectuel qui dura huit ans) construire une nouvelle théorie de la gravitation, basée sur une ample

généralisation de la théorie de la relativité de 1905, en partant seulement du principe d'équivalence énoncé ci-dessus. La première étape de ce processus a consisté à comprendre que le principe d'équivalence suggérait une modification profonde de la structure chrono-géométrique de l'espace-temps de Poincaré-Minkowski rappelé dans l'équation (6.1) ci-dessus. En effet, soit  $X^\alpha$ ,  $\alpha = 0, 1, 2, 3$ , les coordonnées d'espace-temps dans un référentiel local en chute libre (ou *référentiel localement inertiel*). Dans un tel référentiel, les lois de la relativité restreinte s'appliquent. En particulier, l'intervalle infinitésimal d'espace-temps  $ds^2 = dL^2 - c^2 dT^2$  entre deux événements infiniment voisins d'un tel référentiel  $X^\alpha$ ,  $X'^\alpha = X^\alpha + dX^\alpha$  (proches du centre de ce référentiel) prend la forme

$$ds^2 = dL^2 - c^2 dT^2 = \eta_{\alpha\beta} dX^\alpha dX^\beta \quad (6.5)$$

où l'on rappelle que les indices répétés  $\alpha$  et  $\beta$  sont sommés sur toutes leurs valeurs ( $\alpha, \beta = 0, 1, 2, 3$ ). On sait aussi qu'en relativité restreinte les densités et flux locaux d'énergie et d'impulsion se rassemblent dans les dix composantes du *tenseur d'énergie-impulsion*  $T^{\alpha\beta}$ . (Par exemple, la densité volumique d'énergie est égale, dans le référentiel  $X^\alpha = (X^0, X^i)$ ,  $i = 1, 2, 3$  à  $T^{00}$ .) La conservation de l'énergie et de l'impulsion se traduit par l'équation  $\partial_\beta T^{\alpha\beta} = 0$ , où  $\partial_\beta = \partial/\partial X^\beta$ .

La théorie de la relativité restreinte nous dit que l'on peut changer de référentiel localement inertiel (tout en restant dans le voisinage d'un point d'espace-temps où l'on a « effacé » la gravitation) par une transformation de Lorentz,  $X'^\alpha = \Lambda^\alpha_\beta X^\beta$ . Sous une telle transformation, l'intervalle infinitésimal  $ds^2$  (Éq. (6.5)) reste invariant, et les 10 composantes du tenseur (symétrique)  $T^{\alpha\beta}$  se transforment selon  $T'^{\alpha\beta} = \Lambda^\alpha_\gamma \Lambda^\beta_\delta T^{\gamma\delta}$ . En revanche, quand on passe d'un référentiel *localement inertiel* (de coordonnées  $X^\alpha$ ) à un référentiel non inertiel *étendu* (de coordonnées  $x^\mu$ ;  $\mu = 0, 1, 2, 3$ ) la transformation reliant les  $X^\alpha$  aux  $x^\mu$  n'est plus une transformation *linéaire* (comme la transformation de Lorentz) mais devient une transformation *non linéaire*  $X^\alpha = X^\alpha(x^\mu)$  d'une forme qui pourra être quelconque. Du coup, la valeur de l'intervalle infinitésimal  $ds^2$  va prendre, quand on l'exprime dans un référentiel étendu général, une forme plus compliquée que la forme très simple, équation (6.5), qu'elle avait dans un référentiel localement en chute libre. En effet, en différenciant les fonctions non linéaires  $X^\alpha = X^\alpha(x^\mu)$  on obtient le lien  $dX^\alpha = \partial X^\alpha / \partial x^\mu dx^\mu$ . En remplaçant cette relation dans (6.5) on obtient ensuite

$$ds^2 = g_{\mu\nu}(x^\lambda) dx^\mu dx^\nu \quad (6.6)$$

où les indices  $\mu, \nu$  sont sommés sur  $0, 1, 2, 3$  et où les dix fonctions  $g_{\mu\nu}(x)$  (symétriques sur les indices  $\mu$  et  $\nu$ ) des quatre variables  $x^\lambda$  sont définies, point par point (c'est-à-dire en considérant pour chaque point  $x^\lambda$  un référentiel localement en chute libre au point  $x$ , de coordonnées locales  $X^\alpha_x$ ) par  $g_{\mu\nu}(x) = \eta_{\alpha\beta} \partial X^\alpha_x(x) / \partial x^\mu \partial X^\beta_x(x) / \partial x^\nu$ . À cause de la non linéarité des fonctions  $X^\alpha(x)$ , les fonctions  $g_{\mu\nu}(x)$  dépendent en général de façon non triviale des coordonnées de  $x$ .

La chrono-géométrie locale de l'espace-temps apparaît ainsi comme donnée non pas par la métrique minkowskienne simple (6.2), avec des coefficients constants  $\eta_{\mu\nu}$ , mais par une métrique quadratique d'un type beaucoup plus général, équation (6.6), avec des coefficients  $g_{\mu\nu}(x)$  variables de point en point. De tels espaces métriques généraux ont été introduits et étudiés par Gauss et Riemann au XIX<sup>e</sup> siècle (dans le cas où la forme quadratique (6.6) est définie positive). Ils portent le nom d'*espaces riemanniens* ou d'*espaces courbes*. (Dans le cas, d'intérêt pour la théorie d'Einstein, où la forme quadratique (6.6) n'est pas définie positive, on parle de métrique pseudo-riemannienne.)

Nous n'avons pas la place ici d'expliquer en détail les diverses structures géométriques induites, dans un espace riemannien, par la donnée de l'intervalle infinitésimal (6.6). Indiquons seulement que la donnée, équation (6.6), de la distance  $ds$  entre deux points infiniment voisins permet, par intégration le long d'une ligne, de définir la longueur d'une ligne quelconque joignant deux points éloignés,  $A$  et  $B$  :  $L_{AB} = \int_A^B ds$ . On peut alors définir la « ligne la plus droite possible » entre deux points donnés  $A$  et  $B$  comme la ligne la plus courte, c'est-à-dire celle qui minimise (ou, en général, extrémise) la distance intégrée  $L_{AB}$ . Ces lignes les plus droites possibles s'appellent *lignes géodésiques*. Pour donner un exemple simple, les géodésiques d'une surface sphérique (comme la surface de la Terre) sont les grands cercles (de rayon égal au rayon de la sphère). Si l'on écrit mathématiquement les conditions pour qu'une ligne, donnée par sa représentation paramétrique  $x^\mu = x^\mu(s)$  où  $s$  est la longueur le long de la ligne, extrémise la longueur totale  $L_{AB}$  on trouve que  $x^\mu(s)$  doit satisfaire l'équation différentielle du second ordre suivante

$$\frac{d^2 x^\lambda}{ds^2} + \Gamma_{\mu\nu}^\lambda(x) \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0 \quad (6.7)$$

où les quantités  $\Gamma_{\mu\nu}^\lambda$ , appelés *coefficients de Christoffel* ou *coefficients de connexion*, sont calculés, en chaque point  $x$ , à partir des *composantes de la métrique*  $g_{\mu\nu}(x)$  selon

$$\Gamma_{\mu\nu}^\lambda \equiv \frac{1}{2} g^{\lambda\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) \quad (6.8)$$

où  $g^{\mu\nu}$  désigne la matrice inverse de  $g_{\mu\nu}$  ( $g^{\mu\sigma} g_{\sigma\nu} = \delta_\nu^\mu$  où le symbole de Kronecker  $\delta_\nu^\mu$  vaut 1 quand  $\mu = \nu$  et 0 sinon) et où  $\partial_\mu \equiv \partial/\partial x^\mu$  désigne la dérivée partielle par rapport à la coordonnée  $x^\mu$ . Pour donner un exemple très simple : dans l'espace-temps de Poincaré-Minkowski les composantes de la métrique sont des constantes  $g_{\mu\nu} = \eta_{\mu\nu}$  (quand on utilise un référentiel inertiel). Du coup, les coefficients de connexion (6.8) sont nuls dans un référentiel inertiel, et l'équation différentielle des géodésiques se réduit à  $d^2 x^\lambda/ds^2 = 0$  dont les solutions sont des lignes droites ordinaires :  $x^\lambda(s) = a^\lambda s + b^\lambda$ . En revanche, dans un espace-temps « courbe » général (c'est-à-dire avec des  $g_{\mu\nu}$  qui dépendent de façon arbitraire du point  $x$ ) les géodésiques ne pourront pas être *globalement* représentées par des lignes droites. On montre cependant qu'il

reste toujours possible, pour des  $g_{\mu\nu}(x)$  quelconques, de changer de coordonnées  $x^\mu \rightarrow X^\alpha(x)$  de telle sorte que les coefficients de connexion  $\Gamma_{\beta\gamma}^\alpha$ , dans le nouveau système  $X^\alpha$  de coordonnées, s'annulent *localement*, en un point donné  $X_0^\alpha$  (ou même le long d'une ligne quelconque). De tels systèmes de coordonnées *localement géodésiques* réalisent mathématiquement le principe d'équivalence d'Einstein : à des termes du deuxième ordre près, les composantes  $g_{\alpha\beta}(X)$  d'une métrique « courbe » dans des coordonnées localement géodésiques  $X^\alpha$  ( $ds^2 = g_{\alpha\beta}(X) dX^\alpha dX^\beta$ ) peuvent être identifiées aux composantes d'une métrique « plate » de Poincaré-Minkowski :  $g_{\alpha\beta}(X) = \eta_{\alpha\beta} + \mathcal{O}((X - X_0)^2)$ , où  $X_0$  est le point autour duquel on développe.

## 6.5 Équations d'Einstein : l'espace-temps élastique

Ayant postulé qu'une théorie relativiste cohérente du champ gravitationnel devait inclure la considération d'une vaste généralisation de l'espace-temps de Poincaré-Minkowski, équation (6.6), Einstein en conclut que les dix mêmes fonctions  $g_{\mu\nu}(x)$  devaient décrire à la fois la géométrie de l'espace-temps et la gravitation. Il s'attela alors à la tâche de trouver quelles équations devaient satisfaire le « champ géométrico-gravitationnel »  $g_{\mu\nu}(x)$ . Trois principes le guidèrent dans cette recherche. Le premier est le *principe de relativité générale* qui affirme qu'en présence d'un champ gravitationnel les lois fondamentales de la physique (y compris celles du champ gravitationnel lui-même) doivent pouvoir s'écrire de la même façon dans n'importe quel système de coordonnées. Le deuxième est que la « source » du champ gravitationnel doit être le tenseur d'énergie-impulsion  $T^{\mu\nu}$ . Le troisième est un principe de *correspondance* avec la physique antérieure : dans la limite où l'on néglige les effets gravitationnels  $g_{\mu\nu}(x) = \eta_{\mu\nu}$  doit être une solution des équations recherchées, et il doit aussi exister une limite (dite *Newtonienne*) où la nouvelle théorie se réduit à la théorie de la gravitation de Newton.

Notons que le principe de relativité générale (contrairement aux apparences et contrairement à ce qu'Einstein crut pendant quelques années) a un statut physique différent du principe de relativité restreinte. Le principe de relativité restreinte était un principe de symétrie de la structure de l'espace-temps qui affirmait que la physique était *la même* dans une classe particulière de référentiels, et donc que certains phénomènes « correspondants » se déroulaient exactement de la même façon dans des référentiels différents (transformations « actives »). En revanche, le principe de relativité générale est un *principe d'indifférence* : les phénomènes ne se déroulent (en général) pas de la même façon dans des systèmes de coordonnées différents, mais aucun des systèmes (étendus) de coordonnées n'a de statut privilégié par rapport aux autres.

Le principe affirmant que le tenseur d'impulsion-énergie  $T^{\mu\nu}$  doit être la source du champ gravitationnel est fondé sur deux idées : les relations



$E = m_i c^2$  et le principe d'équivalence faible  $m_i = m_g$  montrent que, dans la limite newtonienne, la source de la gravitation, la masse gravitationnelle  $m_g$ , est égale, au facteur  $c^{-2}$  près, à l'énergie totale du corps considéré c'est-à-dire à l'intégrale sur l'espace de la densité d'énergie  $T^{00}$ . Donc au moins une des composantes du tenseur  $T^{\mu\nu}$  doit jouer le rôle de source pour le champ gravitationnel. Mais comme ce dernier est décrit, d'après Einstein, par les dix composantes de la métrique  $g_{\mu\nu}$ , il est naturel de supposer que la source de  $g_{\mu\nu}$  doit aussi avoir dix composantes, ce qui est justement le cas du tenseur (symétrique)  $T^{\mu\nu}$ .

Après de longues années d'un travail conceptuel ardu, Einstein écrivit, en novembre 1915 [2, 5], la forme finale de la théorie de la relativité générale [6].

Les « équations d'Einstein » sont des équations aux dérivées partielles non-linéaires du second ordre pour le champ géométrico-gravitationnel  $g_{\mu\nu}$ , contenant le tenseur d'énergie impulsion  $T_{\mu\nu} \equiv g_{\mu\kappa} g_{\nu\lambda} T^{\kappa\lambda}$  au second membre. Elles s'écrivent comme

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \quad (6.9)$$

où  $G$  est la constante (newtonienne) de la gravitation,  $c$  la vitesse de la lumière, et  $R \equiv g^{\mu\nu} R_{\mu\nu}$  où le *tenseur de Ricci*  $R_{\mu\nu}$  se calcule en fonction des coefficients de connexion  $\Gamma_{\mu\nu}^\lambda$  (6.8) de la façon suivante

$$R_{\mu\nu} \equiv \partial_\alpha \Gamma_{\mu\nu}^\alpha - \partial_\nu \Gamma_{\mu\alpha}^\alpha + \Gamma_{\beta\alpha}^\alpha \Gamma_{\mu\nu}^\beta - \Gamma_{\beta\nu}^\alpha \Gamma_{\mu\alpha}^\beta \quad (6.10)$$

On démontre que les trois principes que nous avons énoncés ci-dessus déterminent uniquement, dans un espace-temps à quatre dimensions, les équations d'Einstein (6.9). Il est d'ailleurs remarquable que ces équations puissent aussi être élaborées à partir de points de vue complètement différents de celui pris par Einstein. Par exemple, dans les années soixante divers auteurs (notamment Feynman, Weinberg et Deser ; voir références dans [4]) montrèrent que les équations d'Einstein pouvaient être obtenues par une approche purement « dynamique », fondée sur la cohérence des interactions d'un champ de spin 2 à longue portée, sans faire appel, comme le fit Einstein, aux notions « géométriques » issues des travaux mathématiques sur les espaces riemanniens. Signalons aussi que si l'on relâche une partie des principes énoncés ci-dessus (comme le fit Einstein en 1917) on peut aboutir à une généralisation des équations (6.9) dans laquelle on rajoute au membre de gauche le terme  $+\Lambda g_{\mu\nu}$  où  $\Lambda$  est une constante appelée *constante cosmologique*. Une telle modification avait été proposée par Einstein en 1917 pour pouvoir écrire une solution cosmologique globalement homogène et *stationnaire*. Einstein rejeta ce terme après les travaux de Friedmann (1922) montrant l'existence de solutions cosmologiques *en expansion* de la relativité générale et après la découverte observationnelle (par Hubble, 1929) du mouvement d'expansion de l'ensemble des galaxies. Cependant les données cosmologiques récentes ont remis au goût du jour cette possibilité (voir chapitre 7), même si la physique

fondamentale d'aujourd'hui tend à considérer qu'un terme du type  $\Lambda g_{\mu\nu}$  doit être considéré comme une contribution physique particulière au membre de droite des équations d'Einstein (plus précisément comme le tenseur d'énergie-impulsion du « vide »,  $T_{\mu\nu}^V = -\frac{c^4}{8\pi G} \Lambda g_{\mu\nu}$ ), plutôt que comme une modification géométrique universelle du membre de gauche.

Commentons le sens physique des équations d'Einstein (6.9). L'idée nouvelle essentielle est que la structure chrono-géométrique de l'espace-temps, équation (6.6), c'est-à-dire la structure qui sous-tend toutes les mesures que l'on peut faire localement de durées,  $dT$ , et de distance,  $dL$ , (rappelons que, localement,  $ds^2 = dL^2 - c^2 dT^2$ ) n'est plus une structure rigide, donnée a priori, une fois pour toutes (comme l'était la structure de l'espace-temps de Poincaré-Minkowski), mais est devenue un *champ*, c'est-à-dire une structure dynamique, ou *élastique*, qui est créée et/ou déformée par la présence d'une distribution d'énergie-impulsion. Voir la figure 6.2 qui visualise la géométrie « élastique » de l'espace-temps de la théorie de la relativité générale en représentant, autour de chaque point  $x$ , le lieu des points (supposés être infiniment proches de  $x$ ) séparés de  $x$  par un intervalle (carré) constant :  $ds^2 = \varepsilon^2$ . Comme dans le cas de l'espace-temps de Poincaré-Minkowski (Fig. 6.1), on obtient ainsi un « champ » d'hyperboloïdes, mais maintenant ce champ d'hyperboloïdes n'a plus une structure « rigide » et homogène.

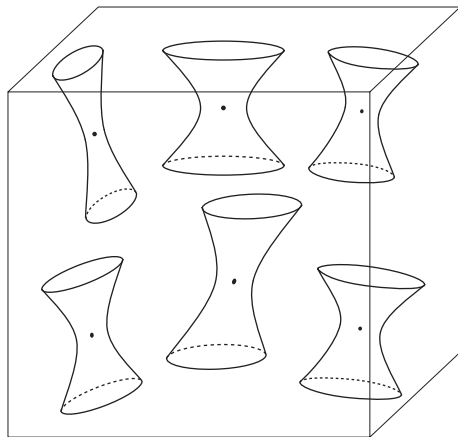


FIG. 6.2 – Géométrie de l'espace-temps « élastique » de la théorie de la relativité générale. Cette géométrie est visualisée en représentant, autour de chaque point  $x$  de l'espace-temps, le lieu des points séparés de  $x$  par un petit intervalle (carré) positif donné.

Le *champ* espace-temps  $g_{\mu\nu}(x)$  décrit à la fois la variabilité de point en point de la chrono-géométrie et tous les effets gravitationnels. L'exemple le plus simple d'*élasticité* de la chrono-géométrie de l'espace-temps est l'effet de la proximité d'une masse sur la « vitesse d'écoulement local du temps ». En termes concrets, si vous séparez deux jumeaux à la naissance, l'un restant

à la surface de la Terre et l'autre allant vivre au sommet d'une très haute montagne (c'est-à-dire plus loin du centre de la Terre), et si vous les réunissez après 100 ans, le jumeau « montagnard » sera plus vieux (aura vécu plus longtemps) que le jumeau resté sur le plancher des vaches. Tout se passe comme si le temps s'écoulait d'autant moins vite que l'on est plus près d'une distribution de masse-énergie. En termes de représentation mathématique cet effet est dû au fait que le coefficient  $g_{00}(x)$  de  $(dx^0)^2$  dans l'équation (6.6) est déformé par rapport à sa valeur en relativité restreinte,  $g_{00}^{\text{Minkowski}} = \eta_{00} = -1$ , pour devenir  $g_{00}^{\text{Einstein}}(x) \simeq -1 + 2GM/c^2 r$ , où  $M$  est la masse de la Terre (dans notre exemple) et  $r$  la distance au centre de la Terre. Dans l'exemple considéré ci-dessus des jumeaux terrestres l'effet est très petit (une différence de vie d'environ une seconde sur 100 ans), mais l'effet est réel et a été vérifié maintes fois en considérant des horloges atomiques (voir références dans [4]). Mentionnons que cet « effet Einstein » a aujourd'hui d'importantes retombées pratiques, par exemple pour la navigation aérienne ou maritime, pour le pilotage des automobiles, ou même d'engins agricoles, etc. En effet le système GPS (Global Positioning System), qui utilise les données transmises par une constellation d'horloges atomiques embarquées sur des satellites, inclut dans son logiciel la déformation einsteinienne de la chronométrie de l'espace-temps. L'effet est seulement de l'ordre du milliardième, mais si on ne le prenait pas en compte, il introduirait dans le système GPS un dysfonctionnement inacceptablement grand, et augmentant continuellement avec le temps. En effet, les performances du système GPS reposent sur la haute stabilité des horloges embarquées, meilleure que  $10^{-13}$ , c'est-à-dire une stabilité 10 000 fois plus grande que le changement apparent de fréquence ( $\sim 10^{-9}$ ) dû à la déformation einsteinienne de la chrono-géométrie.

## 6.6 Limite des champs faibles et limite newtonienne

Pour comprendre les conséquences physiques des équations d'Einstein (6.9), il est utile de commencer en considérant le cas limite des champs géométrico-gravitationnels *faibles*, c'est-à-dire le cas où  $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$ , avec des quantités  $h_{\mu\nu}(x)$  très petites par rapport à l'unité :  $|h_{\mu\nu}(x)| \ll 1$ . Dans ce cas, un calcul simple (que nous encourageons le lecteur à faire) à partir des définitions (6.8) et (6.10) ci-dessus, conduit à la forme explicite suivante des équations d'Einstein (où l'on néglige les termes d'ordre  $h^2$  et  $hT$ ) :

$$\square h_{\mu\nu} - \partial_\mu \partial^\alpha h_{\alpha\nu} - \partial_\nu \partial^\alpha h_{\alpha\mu} + \partial_{\mu\nu} h^\alpha_\alpha = -\frac{16\pi G}{c^4} \tilde{T}_{\mu\nu} \quad (6.11)$$

où  $\square = \eta^{\mu\nu} \partial_{\mu\nu} = \Delta - \partial_0^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2 - c^{-2} \partial^2/\partial t^2$  désigne le d'Alembertien « plat » (opérateur des ondes ;  $x^\mu = (ct, x, y, z)$ ), et où les indices en position haute ont été élevés par l'inverse  $\eta^{\mu\nu}$  de la métrique plate  $\eta_{\mu\nu}$

(numériquement  $\eta^{\mu\nu} = \eta_{\mu\nu}$ , c'est-à-dire  $-\eta^{00} = \eta^{11} = \eta^{22} = \eta^{33} = +1$ ). Par exemple  $\partial^\alpha h_{\alpha\nu}$  désigne  $\eta^{\alpha\beta} \partial_\alpha h_{\beta\nu}$  et  $h_\alpha^\alpha \equiv \eta^{\alpha\beta} h_{\alpha\beta} = -h_{00} + h_{11} + h_{22} + h_{33}$ . La « source »  $\tilde{T}_{\mu\nu}$  apparaissant au membre de droite de (6.11) désigne (quand l'espace-temps est à quatre dimensions) la combinaison  $\tilde{T}_{\mu\nu} \equiv T_{\mu\nu} - \frac{1}{2} T_\alpha^\alpha \eta_{\mu\nu}$ .

L'approximation « linéarisée » (6.11) des équations d'Einstein est analogue aux équations de Maxwell

$$\square A_\mu - \partial_\mu \partial^\alpha A_\alpha = -4\pi J_\mu \quad (6.12)$$

reliant le quadri-potential électromagnétique  $A_\mu \equiv \eta_{\mu\nu} A^\nu$  (où  $A^0 = V$ ,  $A^i = \mathbf{A}$ ,  $i = 1, 2, 3$ ) à la densité de quadri-courant  $J_\mu \equiv \eta_{\mu\nu} J^\nu$  (où  $J^0 = \rho =$  densité de charge et  $J^i = \mathbf{J} =$  densité de courant). Une autre analogie est que la structure du membre de gauche des équations de Maxwell (6.12) implique que la « source »  $J_\mu$  apparaissant au membre de droite doit satisfaire  $\partial^\mu J_\mu = 0$  ( $\partial^\mu \equiv \eta^{\mu\nu} \partial_\nu$ ), qui traduit la conservation de la charge électrique. De même la structure du membre de gauche des équations d'Einstein linéarisées (6.11) implique que le « source »  $T_{\mu\nu} = \tilde{T}_{\mu\nu} - \frac{1}{2} \tilde{T}_\alpha^\alpha \eta_{\mu\nu}$  doit satisfaire  $\partial^\mu T_{\mu\nu} = 0$ , qui traduit la conservation de l'énergie et de l'impulsion de la matière. (La structure du membre de gauche des équations d'Einstein exactes (6.9) implique que la source  $T_{\mu\nu}$  doit satisfaire l'équation plus compliquée  $\partial_\mu T^{\mu\nu} + \Gamma_{\sigma\mu}^\nu T^{\sigma\nu} + \Gamma_{\sigma\mu}^\nu T^{\mu\sigma} = 0$ , où les termes en  $\Gamma T$  peuvent être interprétés comme décrivant un échange d'énergie et d'impulsion entre la matière et le champ gravitationnel.) La différence majeure est que, dans le cas de l'électromagnétisme, le champ  $A_\mu$  et sa source  $J_\mu$  ont un seul indice d'espace-temps, alors que dans le cas gravitationnel le champ  $h_{\mu\nu}$  et sa source  $\tilde{T}_{\mu\nu}$  ont deux indices d'espace-temps. Nous reviendrons plus loin sur cette analogie/différence entre  $A_\mu$  et  $h_{\mu\nu}$  qui suggère l'existence d'une certaine parenté entre gravitation et électromagnétisme.

On retrouve la théorie newtonienne de la gravitation comme cas limite de la théorie einsteinienne en supposant non seulement que le champ gravitationnel est une faible déformation de l'espace-temps plat de Minkowski ( $h_{\mu\nu} \ll 1$ ) mais aussi que le champ  $h_{\mu\nu}$  est lentement variable ( $\partial_0 h_{\mu\nu} \ll \partial_i h_{\mu\nu}$ ) et que sa source  $T_{\mu\nu}$  est non relativiste ( $T_{ij} \ll T_{0i} \ll T_{00}$ ). Sous ces conditions l'équation (6.11) conduit à une équation pour la composante purement temporelle,  $h_{00}$ , du champ espace-temps du type de celle de Poisson,

$$\Delta h_{00} = -\frac{16\pi G}{c^4} \tilde{T}_{00} = -\frac{8\pi G}{c^4} (T_{00} + T_{ii}) \simeq -\frac{8\pi G}{c^4} T_{00} \quad (6.13)$$

où  $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$  est le Laplacien. Rappelons que, d'après Laplace et Poisson, la théorie newtonienne de la gravitation se résume à dire que le champ gravitationnel est décrit par un seul potentiel  $U(x)$ , engendré par la densité volumique de masse  $\rho(x)$  selon l'équation de Poisson  $\Delta U = -4\pi G\rho$ , et déterminant l'accélération d'une particule d'épreuve plongée dans le champ

extérieur  $U(x)$  selon l'équation  $d^2x^i/dt^2 = \partial_i U(x) \equiv \partial U/\partial x^i$ . D'après  $m_i = m_g = E/c^2$  on peut identifier  $\rho = T^{00}/c^2$ . On voit alors que (6.13) reproduit l'équation de Poisson si  $h_{00} = +2U/c^2$ . Il reste donc à vérifier que la théorie d'Einstein prédit bien qu'une particule d'épreuve non relativiste est accélérée par un champ d'espace-temps selon  $d^2x^i/dt^2 \simeq \frac{1}{2}c^2 \partial_i h_{00}$ . Einstein comprit que c'était une conséquence du principe d'équivalence. En effet, tel qu'on l'a discuté au paragraphe 6.4, le principe d'équivalence dit que le champ gravitationnel est (localement) effacé dans un référentiel localement inertiel  $X^\alpha$  (où  $g_{\alpha\beta}(X) = \eta_{\alpha\beta} + \mathcal{O}((X - X_0)^2)$ ). Dans un tel référentiel, les lois de la relativité restreinte s'appliquent au point  $X_0$ . En particulier un corps isolé (et électriquement neutre) doit y satisfaire au principe d'inertie : son centre de masse se déplace en ligne droite à vitesse constante. Autrement dit, il satisfait à l'équation du mouvement  $d^2X^\alpha/ds^2 = 0$ . En repassant dans un système de coordonnées (étendu) quelconque  $x^\mu$ , on vérifie que cette équation du mouvement inertiel local se transforme dans l'équation des géodésiques (6.7). Donc (6.7) décrit la chute des corps, telle qu'elle est observée dans un référentiel étendu quelconque (par exemple un référentiel au repos par rapport à la Terre, ou au repos par rapport au barycentre du système solaire). On en conclut que l'analogue relativiste du champ newtonien d'accélération gravitationnelle,  $\mathbf{g}(x) = \nabla U(x)$ , est  $g^\lambda(x) \equiv -c^2 \Gamma_{\mu\nu}^\lambda dx^\mu/ds dx^\nu/ds$ . En considérant une particule dont le mouvement est lent par rapport à la vitesse de la lumière ( $dx^i/ds \ll dx^0/ds \simeq 1$ ) on vérifie facilement que  $g^i(x) \simeq -c^2 \Gamma_{00}^i$ . En utilisant enfin la définition (6.8) des  $\Gamma_{\mu\nu}^\alpha$ , et l'hypothèse des champs faibles, on vérifie bien que  $g^i(x) \simeq \frac{1}{2}c^2 \partial_i h_{00}$ , en parfait accord avec l'identification  $h_{00} = 2U/c^2$  anticipée ci-dessus. Nous encourageons le lecteur à vérifier par lui-même ce résultat qui contient l'essence même de la théorie d'Einstein : le mouvement gravitationnel n'est plus décrit comme étant dû à une force, mais est identifié au mouvement « le plus inertiel possible » dans un espace-temps dont la chrono-géométrie est déformée en présence d'une distribution d'énergie et d'impulsion.

Retrouver la théorie newtonienne comme cas limite de la théorie einsteinienne est évidemment une nécessité pour considérer sérieusement cette nouvelle théorie. Mais bien sûr, dès le début, Einstein a exploré les conséquences observationnelles de la relativité générale qui vont au-delà de la description newtonienne de la gravitation. Nous en avons déjà mentionné une ci-dessus : le fait que  $g_{00} = \eta_{00} + h_{00} \simeq -1 + 2U(x)/c^2$  implique une distortion des mesures relatives de temps au voisinage des corps massifs. Einstein prédit dès 1907 (c'est-à-dire dès qu'il conçut le principe d'équivalence, et bien avant d'avoir obtenu les équations de champ de la relativité générale) l'existence d'une telle distortion des mesures de temps et de fréquence en présence d'un champ gravitationnel extérieur. Il réalisa que cela aurait des conséquences observables pour la fréquence, observée sur Terre, des raies spectrales émises à la surface du Soleil. Spécifiquement, une raie spectrale de fréquence (propre locale)  $\nu_0$  émise en un point  $x_0$  où le potentiel gravitationnel (stationnaire) vaut  $U(x_0)$  et observée (via des signaux électromagnétiques) en un point  $x$  où le potentiel

vaut  $U(\mathbf{x})$  doit apparaître avoir la fréquence  $\nu$  telle que

$$\frac{\nu}{\nu_0} = \sqrt{\frac{g_{00}(x_0)}{g_{00}(x)}} \simeq 1 + \frac{1}{c^2} [U(\mathbf{x}) - U(\mathbf{x}_0)] \quad (6.14)$$

Dans le cas où le point d'émission  $x_0$  est dans un puits de potentiel gravitationnel plus profond que le point d'observation  $x$  (c'est-à-dire  $U(\mathbf{x}_0) > U(\mathbf{x})$ ) on a  $\nu < \nu_0$ , c'est-à-dire un effet de *rougissement* des fréquences. Cet effet, prédit par Einstein dès 1907, n'a été vérifié, de façon non ambiguë, qu'à partir des années 60 : expériences de Pound et collaborateurs sur une hauteur d'une vingtaine de mètres. La vérification la plus précise (niveau  $\sim 10^{-4}$ ) est due à Vessot et al., qui comparèrent un maser à hydrogène, embarqué sur une fusée ayant monté jusqu'à  $\sim 10\,000$  km d'altitude, à une horloge de construction similaire sur le sol. D'autres expériences comparèrent les temps lus par des horloges embarquées sur des avions à des horloges restées sur le sol. (Pour les références à ces expériences voir [4].) Comme on l'a déjà dit, l'« effet Einstein » (6.14) doit être incorporé de façon cruciale dans les logiciels des systèmes de positionnement par satellites comme GPS.

Einstein indiqua, aussi, dès 1907 que le principe d'équivalence suggérait que les rayons lumineux devaient être défléchis par un champ gravitationnel. En effet, une généralisation du raisonnement, fondé sur le principe d'équivalence, donné ci-dessus pour le mouvement des particules dans un champ gravitationnel extérieur montre que la lumière doit elle aussi suivre les mouvements « les plus inertiels possibles », c'est-à-dire des géodésiques de l'espace-temps courbe. Les rayons lumineux doivent donc satisfaire à l'équation des géodésiques (6.7). (La seule différence avec les géodésiques suivies par les particules matérielles est que le paramètre  $s$  de l'équation (6.7) ne peut plus être pris égal à la « longueur » le long de la géodésique, car une géodésique « de lumière » doit aussi satisfaire la contrainte  $g_{\mu\nu}(x) dx^\mu dx^\nu = 0$  exprimant que sa vitesse est égale à  $c$ , quand elle est mesurée dans un référentiel localement inertiel.) En partant de l'équation (6.7) on peut donc calculer quelle est la déflexion de la lumière quand elle passe au voisinage d'une masse importante (comme le Soleil). Pour faire ce calcul, on s'aperçoit cependant rapidement qu'il faut connaître plus que la composante  $h_{00}$  du champ gravitationnel. Les autres composantes de  $h_{\mu\nu}$ , et notamment les composantes spatiales  $h_{ij}$  interviennent de façon cruciale dans le calcul. C'est pourquoi c'est seulement en novembre 1915, après avoir obtenu la forme (essentiellement) finale de sa théorie, qu'Einstein put prédire la valeur complète de la déflexion de la lumière par le Soleil. En partant des équations d'Einstein linéarisées (6.11), et en continuant à faire les simplifications « non relativistes » indiquées ci-dessus ( $T_{ij} \ll T_{0i} \ll T_{00}$ ;  $\partial_0 h \ll \partial_i h$ ), il est facile de voir que la composante spatiale  $h_{ij}$  peut, comme  $h_{00}$ , s'écrire (après un choix commode de coordonnées) en termes du potentiel newtonien  $U$  selon  $h_{ij}(x) \simeq +2U(x)\delta_{ij}/c^2$ , où  $\delta_{ij}$  vaut 1 si  $i = j$  et 0 sinon ( $i, j = 1, 2, 3$ ). En insérant ce résultat, ainsi que le résultat précédent  $h_{00} = +2U/c^2$ , dans l'équation géodésique (6.7) pour le

mouvement de la lumière, on trouve (comme le fit Einstein en 1915) que la relativité générale prédit que le Soleil doit défléchir un rayon lumineux d'un angle  $\theta = 4GM/(c^2b)$  où  $b$  est le paramètre d'impact du rayon (c'est-à-dire sa distance minimale au Soleil). Comme l'on sait, la confirmation (avec une faible précision) de cet effet en 1919 rendit célèbre la théorie de la relativité générale et son auteur.

## 6.7 Approximation post-newtonienne et confirmations expérimentales dans le régime des champs gravitationnels faibles et quasi-stationnaires

On a déjà indiqué ci-dessus quelques confirmations expérimentales de la théorie de la relativité générale. À l'heure actuelle, la très grande précision de certaines mesures de durée ou de fréquence dans le système solaire nécessite de tenir compte très soigneusement des modifications que la relativité générale apporte à la description newtonienne de l'espace-temps. Par conséquent, la relativité générale est utilisée dans un grand nombre de situations, depuis la recherche astronomique ou géophysique (interférométrie radio à très longue base, poursuite radar des planètes, poursuite laser de la Lune ou de satellites artificiels) jusqu'aux applications météorologiques, géodésiques ou autres (définition du temps atomique international, cartographie de précision, système GPS). Pour ce faire, une méthode d'approximation, dite *post-newtonienne*, a été développée. Cette méthode consiste à compléter la limite newtonienne esquissée ci-dessus en gardant les termes supérieurs dans le petit paramètre ( $v$  désignant une vitesse caractéristique des éléments du système considéré)

$$\varepsilon \sim \frac{v^2}{c^2} \sim |h_{\mu\nu}| \sim |\partial_0 h / \partial_i h|^2 \sim |T^{0i}/T^{00}|^2 \sim |T^{ij}/T^{00}|$$

Pour toutes les applications présentes de la relativité générale dans le système solaire, il suffit d'inclure la *première approximation post-newtonienne*, c'est-à-dire de garder les corrections relatives d'ordre  $\varepsilon$  aux prédictions newtoniennes. Comme pendant longtemps la théorie de la relativité générale était mal vérifiée on a trouvé utile (selon les travaux pionniers d'A. Eddington, généralisés dans les années soixante par K. Nordtvedt et C.M. Will) d'étudier non seulement les prédictions précises des équations (6.9) définissant la théorie d'Einstein, mais de considérer aussi des déviations possibles à ces prédictions. Ces déviations possibles furent paramétrisées au moyen de plusieurs paramètres (non dimensionnés) dits « post-newtoniens ». Parmi ces paramètres, deux jouent un rôle clé :  $\gamma$  et  $\beta$ . Le paramètre  $\gamma$  paramétrise une déviation possible de la relativité générale intervenant dès le niveau linéarisé, c'est-à-dire modifiant l'approximation linéarisée donnée ci-dessus. Plus préci-

sément, il est défini en écrivant que l'écart  $h_{ij} \equiv g_{ij} - \delta_{ij}$  entre la métrique d'espace et la métrique euclidienne pourrait valoir (dans un système de coordonnées adéquat), non pas la valeur  $h_{ij}^{\text{GR}} = 2U \delta_{ij}/c^2$  qu'il prend en relativité générale, mais la valeur  $h_{ij} = 2\gamma U \delta_{ij}/c^2$ , différant d'un facteur  $\gamma$ . Donc, par définition  $\gamma$  vaut 1 en relativité générale, et  $\gamma - 1$  mesure une déviation possible par rapport à cette théorie. Quant au paramètre  $\beta$  (ou plutôt  $\beta - 1$ ) il mesure une déviation possible (par rapport à la relativité générale) dans la valeur de  $h_{00} \equiv g_{00} - \eta_{00}$ . La valeur de  $h_{00}$  en relativité générale est  $h_{00}^{\text{GR}} = 2U/c^2 - 2U^2/c^4$ , où le premier terme (discuté ci-dessus) reproduit l'approximation newtonienne (et ne peut donc pas être modifié si l'idée est de paramétriser la physique gravitationnelle au-delà des prédictions newtoniennes) et où le deuxième terme est obtenu en résolvant les équations d'Einstein (6.9) au deuxième ordre d'approximation. On écrit alors un  $h_{00}$  d'un type paramétrisé plus général,  $h_{00} = 2U/c^2 - 2\beta U^2/c^4$ , où, par définition,  $\beta$  vaut 1 en relativité générale. Signalons enfin que les paramètres  $\gamma - 1$  et  $\beta - 1$  paramétrisent complètement le régime post-newtonien des alternatives théoriques les plus simples à la relativité générale, c'est-à-dire les théories tenseur-scalaires de la gravitation. Dans ces théories, l'interaction gravitationnelle est transportée par deux champs à la fois : un champ tensoriel (spin 2) de masse nulle couplé à  $T^{\mu\nu}$ , et un champ scalaire  $\varphi$  (spin 0) de masse nulle couplé à la trace  $T_\alpha^\alpha$ . Le paramètre  $-(\gamma - 1)$  joue alors le rôle clé de mesurer le rapport entre le couplage scalaire et le couplage tensoriel.

Toutes les expériences réalisées à ce jour dans le système solaire sont compatibles avec les prédictions de la relativité générale. Quand elles sont interprétées en termes des paramètres post-newtoniens (et « post-einsteinien »)  $\gamma - 1$  et  $\beta - 1$ , elles conduisent à de fortes contraintes sur les déviations possibles de la théorie d'Einstein. Citons parmi les tests réalisés dans le système solaire, la déflexion d'ondes électromagnétiques au voisinage du soleil, le retard gravitationnel (effet Shapiro) d'échos radars sur la station Viking posée sur Mars, l'analyse globale de la dynamique du système solaire (incluant l'avance des périhélie planétaires), la mesure sub-centimétrique de la distance Terre-Lune obtenue à partir d'échos lasers sur les réflecteurs déposés sur la Lune, etc. À l'heure actuelle (2004) le test le plus précis de la relativité générale a été obtenu en 2003 en mesurant le rapport  $1 + y \equiv f/f_0$  entre la fréquence  $f_0$  d'ondes radio envoyées depuis la Terre vers la sonde spatiale Cassini et la fréquence  $f$  d'ondes radio cohérentes renvoyées (avec la même fréquence locale) par Cassini vers la Terre et comparées (sur Terre) à la fréquence d'émission  $f_0$ . La contribution principale à la petite quantité  $y$  est un effet égal, en relativité générale, à  $y_{\text{GR}} = 8(GM/c^3 b) db/dt$  (où  $b$  est, comme ci-dessus, le paramètre d'impact) dû à la propagation des ondes radios dans la géométrie d'espace-temps déformée par le Soleil :  $ds^2 \simeq -(1 - 2U/c^2) c^2 dt^2 + (1 + 2U/c^2)(dx^2 + dy^2 + dz^2)$ , où  $U = GM/r$ . La valeur maximale du changement de fréquence prédit par la relativité générale était seulement  $|y_{\text{GR}}| \lesssim 2 \times 10^{-10}$  pour les meilleures observations, mais grâce à une excellente stabilité de fréquence  $\sim 10^{-14}$  (après correction des perturbations causées par la couronne solaire) et à un relativement



grand nombre de mesures individuelles réparties sur 18 jours, cette expérience a pu vérifier la théorie d'Einstein au niveau remarquable de  $\sim 10^{-5}$  [7]. Plus précisément, quand cette expérience est interprétée en termes des paramètres post-newtoniens  $\gamma - 1$  et  $\beta - 1$ , elle donne la limite suivante sur le paramètre  $\gamma - 1$  [7]

$$\gamma - 1 = (2,1 \pm 2,3) \times 10^{-5} \quad (6.15)$$

Quant à la meilleure limite actuelle sur le paramètre  $\beta - 1$ , elle est meilleure que  $10^{-3}$  et provient de la non-observation, dans les données des échos lasers sur la Lune, d'une éventuelle polarisation de l'orbite de la Lune en direction du Soleil (effet Nordtvedt ; voir [4] pour les références)

$$4(\beta - 1) - (\gamma - 1) = -0,0007 \pm 0,0010 \quad (6.16)$$

Bien que la théorie de la relativité générale soit une des mieux vérifiées de la physique, les scientifiques continuent de concevoir et de planifier des tests nouveaux, ou encore plus fins, de cette théorie. C'est en particulier le cas de la mission spatiale Gravity Probe B (envoyée par la NASA en avril 2004) dont le but principal est d'observer directement une prédiction de la relativité générale qui dit (intuitivement) que l'espace est non seulement « élastique », mais aussi « fluide ». Au dix-neuvième siècle Foucault avait inventé à la fois le gyroscope et son célèbre pendule pour rendre directement observable l'espace absolu (et rigide) de Newton. Ses expériences montraient en effet que, par exemple, un gyroscope à la surface de la Terre continuait, malgré la rotation de la Terre, à s'aligner avec une direction « fixe » par rapport aux étoiles lointaines. Or, quand Lense et Thirring analysèrent, en 1918, certaines des conséquences des équations (linéarisées) d'Einstein (6.11), ils trouvèrent que la relativité générale prédit, entre autres, le phénomène suivant : la rotation de la Terre (ou de n'importe quelle boule de matière) crée une déformation particulière de la chrono-géométrie de l'espace-temps. Cette déformation est décrite par les composantes « gravito-magnétiques »  $h_{0i}$  de la métrique, et induit un effet analogue à l'effet d'« entraînement en rotation » que cause une boule de matière tournant dans un fluide : la rotation de la Terre entraîne, d'une façon minime, tout l'espace autour d'elle à « tourner » continuellement comme le ferait un fluide. Cette « rotation de l'espace » se traduit, de façon observable, par une violation des effets prévus par Newton et confirmés par les expériences de Foucault : en particulier, un gyroscope ne s'aligne plus avec une direction « fixe dans l'espace absolu », mais son axe de rotation est « entraîné » par le mouvement de rotation de l'espace local où il se trouve. Cet effet était bien trop petit pour être visible dans les expériences de Foucault. Son observation par Gravity Probe B [8] est importante pour rendre sensible au plus grand nombre la notion einsteinienne révolutionnaire d'un espace-temps fluide.

Jusqu'ici nous n'avons discuté que le régime des champs gravitationnels faibles et lentement variables. La théorie de la relativité générale prévoit l'apparition de nouveaux phénomènes quand le champ gravitationnel devient fort

et/ou rapidement variable. (Nous ne discuterons pas ici les aspects cosmologiques de la gravitation relativiste ; voir chapitre 7.)

## 6.8 Champs gravitationnels forts. Trous noirs

Le régime des champs gravitationnels forts se rencontre dans la physique des *corps gravitationnellement condensés*. Ce vocable désigne les états finals de l'évolution des étoiles, et tout particulièrement les étoiles à neutrons et les trous noirs. Rappelons que la plus grande partie de la vie d'une étoile se passe à brûler, de façon lente, son carburant nucléaire. Ce processus conduit à structurer l'étoile en une suite de couches de composition nucléaire différenciée, entourant un cœur de plus en plus dense (structure « en oignon »). Quand la masse initiale de l'étoile est suffisamment grande, ce processus finit par conduire à un phénomène catastrophique : le cœur, déjà beaucoup plus dense que la matière ordinaire, s'effondre sur lui-même, sous l'influence de son auto-attraction gravitationnelle. (Cette implosion de la partie centrale de l'étoile est accompagnée, dans beaucoup de cas, par une explosion des couches extérieures de l'étoile (phénomène de supernova).) Selon la masse qui s'effondre avec le cœur de l'étoile, cet effondrement peut donner lieu soit à la formation d'une étoile à neutrons, soit à celle d'un trou noir.

Une *étoile à neutrons* condense une masse de l'ordre de la masse du Soleil dans un rayon de l'ordre de 10 km. La densité à l'intérieur d'une étoile à neutrons (ainsi appelée car les neutrons dominent dans sa composition nucléaire) dépasse cent millions de tonnes par centimètre cube ( $10^{14}$  g/cm<sup>3</sup>) ! Elle est de l'ordre de grandeur de la densité à l'intérieur des noyaux des atomes. Ce qui est important pour notre propos est que la déformation de la métrique de Minkowski au voisinage immédiat d'une étoile à neutrons, mesurée par  $h_{00} \sim h_{ii} \sim 2GM/c^2R$ , où  $R$  est le rayon de l'étoile, n'est plus du tout une petite quantité, comme elle l'était dans le système solaire. En effet, alors que  $h \sim 2GM/c^2R$  était de l'ordre de  $10^{-9}$  pour la Terre, et de  $10^{-6}$  pour le Soleil, on trouve que  $h \sim 0,4$  pour une étoile à neutrons typique ( $M \simeq 1,4 M_{\odot}$ ,  $R \sim 10$  km). On en conclut qu'il n'est plus possible, comme c'était le cas dans le système solaire, d'étudier la structure et la physique des étoiles à neutrons en utilisant l'approximation post-newtonienne évoquée ci-dessus. Il faut considérer les équations d'Einstein exactes (6.9), avec toute leur structure non-linéaire. À cause de ceci, on s'attend à ce que des observations concernant des étoiles à neutrons puissent nous permettre de confirmer (ou d'infirmer) la théorie de la relativité générale dans son régime fortement non-linéaire. Nous discuterons de tels tests paragraphe 9 à propos des observations de pulsars binaires.

Un *trou noir* est le résultat d'un effondrement *continué*, c'est-à-dire qui ne s'est pas arrêté à la formation d'une étoile ultra-dense (comme une étoile à neutron). (Le concept physique de trou noir a été introduit par J.R. Oppenheimer et H. Snyder en 1939. La structure géométrique globale des trous noirs

n'a été comprise que des années plus tard, notamment grâce aux travaux de R. Penrose. Pour une revue historique du concept de trou noir voir [9].) C'est une structure particulière d'espace-temps courbe caractérisée par l'existence d'une frontière (dite « surface du trou noir » ou « horizon ») entre une région extérieure, d'où il est possible d'émettre des signaux à l'infini, et une région (d'espace-temps) intérieure, où tout signal émis reste piégé (Fig. 6.3).

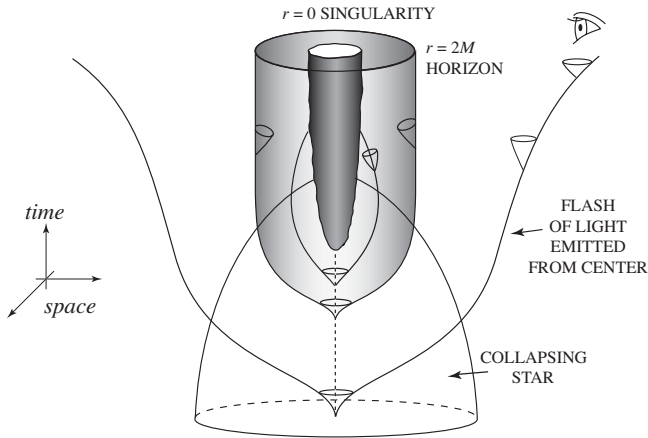


FIG. 6.3 – Représentation schématique de l'espace-temps d'un trou noir engendré par l'effondrement d'une étoile sphérique. Chaque cône représente l'histoire spatio-temporelle d'un éclair lumineux émis à partir d'un point à un certain instant. (Un tel « champ de cônes » est obtenu en prenant la limite  $\varepsilon^2 = 0$  de la figure 6.2, et en ne gardant que la partie supérieure, c'est-à-dire dirigée vers le futur, des doubles cônes obtenus comme limites des hyperboloïdes de la figure 6.2.) L'intérieur du trou noir est indiqué en grisé, sa frontière extérieure étant la « surface du trou noir » ou « horizon ». La « frontière intérieure » (indiquée en grisé foncé) de la région intérieure du trou noir est une singularité de l'espace-temps du type big-crunch.

Les cônes indiqués sur cette figure sont ce qu'on appelle des « cônes de lumière ». Ils sont définis comme le lieu des points (infiniment voisins de  $x$ ) tels que  $ds^2 = 0$ , avec  $dx^0 = cdt \geq 0$ . Chacun représente le début de l'histoire spatio-temporelle d'un éclair lumineux émis d'un certain point d'espace-temps. Les cônes dont le sommet est situé à l'extérieur de l'horizon (zone claire) vont évoluer en s'évasant jusqu'à l'infini, représentant ainsi la possibilité pour des signaux électromagnétiques d'atteindre l'infini.

En revanche, les cônes dont le sommet est situé à l'intérieur de l'horizon (zone grisée) vont évoluer sans jamais réussir à sortir de la zone grisée. Il est donc impossible d'émettre, depuis la zone grisée, un signal électromagnétique atteignant l'infini. L'horizon, c'est-à-dire la frontière entre la zone grisée et la zone claire, est, quant à lui, l'histoire d'un éclair lumineux particulier, émis du centre de l'étoile au cours de l'effondrement, et tel qu'il se stabilise

asymptotiquement en un cylindre d'espace-temps. Ce cylindre d'espace-temps (horizon asymptotique) représente donc l'histoire spatio-temporelle d'une bulle de lumière qui, vue localement, se déplace vers l'extérieur à la vitesse  $c$ , mais qui, globalement, fait du « sur-place ». Ce comportement remarquable est une illustration frappante du caractère « fluide » de l'espace-temps einsteinien. En effet, on pourrait comparer la situation précédente à ce qui peut se passer autour de la bonde ouverte d'un évier en train de se vider : une onde peut se déplacer, vers l'extérieur, sur l'eau tout en faisant du sur-place par rapport à l'évier à cause du mouvement de chute de l'eau en direction de la bonde.

Notons que le développement temporel de la région intérieure est limité, se terminant (surface gris foncé) par une *singularité* où la courbure devient infinie et où la description classique de l'espace et du temps perd son sens. Cette singularité est localement semblable à l'inverse temporel d'une singularité cosmologique du type big bang. On l'appelle un *big crunch*. C'est un bord de l'espace-temps, au-delà duquel l'espace-temps cesse d'exister. L'apparition de singularités associées à des régions de champ gravitationnel fort est un phénomène générique en relativité générale, comme le montrent des théorèmes dus à R. Penrose et S.W. Hawking.

Les trous noirs ont des propriétés remarquables. D'abord, un théorème d'*unicité* (dû à W. Israel, B. Carter, D.C. Robinson, G. Bunting et P.O. Mazur) affirme qu'un trou noir stationnaire isolé (en théorie d'Einstein-Maxwell) est complètement décrit par trois paramètres : sa masse  $M$ , son moment cinétique  $J$  et sa charge électrique  $Q$ . La solution exacte (dite de Kerr-Newman) des équations d'Einstein (6.11) décrivant un trou noir de paramètres  $M, J, Q$  est connue explicitement. Contentons nous d'écrire la géométrie d'espace-temps du cas particulier le plus simple de trou noir : celui où  $J = Q = 0$  et où le trou noir n'est décrit que par sa masse (solution découverte par K. Schwarzschild dès janvier 1916) :

$$ds^2 = - \left( 1 - \frac{2GM}{c^2 r} \right) c^2 dt^2 + \frac{dr^2}{1 - \frac{2GM}{c^2 r}} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \quad (6.17)$$

On voit que la composante purement temporelle de la métrique,  $g_{00} = -(1 - 2GM/c^2 r)$ , s'annule quand la coordonnée radiale  $r$  vaut  $r = r_H \equiv 2GM/c^2$ . D'après la première équation (6.14), il semblerait donc que la lumière émise d'un point quelconque de la sphère  $r_0 = r_H$  subisse, quand elle est vue par un observateur situé n'importe où à l'extérieur (en  $r > r_H$ ), un rougissement infini de sa fréquence d'émission ( $\nu/\nu_0 = 0$ ). En fait, la sphère  $r_H = 2GM/c^2$  est l'*horizon* du trou noir de Schwarzschild, et aucune particule (susceptible d'émettre de la lumière) ne peut rester au repos en  $r = r_H$  (ni a fortiori en  $r < r_H$ ). Pour étudier ce qui se passe sur l'horizon ( $r = r_H$ ), ou à l'intérieur d'un trou noir de Schwarzschild ( $r < r_H$ ), il faut utiliser d'autres coordonnées d'espace-temps que les coordonnées  $(t, r, \theta, \varphi)$  utilisées dans l'équation (6.17). La singularité de « big crunch » à l'intérieur d'un trou noir de Schwarzschild est située, dans les coordonnées de (6.17), en  $r = 0$  (qui ne désigne pas,

comme on pourrait le croire, un point dans l'espace, mais bien un instant dans le temps).

La métrique d'espace-temps d'un trou noir, comme l'équation (6.17) dans le cas simple où  $J = Q = 0$ , permet d'étudier l'influence d'un trou noir sur les particules et les champs présents dans son voisinage. On trouve ainsi qu'un trou noir est un puits de potentiel gravitationnel si profond que toute particule, ou onde, qui pénètre à l'intérieur du trou noir (dans la région  $r < r_H$ ) ne peut plus jamais ressortir du trou noir, et que l'énergie totale de la particule ou onde qui « tombe » dans le trou noir vient s'ajouter à la masse-énergie totale  $M$  du trou noir. En étudiant (à la suite de R. Penrose) de tels processus d'« accréation » d'un trou noir par chute de particules, D. Christodoulou et R. Ruffini montrèrent qu'un trou noir est non seulement un puits de potentiel, mais aussi un objet physique possédant une *énergie libre* importante qu'il est possible, en principe, d'extraire. L'*énergétique des trous noirs* est résumée dans la « formule de masse » de Christodoulou et Ruffini (en unités où  $c = 1$ )

$$M^2 = \left( M_{\text{irr}} + \frac{Q^2}{4GM_{\text{irr}}} \right)^2 + \frac{J^2}{4G^2M_{\text{irr}}^2} \quad (6.18)$$

où  $M_{\text{irr}}$  désigne la *masse irréductible* du trou noir, une quantité *qui ne peut qu'augmenter de manière irréversible*. On déduit de (6.18) qu'un trou noir tournant ( $J \neq 0$ ) et/ou chargé ( $Q \neq 0$ ) possède une énergie libre  $M - M_{\text{irr}} > 0$  qui peut, en principe, être extraite par les processus réduisant son moment cinétique et/ou sa charge électrique. De tels processus d'extraction d'énergie de trous noirs pourraient être à l'origine de certains phénomènes astrophysiques ultra-énergétiques (comme les quasars ou les « gamma ray bursts »). Notons que, selon l'équation (6.18), les trous noirs (tournants ou chargés) sont les plus grands réservoirs d'énergie libre de l'Univers : en effet, 29 % de leur énergie de masse peut être stockée sous forme d'énergie cinétique de rotation, et jusqu'à 50 % peut être stockée sous forme d'énergie électrique. Ces pourcentages sont très supérieurs aux quelques pour cent d'énergie de liaison nucléaire qui sont à l'origine de toute la lumière émise par les étoiles pendant leur vie. Bien qu'il n'y ait pas, à ce jour, de preuve irréfutable de l'existence de trous noirs dans l'Univers, tout un faisceau de présomptions très fortes militent en faveur de leur existence. En particulier, plus d'une douzaine de systèmes binaires, émetteurs de rayons X, de notre Galaxie sont probablement constitués d'un trou noir et d'une étoile ordinaire. En outre, le centre de notre Galaxie semble contenir une concentration très compacte de masse  $\sim 3 \times 10^6 M_{\odot}$  qui est probablement un trou noir. (Pour une revue des données observationnelles conduisant à ces conclusions voir par exemple le paragraphe 7.6 du livre récent de N. Straumann [6].)

Le fait qu'une quantité associée à un trou noir, ici la masse irréductible  $M_{\text{irr}}$ , ou, selon un résultat plus général dû à S.W. Hawking l'aire totale  $A$  de la surface d'un trou noir ( $A = 16 \pi G^2 M_{\text{irr}}^2$ ), ne puisse évoluer qu'en augmentant d'une façon irréversible, est réminiscent de la deuxième loi de la thermodynamique. Ce résultat a suggéré à J.D. Bekenstein d'interpréter l'aire de

l'horizon,  $A$ , comme étant proportionnelle à l'*entropie* du trou noir. Une telle interprétation thermodynamique est renforcée par l'étude de l'accroissement de  $A$  sous l'influence de perturbations externes, accroissement que l'on peut en effet attribuer à des propriétés dissipatives locales de la surface du trou noir : notamment une viscosité surfacique et une résistivité électrique surfacique égale à 377 ohms (selon les travaux de T. Damour et R.L. Znajek). Ces interprétations « thermodynamiques » des propriétés des trous noirs restent de simples analogies au niveau de la physique classique, mais un résultat remarquable de Hawking a montré qu'elles avaient un contenu réel au niveau de la physique quantique. En 1974, Hawking découvrit que la présence d'un horizon dans l'espace-temps d'un trou noir affectait la définition d'une particule quantique, et conduisait un trou noir à émettre, de façon continue, un flux de particules ayant le spectre caractéristique (spectre de Planck) d'une émission thermique à la température  $T = 4 \hbar G \partial M / \partial A$ , où  $\hbar$  désigne la constante de Planck réduite. En utilisant la relation thermodynamique générale reliant la température à l'énergie  $E = M$  et à l'entropie  $S$ ,  $T = \partial M / \partial S$ , on voit que le résultat de Hawking suggère (en conformité avec les idées de Bekenstein) qu'un trou noir possède une *entropie*  $S$  égale à (toujours avec  $c = 1$ )

$$S = \frac{1}{4} \frac{A}{\hbar G} \quad (6.19)$$

La formule de Bekenstein-Hawking (6.19) suggère une connexion inattendue, et peut être profonde, entre gravitation, thermodynamique et théorie quantique (voir paragraphe 6.11).

## 6.9 Pulsars binaires et confirmations expérimentales du régime des champs gravitationnels forts et radiatifs

Les « pulsars binaires » sont des systèmes binaires constitués d'un pulsar (étoile à neutrons en rotation rapide sur elle-même) et d'une étoile compagnon très dense (étoile à neutrons ou naine blanche). Le premier système de ce type (appelé PSR B1913+16) a été découvert par R.A. Hulse et J.H. Taylor en 1974 [10]. On en connaît aujourd'hui une dizaine. Certains d'entre eux (et notamment le premier découvert PSR B1913+16) se sont révélés être des sondes remarquables de la gravitation relativiste, et, notamment, du régime des champs gravitationnels forts et/ou radiatifs. La raison pour laquelle un pulsar binaire permet de sonder la physique des champs gravitationnels forts est que, comme on l'a déjà indiqué ci-dessus, la déformation de la géométrie de l'espace-temps au voisinage d'une étoile à neutrons n'est plus, comme dans le système solaire, une petite quantité, mais qu'elle est de l'ordre de l'unité :  $h_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu} \sim 2GM/c^2 R \sim 0,4$ . (Notons que cette valeur est seulement 2,5 fois plus petite que dans le cas extrême d'un trou noir pour

lequel  $2GM/c^2R = 1$ .) En outre, le fait que l'interaction gravitationnelle se propage à la vitesse de la lumière (comme l'indique la présence de l'opérateur des ondes,  $\square = \Delta - c^{-2}\partial^2/\partial t^2$  dans (6.11)) entre le pulsar et son compagnon, se trouve jouer un rôle observationnellement significatif dans certains pulsars binaires.

Indiquons comment les données observationnelles des pulsars binaires sont utilisées pour sonder le régime des champs gravitationnels forts ( $h_{\mu\nu}$  de l'ordre de l'unité) et/ou radiatifs (propagation d'effets à la vitesse  $c$ ). Essentiellement, un pulsar joue le rôle d'une *horloge* très stable. En effet, le « phénomène pulsar » est dû à la rotation d'un faisceau d'ondes électromagnétiques, créé au voisinage des deux pôles magnétiques d'une étoile à neutrons fortement magnétisée (champ magnétique de l'ordre de  $10^{12}$  gauss, c'est-à-dire  $10^{12}$  fois le champ magnétique terrestre). L'axe magnétique d'un pulsar n'étant pas aligné avec son axe de rotation, la rotation rapide du pulsar fait tourner en bloc la magnétosphère (proche) du pulsar et par là même le faisceau d'ondes électromagnétiques engendré près des pôles magnétiques. Le pulsar est donc analogue à un phare qui balaie l'espace d'un double faisceau d'ondes électromagnétiques (un par pôle). Comme pour un phare, on ne voit le pulsar depuis la Terre que quand l'un de ses faisceaux balaie la Terre engendrant ainsi une bouffée de bruit électromagnétique à chaque tour du pulsar sur lui-même (dans certains cas, on voit même, après chaque demi-tour, une bouffée secondaire liée à l'émission du deuxième pôle). On peut mesurer le temps d'arrivée sur Terre (du centre) de chaque bouffée de bruit électromagnétique. Les données observationnelles essentielles d'un pulsar sont ainsi constituées de la suite discrète régulière des *temps d'arrivée* sur Terre de ces bouffées ou « impulsions ». Cette suite est analogue aux signaux d'une horloge : tic, tic, tic, ... On trouve observationnellement que certains pulsars (et notamment ceux qui sont dans des systèmes binaires) définissent ainsi des horloges d'une stabilité comparable aux meilleures horloges atomiques [11]. Dans le cas d'un pulsar solitaire, la suite de ses temps d'arrivées est (essentiellement) une « suite arithmétique » régulière,  $T_N = aN + b$ , où  $N$  est un nombre entier qui est le numéro de l'impulsion considérée, et où  $a$  est égal à la période de rotation du pulsar sur lui-même. Dans le cas d'un pulsar binaire, la suite de ses temps d'arrivée est un signal beaucoup plus riche, disons,  $T_N = aN + b + \Delta_N$ , où  $\Delta_N$  mesure la déviation par rapport à une suite arithmétique régulière. Cette déviation est due (après soustraction d'effets non liés à la période orbitale du pulsar) à tout un ensemble d'effets physiques liés au mouvement orbital du pulsar autour de son compagnon, ou plus exactement, autour du centre de masse du système binaire. Certains de ces effets pourraient être prévus par une description purement *képlérienne* du mouvement du pulsar dans l'espace, et sont analogues à l'« effet Rømer » qui permit à ce dernier de déterminer pour la première fois la vitesse de la lumière à partir des temps d'arrivées sur Terre de signaux lumineux venant des satellites de Jupiter (les signaux lumineux venant d'un corps se déplaçant sur une orbite sont « en retard » du temps mis par la lumière pour traverser cette orbite et parvenir à la Terre).

D'autres effets ne peuvent être prévus et calculés qu'en utilisant une description *relativiste* soit du mouvement orbital du pulsar, soit de la propagation des signaux électromagnétiques entre le pulsar et la Terre. Par exemple, il faut tenir compte des faits suivants : (i) l'« horloge pulsar » se déplace à une vitesse importante (de l'ordre de  $300 \text{ km/s} \sim 10^{-3}c$ ) et est plongée dans le potentiel gravitationnel variable dû au compagnon ; (ii) l'orbite du pulsar n'est pas une simple ellipse képlérienne mais (en relativité générale) une orbite plus compliquée qui trace une rosace autour du centre de masse ; (iii) la propagation des signaux électromagnétiques entre le pulsar et la Terre s'effectue dans un espace-temps courbé à la fois par le pulsar et son compagnon, ce qui conduit à des effets de retard relativiste particuliers ; etc. La prise en compte des effets relativistes dans la description théorique des temps d'arrivée des signaux émis par des pulsars binaires conduit alors à écrire ce qu'on appelle une « formule de chronométrage ». Cette formule de chronométrage (due à T. Damour et N. Deruelle) permet essentiellement de paramétriser la suite des temps d'arrivée,  $T_N = aN + b + \Delta_N$ , c'est-à-dire de paramétriser  $\Delta_N$ , en fonction d'un ensemble de « paramètres phénoménologiques », qui comprennent non seulement des paramètres dits « képlériens » (comme la période orbitale  $P$ , la projection  $x_A = a_A \sin i$  du demi-grand axe de l'orbite du pulsar sur la ligne de visée, et l'excentricité  $e$ ), mais aussi des paramètres *post-képlériens* associés aux effets relativistes mentionnés ci-dessus. Par exemple, l'effet (i) discuté ci-dessus est paramétrisé par une quantité notée  $\gamma_T$  ; l'effet (ii) par (entre autres) les quantités  $\dot{\omega}$ ,  $\dot{P}$  ; l'effet (iii) par les quantités  $r$ ,  $s$  ; etc.

La façon dont les observations des pulsars binaires permet de tester les théories relativistes de la gravitation est alors la suivante. Un ajustage (de moindres carrés) entre les données observationnelles de chronométrage,  $\Delta_N^{\text{obs}}$ , et la formule théorique, paramétrisée de chronométrage,  $\Delta_N^{\text{th}}(P, x_A, e; \gamma_T, \dot{\omega}, \dot{P}, r, s)$ , permet de déterminer les valeurs observationnelles des paramètres képlériens ( $P^{\text{obs}}, x_A^{\text{obs}}, e^{\text{obs}}$ ) et post-képlériens ( $\gamma_T^{\text{obs}}, \dot{\omega}^{\text{obs}}, \dot{P}^{\text{obs}}, r^{\text{obs}}, s^{\text{obs}}$ ). La théorie de la relativité générale prédit la valeur de chaque paramètre post-képlérien en fonction des paramètres képlériens, et des deux masses du système binaire : la masse  $m_A$  du pulsar, et la masse  $m_B$  du compagnon. Par exemple, la valeur théorique prédite par la relativité générale pour le paramètre  $\gamma_T$  est  $\gamma_T^{\text{GR}}(m_A, m_B) = en^{-1}(GMn/c^3)^{2/3} m_B(m_A + 2m_B)/M^2$ , où  $e$  est l'excentricité,  $n = 2\pi/P$  la fréquence orbitale, et  $M \equiv m_A + m_B$ . On voit alors que la mesure observationnelle d'un paramètre post-képlérien, par exemple  $\gamma_T^{\text{obs}}$ , détermine, si l'on suppose que la relativité générale est correcte, une *courbe* dans le plan  $(m_A, m_B)$  des deux masses :  $\gamma_T^{\text{GR}}(m_A, m_B) = \gamma_T^{\text{obs}}$ , dans notre exemple. La mesure de deux paramètres post-képlériens donne ainsi deux courbes dans le plan  $(m_A, m_B)$  et permet, génériquement, de déterminer les valeurs des deux masses  $m_A$  et  $m_B$ , en considérant l'intersection des deux courbes. On obtient un test de la relativité générale à partir du moment où l'on mesure observationnellement trois, ou plus, paramètres post-képlériens : si les trois (ou plus) courbes s'intersectent en un point du plan des masses, la théorie de la relativité générale est



confirmée, mais si ça n'est pas le cas la théorie est infirmée. À l'heure actuelle, quatre pulsars binaires distincts ont permis de tester la relativité générale. Ces quatre pulsars binaires « relativistes » sont : le premier pulsar binaire PSR B1913+16, le pulsar PSR B1534+12 (découvert par A. Wolszczan en 1991), et deux pulsars découverts récemment : PSR J1141–6545 (découvert en 1999 par V.M. Kaspi et al. et dont les premiers résultats de chronométrage sont dus à M. Bailes et al. en 2003), et PSR J0737–3039 (découvert en 2003 par M. Burgay et al., et dont les premiers résultats de chronométrage sont dus à A.G. Lyne et al. et M. Kramer et al.). À l'exception de PSR J1141–6545 dont le compagnon est une naine blanche, les compagnons des trois autres pulsars sont des étoiles à neutrons. Dans le cas de PSR J0737–3039 le compagnon se trouve être aussi un pulsar visible depuis la Terre.

Dans le système PSR B1913+16 on a mesuré *trois* paramètres post-képlériens ( $\dot{\omega}, \gamma_T, \dot{P}$ ), ce qui donne *un* test de la théorie. Dans le système PSR J1141–65 on a mesuré *quatre* paramètres post-képlériens ( $\dot{\omega}, \gamma_T, \dot{P}, s$ ), ce qui donne *deux* tests de la théorie. (Le paramètre  $s$  est mesuré à la fois par le chronométrage et (avec plus de précision) par des phénomènes de scintillation.) Dans les deux autres (PSR B1534+12 et PSR J0737–3039) on a mesuré jusqu'à *cinq* paramètres post-képlériens<sup>4</sup>, de sorte que chacun de ces pulsars donne *trois* tests de la théorie. Il est remarquable que tous ces tests ont confirmé la relativité générale (voir Fig. 6.4, et, pour les références et les détails [4, 11–13]).

Notez que, dans la figure 6.4, certains paramètres post-képlériens sont mesurés avec une précision si élevée qu'ils définissent en effet des courbes très minces dans le plan  $m_A, m_B$ . En revanche, certains d'entre eux ne sont mesurés qu'avec une précision fractionnaire peu élevée et définissent alors des « courbes épaisses », c'est-à-dire des « bandes » dans le plan des masses (voir, par exemple, les bandes associées à  $\dot{P}$ ,  $r$  et  $s$  dans le cas de PSR B1534+12). Dans tous les cas, la théorie est confirmée quant toutes les bandes (épaisses ou minces) ont une intersection commune non vide. (On doit aussi noter que les bandes représentées dans la figure 6.4 n'utilisent que les barres d'erreur « à un sigma », c'est-à-dire à 68 % de niveau de confiance. Donc, le fait que la bande  $\dot{P}$  dans PSR B1534+12 est un petit peu disjointe de l'intersection des autres bandes, n'est pas significatif : une figure « à deux sigmas » montrerait un excellent accord entre l'observation et la relativité générale.)

Au vu des arguments présentés ci-dessus, tous les tests représentés dans la figure 6.4 confirment la validité de la relativité générale dans le régime des champs gravitationnels forts ( $h_{\mu\nu} \sim 1$ ). En outre, les trois<sup>5</sup> tests utilisant des mesures du paramètre  $\dot{P}$  (dans PSR B1913+16, PSR B1534+12 et PSR J1141–6545) sont des confirmations expérimentales directes du fait que l'interaction gravitationnelle se propage à la vitesse  $c$  entre le compagnon et le

<sup>4</sup>Dans le cas de PSR J0737–3039, l'un des cinq paramètres mesurés est le rapport  $x_A/x_B$  entre un paramètre képlérien du pulsar et son analogue pour le compagnon, qui se trouve être aussi un pulsar.

<sup>5</sup>PSR J0737–3039 devrait bientôt permettre aussi une mesure du paramètre  $\dot{P}$ .

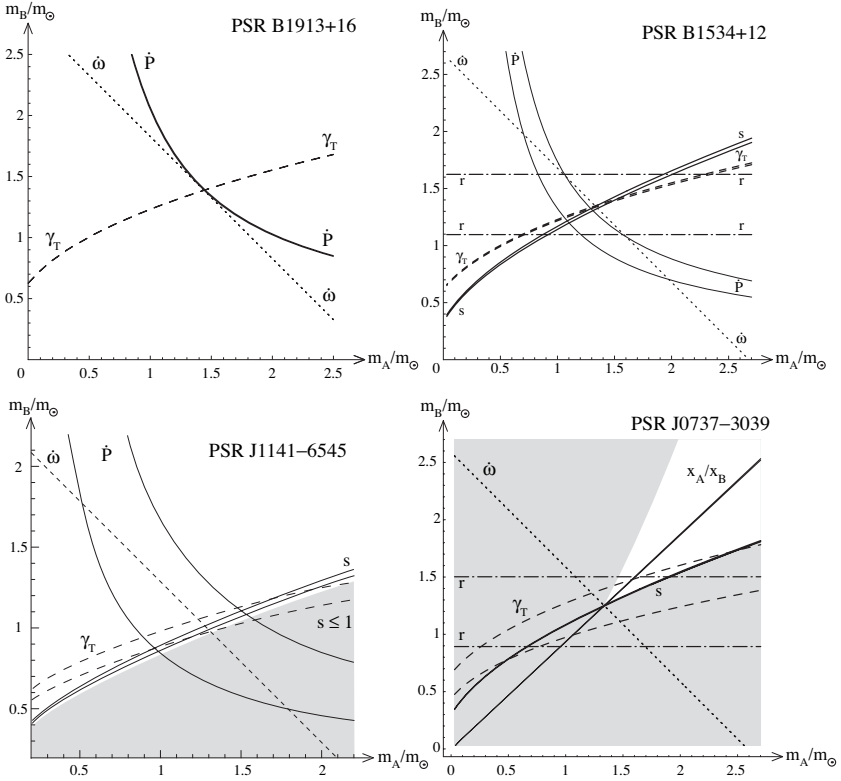


FIG. 6.4 – Tests de la relativité générale obtenus à partir des observations de quatre pulsars binaires. Pour chaque pulsar binaire on a tracé les « courbes », dans le plan des deux masses ( $m_A =$  masse du pulsar,  $m_B =$  masse du compagnon), définies en égalant l'expression théorique, prédite par la relativité générale, de divers paramètres post-képlériens à leur valeur observationnelle, déterminée par ajustage de moindres carrés à la formule théorique paramétrisée de chronométrage. Chaque « courbe » est en fait une « bande » dont l'épaisseur est donnée par la précision (un  $\sigma$ ) avec laquelle le paramètre post-képlérien correspondant est mesuré. Pour certains paramètres, ces bandes sont trop fines pour être visibles. Les zones grisées correspondraient à un sinus de l'inclinaison du plan orbital par rapport au plan du ciel plus grand que 1, et sont donc physiquement exclues.

pulsar. En effet,  $\dot{P}$  désigne la variation séculaire  $\langle dP/dt \rangle$  de la période orbitale. Des calculs théoriques détaillés du mouvement de deux objets gravitationnellement condensés, en relativité générale, et en tenant compte des effets liés à la propagation à vitesse finie de l'interaction gravitationnelle [14], ont montré qu'un des effets observables de cette propagation était une diminution

séculaire de la période orbitale donnée par la formule

$$\dot{P}^{\text{GR}}(m_A, m_B) = -\frac{192\pi}{5} \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1 - e^2)^{7/2}} \left(\frac{GMn}{c^3}\right)^{5/3} \frac{m_A m_B}{M^2}$$

L'origine physique directe de cette diminution de la période orbitale est due à la modification, induite par la relativité générale, de la loi newtonienne habituelle d'attraction gravitationnelle entre deux corps :  $F_{\text{Newton}} = G m_A m_B / r_{AB}^2$ . Au lieu d'une telle loi simple, la relativité générale prédit une loi de force plus compliquée qui peut se développer sous la forme symbolique

$$F_{\text{Einstein}} = \frac{G m_A m_B}{r_{AB}^2} \left( 1 + \frac{v^2}{c^2} + \frac{v^4}{c^4} + \frac{v^5}{c^5} + \frac{v^6}{c^6} + \frac{v^7}{c^7} + \dots \right) \quad (6.20)$$

où, par exemple, «  $v^2/c^2$  » représente tout un ensemble de termes d'ordre  $v_A^2/c^2$ ,  $v_B^2/c^2$ ,  $v_A v_B/c^2$ , ou encore  $G m_A/c^2 r$  ou  $G m_B/c^2 r$ . Ici  $v_A$  désigne la vitesse du corps  $A$ ,  $v_B$  celle du corps  $B$  et  $r_{AB}$  la distance entre les deux corps. Particulièrement important est le terme d'ordre  $v^5/c^5$  dans l'équation (6.20). Ce terme est une conséquence directe de la propagation à vitesse finie de l'interaction gravitationnelle entre  $A$  et  $B$ , et son calcul montre qu'il contient une composante qui est opposée à la vitesse relative  $\mathbf{v}_A - \mathbf{v}_B$  des deux corps et qui, du coup, freine le mouvement orbital de chaque corps et le fait évoluer vers une orbite plus rapprochée de son compagnon (ayant du coup une période orbitale plus courte). Ce terme de « freinage » (qui est corrélé à l'émission d'ondes gravitationnelles)  $\delta F_{\text{Einstein}} \sim v^5/c^5 F_{\text{Newton}}$  induit une diminution séculaire de la période orbitale  $\dot{P}^{\text{GR}} \sim -(v/c)^5 \sim -10^{-12}$  qui est très petit, mais dont la réalité est vérifiée avec une précision fractionnaire d'ordre  $10^{-3}$  dans PSR B1913+16, et d'ordre 20 % dans PSR B1534+12 et PSR J1141–6545 [4, 11, 13].

Pour finir cette brève évocation des tests de la gravitation relativiste dans les pulsars binaires, indiquons qu'il existe un analogue, pour le régime des champs gravitationnels forts, du formalisme de paramétrisation des déviations possibles à la relativité générale mentionné au paragraphe 6.6 dans le cadre des champs gravitationnels faibles (paramètres post-newtoniens  $\gamma - 1$ ,  $\beta - 1$ ). Cet analogue est obtenu en considérant une famille à deux paramètres de théories relativistes de la gravitation qui diffèrent de la relativité générale en supposant que l'interaction gravitationnelle est propagée non seulement par un champ tensoriel  $g_{\mu\nu}$  mais aussi par un champ scalaire  $\varphi$ . Une telle classe de théories tenseur-scalaires de la gravitation permet de décrire à la fois les déviations possibles dans le système solaire, et dans les pulsars binaires. Elle permet aussi de démontrer explicitement que les pulsars binaires testent bien des effets de champs forts qui vont au-delà des tests de champs faibles du système solaire en exhibant des classes de théories qui sont compatibles avec toutes les observations dans le système solaire mais qui sont incompatibles avec les observations des pulsars binaires [4, 13].

## 6.10 Ondes gravitationnelles : propagation, génération et détection

Dès qu'il eut fini de construire la théorie de la relativité générale, Einstein réalisa qu'elle impliquait l'existence d'ondes de déformation de la géométrie de l'espace-temps, c'est-à-dire d'« ondes gravitationnelles » [2, 15]. Mathématiquement, ces ondes sont des analogues (avec le remplacement  $A_\mu \rightarrow h_{\mu\nu}$ ) des ondes électromagnétiques, mais conceptuellement elles signifient quelque chose de remarquable : elles exemplifient, de la manière la plus pure possible, le caractère « élastique » de l'espace-temps en relativité générale. Avant Einstein l'espace-temps était une structure rigide, donnée a priori, et qui n'était pas influencée par le contenu matériel de l'Univers. Après Einstein, une distribution de matière (ou plus généralement de masse-énergie) qui change au cours du temps, disons pour fixer les idées un système binaire de deux étoiles à neutrons ou de deux trous noirs, va non seulement déformer la chrono-géométrie de l'espace-temps dans son voisinage immédiat, mais cette déformation va se propager dans toutes les directions à partir du système considéré et va s'éloigner à l'infini sous forme d'une onde dont les oscillations vont refléter les variations temporelles de la distribution de matière. On voit alors que l'étude de ces ondes gravitationnelles pose trois problèmes séparés : celui de la génération, celui de la propagation, et, enfin, celui de la détection d'un tel rayonnement gravitationnel. Ces trois problèmes sont activement étudiés à l'heure actuelle, car on espère détecter bientôt des ondes gravitationnelles, et pouvoir ainsi avoir des renseignements nouveaux sur l'Univers [16]. Contentons nous ici d'une introduction élémentaire à ce domaine de recherche.

Considérons d'abord le cas le plus simple d'ondes gravitationnelles très faibles, et en dehors de leurs sources matérielles. La géométrie d'un tel espace-temps peut s'écrire, comme au paragraphe 6.6, selon  $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$ , où  $h_{\mu\nu} \ll 1$ . Au premier ordre en  $h$ , et en dehors de la source (c'est-à-dire dans un domaine où  $T_{\mu\nu}(x) = 0$ ), la perturbation de la géométrie,  $h_{\mu\nu}(x)$ , satisfait une équation homogène obtenue en remplaçant le second membre de l'équation (6.11) par zéro. On démontre que l'on peut simplifier cette équation par un choix adéquat du système de coordonnées. Dans un système de coordonnées dit « transverse et sans trace » (TT) les seules composantes non nulles d'une onde gravitationnelle générale sont ses composantes spatiales  $h_{ij}^{\text{TT}}$ ,  $i, j = 1, 2, 3$  (c'est-à-dire  $h_{00}^{\text{TT}} = 0 = h_{0i}^{\text{TT}}$ ), et celles-ci satisfont

$$\square h_{ij}^{\text{TT}} = 0, \quad \partial_j h_{ij}^{\text{TT}} = 0, \quad h_{jj}^{\text{TT}} = 0 \quad (6.21)$$

La première équation (6.21), où apparaît l'opérateur des ondes  $\square = \Delta - c^{-2} \partial_t^2$ , montre que les ondes gravitationnelles se propagent (comme les ondes électromagnétiques) à la vitesse  $c$ . Si l'on considère pour simplifier une onde monochromatique plane ( $h_{ij}^{\text{TT}} = \zeta_{ij} \exp(i \mathbf{k} \cdot \mathbf{x} - i \omega t)$  + conjugué complexe ; avec  $\omega = c |\mathbf{k}|$ ), la deuxième équation (6.21) montre que le tenseur  $\zeta_{ij}$  mesurant la polarisation d'une onde gravitationnelle n'a de composantes non nulles

que dans le plan orthogonal à la direction de propagation de l'onde :  $\zeta_{ij} k^j = 0$ . Enfin, la troisième équation (6.21) montre que le tenseur de polarisation  $\zeta_{ij}$  est à trace nulle :  $\zeta_{jj} = 0$ . De façon concrète, cela veut dire que si une onde gravitationnelle se propage dans la direction  $z$ , sa polarisation est décrite par une matrice  $2 \times 2$ ,  $\begin{pmatrix} \zeta_{xx} & \zeta_{xy} \\ \zeta_{yx} & \zeta_{yy} \end{pmatrix}$ , qui est symétrique et à trace nulle. Une telle matrice de polarisation ne contient donc que deux composantes (complexes) indépendantes :  $\zeta_+ \equiv \zeta_{xx} = -\zeta_{yy}$ , et  $\zeta_\times \equiv \zeta_{xy} = \zeta_{yx}$ . C'est le même nombre de composantes (complexes) indépendantes que pour une onde électromagnétique. En effet, dans une jauge transverse, une onde électromagnétique n'a que des composantes spatiales  $A_i^T$  qui satisfont

$$\square A_i^T = 0, \quad \partial_j A_j^T = 0 \quad (6.22)$$

Comme ci-dessus la première équation (6.22) veut dire qu'une onde électromagnétique se propage à la vitesse  $c$ , et la seconde équation montre qu'une onde électromagnétique monochromatique plane ( $A_i^T = \zeta_i \exp(i \mathbf{k} \cdot \mathbf{x} - i \omega t) + \text{c.c.}$ ;  $\omega = c|\mathbf{k}|$ ) est décrite par un vecteur de polarisation  $\zeta_i$  qui est orthogonal à la direction de propagation :  $\zeta_j k^j = 0$ . Pour une onde se propageant dans la direction  $z$  un tel vecteur n'a que deux composantes (complexes) indépendantes :  $\zeta_x$  et  $\zeta_y$ . C'est bien le même nombre de composantes que pour une onde gravitationnelle, mais on voit que les deux quantités mesurant la polarisation d'une onde gravitationnelle  $\zeta_+ = \zeta_{xx} = -\zeta_{yy}$ ,  $\zeta_\times = \zeta_{xy} = \zeta_{yx}$  sont mathématiquement très différentes des deux quantités  $\zeta_x, \zeta_y$  mesurant la polarisation d'une onde électromagnétique (voir cependant le paragraphe 6.11). Nous avons discuté ici la propagation d'une onde gravitationnelle dans un espace-temps de fond décrit par la métrique de Minkowski  $\eta_{\mu\nu}$ . On peut aussi considérer la propagation d'une onde dans un espace-temps de fond courbe, c'est-à-dire étudier les solutions des équations d'Einstein (6.9) de la forme  $g_{\mu\nu}(x) = g_{\mu\nu}^B(x) + h_{\mu\nu}(x)$  où  $h_{\mu\nu}$  est non seulement petit, mais varie sur les échelles de temps et d'espace beaucoup plus courtes que la métrique de fond  $g_{\mu\nu}^B(x)$ . Une telle étude est, par exemple, nécessaire pour comprendre la propagation des ondes gravitationnelles dans l'Univers cosmologique.

Le problème de la *génération* consiste à chercher le lien entre l'amplitude tensorielle  $h_{ij}^{\text{TT}}$  du rayonnement gravitationnel dans la zone d'ondes, et le mouvement et la structure de la source. Si l'on considère le cas le plus simple d'une source suffisamment peu condensée pour n'engendrer que des ondes partout faibles ( $g_{\mu\nu} - \eta_{\mu\nu} = h_{\mu\nu} \ll 1$ ), on peut utiliser l'approximation linéarisée des équations d'Einstein (6.9), c'est-à-dire les équations (6.11). On peut résoudre les équations (6.11) par la même technique utilisée pour résoudre les équations de Maxwell (6.12) : on fixe le système de coordonnées en imposant  $\partial^\alpha h_{\alpha\mu} - \frac{1}{2} \partial_\mu h_\alpha^\alpha = 0$  (analogue de la condition de jauge de Lorentz  $\partial^\alpha A_\alpha = 0$ ), puis on inverse l'opérateur des ondes en utilisant les potentiels retardés. Enfin, on doit étudier la forme asymptotique, à l'infini, de l'onde émise, et la mettre sous la forme réduite d'une amplitude transverse et sans trace  $h_{ij}^{\text{TT}}$  satisfaisant les équations (6.21) (analogue à une onde

électromagnétique transverse  $A_i^T$  satisfaisant (6.22)). On trouve alors que, de même que la conservation de la charge implique qu'il n'y a pas d'émission électrique de type monopolaire, mais seulement dipolaire ou d'un ordre de polarité plus élevé, de même la conservation de l'énergie-impulsion implique l'absence de rayonnements gravitationnels monopolaires *et* dipolaires. Pour une source lentement variable ( $v/c \ll 1$ ), le rayonnement gravitationnel dominant est de type « quadrupolaire ». Il est donné, dans la zone d'ondes, par une expression de la forme

$$h_{ij}^{TT}(t, r, \mathbf{n}) \simeq \frac{2G}{c^4 r} \frac{\partial^2}{\partial t^2} [I_{ij}(t - r/c)]^{TT} \quad (6.23)$$

Ici  $r$  désigne la distance au centre de masse de la source,  $I_{ij}(t) \equiv \int d^3x c^{-2} T^{00}(t, \mathbf{x}) (x^i x^j - \frac{1}{3} \mathbf{x}^2 \delta^{ij})$  le moment quadrupolaire de la distribution de masse-énergie, et l'indice supérieur TT désigne une opération algébrique de projection du tenseur quadrupolaire  $I_{ij}$  (qui est une matrice  $3 \times 3$ ) qui n'en retient que la partie orthogonale à la direction de propagation locale de l'onde  $n^i \equiv x^i/r$ , et sans trace ( $I_{ij}^{TT}$  est donc, localement, une matrice  $2 \times 2$  symétrique et sans trace du même type que  $\zeta_{ij}$  ci-dessus). La formule (6.23) (qui a essentiellement été obtenue par Einstein en 1918 [15]) n'est que la première approximation d'un développement en puissances de  $v/c$ , où  $v$  désigne une vitesse interne caractéristique de la source. La perspective de pouvoir détecter bientôt des ondes gravitationnelles ont motivé les théoriciens à améliorer la formule (6.23) : (i) en décrivant les termes d'ordre supérieur en  $v/c$ , jusqu'à un ordre très élevé, et (ii) en utilisant de nouvelles méthodes d'approximation permettant de traiter des sources contenant des régions de champs gravitationnels forts (comme, par exemple, un système binaire de deux trous noirs ou deux étoiles à neutrons) (voir ci-dessous pour des résultats très récents).

Enfin, le problème de la *détection*, dont le pionnier fut Joseph Weber dans les années soixante, suscite de nos jours de très actives recherches expérimentales. Le principe de tous les détecteurs est qu'une onde gravitationnelle d'amplitude  $h_{ij}^{TT}$  induit, lors de son passage, un changement de la distance  $L$  entre deux corps de l'ordre de  $\delta L \sim hL$ . Une façon de voir ceci est de considérer l'action d'une onde  $h_{ij}^{TT}$  sur deux particules libres, au repos avant l'arrivée de l'onde aux positions  $x_1^i$  et  $x_2^i$  respectivement. Comme on l'a vu chaque particule, en présence de l'onde, va suivre un mouvement géodésique dans la géométrie  $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$  (avec  $h_{00} = h_{0i} = 0$  et  $h_{ij} = h_{ij}^{TT}$ ). En écrivant l'équation des géodésiques, équation (6.7), on trouve qu'elle se réduit simplement (au premier ordre en  $h$ ) à  $d^2 x^i / ds^2 = 0$ . Donc, des particules initialement au repos ( $\dot{x}^i = \text{const.}$ ) restent au repos dans un système de coordonnées transverse et sans trace! Cela ne veut cependant pas dire que l'onde gravitationnelle n'a aucun effet observable. En effet, comme la géométrie spatiale est perturbée par le passage de l'onde,  $g_{ij}(t, \mathbf{x}) = \delta_{ij} + h_{ij}^{TT}(t, \mathbf{x})$ , on trouve que la distance physique entre les deux particules  $x_1^i, x_2^i$  (observable, par exemple, en mesurant le temps d'aller retour de la lumière entre les deux particules) varie, lors du passage de l'onde, selon  $L^2 = (\delta_{ij} + h_{ij}^{TT})(x_2^i - x_1^i)(x_2^j - x_1^j)$ .

Le problème de la détection d'une onde gravitationnelle est ainsi ramené au problème de détecter un petit déplacement relatif  $\delta L/L \sim h$ . En utilisant la formule (6.23), on trouve que l'ordre de grandeur de  $h$ , pour les sources astrophysiques connues ou espérées (par exemple, système très rapproché de deux étoiles à neutrons ou deux trous noirs), situées à des distances telles qu'on puisse espérer voir plusieurs événements par an ( $r \gtrsim 600$  millions d'années lumière), est vraiment extrêmement petit :  $h \lesssim 10^{-22}$  pour des signaux dont la fréquence caractéristique est d'une centaine de Hertz. Plusieurs types de détecteurs ont été développés depuis les travaux pionniers de J. Weber [16]. À l'heure actuelle, les détecteurs qui devraient réussir dans le futur proche à détecter des amplitudes  $h \sim \delta L/L \sim 10^{-22}$  sont de grands interféromètres, du type Michelson ou Fabry-Pérot, ayant des bras kilométriques et dans lesquels est injecté un faisceau laser monochromatique très puissant. De tels détecteurs interférométriques terrestres existent actuellement aux USA (détecteurs LIGO [17]), en Europe (détecteurs VIRGO [18] et GEO 600 [19]) et ailleurs (comme le détecteur TAMA au Japon). En outre, le projet spatial international LISA [20], constitué d'un interféromètre entre satellites distants de plusieurs millions de kilomètres, devrait permettre, dans une dizaine d'années, de détecter des ondes gravitationnelles de basse fréquence ( $\sim$  centième ou millième de Hertz). L'ensemble de ces détecteurs d'ondes gravitationnelles promet d'apporter des renseignements inestimables pour l'astronomie en ouvrant sur l'Univers une nouvelle « fenêtre » beaucoup plus transparente que les diverses « fenêtres » électromagnétiques (ou neutriniques) qui ont tant renouvelé notre connaissance de l'Univers au vingtième siècle.

L'extrême petitesse des signaux gravitationnels attendus a conduit de nombreux expérimentateurs à dépenser, depuis de nombreuses années, des trésors d'ingénuité et de savoir-faire pour développer des technologies assez précises et fiables (voir [17–20]). Pour finir, indiquons aussi que de grands efforts théoriques sont dépensés, à la fois pour calculer les prédictions de la relativité générale pour les ondes gravitationnelles émises par certaines sources, et pour développer des méthodes adaptées d'extraction du signal gravitationnel du bruit de fond présent dans les détecteurs. Par exemple, l'une des sources les plus prometteuses pour les détecteurs terrestres est le train d'ondes gravitationnelles émis par un système de deux trous noirs, et notamment la partie finale (la plus intense) de ce train d'ondes, qui est émise dans les dernières orbites de ce système, et lors de la coalescence finale des deux trous noirs pour former un seul trou noir plus massif. On a vu ci-dessus (voir paragraphe 6.9) que la propagation à la vitesse finie de l'interaction gravitationnelle entre les deux corps d'un système binaire donne lieu à une accélération progressive de la fréquence orbitale, liée à un rapprochement progressif des deux corps. On parle ici des étapes finales d'un tel processus où les deux corps sont si rapprochés qu'ils orbitent l'un autour de l'autre, dans un mouvement de spirale qui s'accélère jusqu'à atteindre (pour les dernières orbites « stables ») des vitesses qui deviennent comparables à la vitesse de la lumière, tout en lui restant cependant légèrement inférieures. Pour pouvoir déterminer, dans

une telle situation, et avec une précision acceptable pour les besoins de la détection, la dynamique d'un système de deux trous noirs, et l'amplitude gravitationnelle  $h_{ij}^{\text{TT}}$  qu'il émet, il a fallu développer à un très haut niveau de précision tout un ensemble de techniques analytiques. Par exemple, il a fallu calculer le développement (6.20) de la force déterminant le mouvement des deux corps à un ordre très élevé et calculer aussi l'amplitude  $h_{ij}^{\text{TT}}$  du rayonnement gravitationnel émis à l'infini avec une précision allant bien au-delà de l'approximation du quadrupole (6.23). Ces calculs sont comparables en complexité aux calculs d'ordres élevés en théorie quantique des champs. Certaines techniques développées pour la théorie quantique des champs se sont d'ailleurs révélées être très utiles pour ces calculs de la théorie (classique) de la relativité générale (comme certaines méthodes de resommation, et l'utilisation mathématique d'un prolongement analytique dans la dimension de l'espace-temps). Pour une entrée dans la littérature de ces méthodes analytiques modernes voir [21], et pour un exemple de résultat d'intérêt physique direct pour la détection obtenu par de telles méthodes voir la figure 6.5 [22] qui représente une composante de l'amplitude gravitationnelle  $h_{ij}^{\text{TT}}(t)$  émise par les étapes ultimes de l'évolution d'un système de deux trous noirs de même masse. Les premières oscillations représentées sur la figure 6.5 sont émises par les dernières orbites quasi-circulaires (mouvement accéléré en spirale de rayon décroissant). Le milieu du signal correspond à une phase où, ayant traversé la dernière orbite stable, les deux trous noirs « tombent » en spiralant rapidement l'un vers l'autre. En effet, contrairement à la théorie de Newton qui prédit que deux corps condensés pourraient orbiter l'un autour de l'autre sur une orbite de rayon arbitrairement petit (essentiellement jusqu'à ce que les corps se touchent), la théorie d'Einstein prédit une loi de force modifiée entre les deux corps, équation (6.20), dont l'analyse montre qu'elle est si attractive qu'elle ne permet plus l'existence d'orbites circulaires stables quand la distance entre les deux corps devient inférieure à environ  $6G(m_A + m_B)/c^2$ . Dans le cas de deux trous noirs, cette distance est assez supérieure aux « rayons » des trous noirs ( $2Gm_A/c^2$  et  $2Gm_B/c^2$ ) pour pouvoir encore traiter analytiquement le début de la « chute en spirale » des deux trous noirs l'un vers l'autre. Les dernières oscillations sur la figure 6.5 sont émises par le trou noir tournant (initialement très déformé) formé par la « fusion » des deux trous noirs séparés en un seul trou noir final.

Malgré la sophistication des résultats analytiques employés pour obtenir les résultats du genre indiqué sur la figure 6.5, certains aspects du signal  $h(t)$  ainsi estimé restent conjecturaux (notamment le signal correspondant à la « chute en spirale » et à la « fusion » des deux trous noirs). De nombreux efforts sont déployés dans le monde pour attaquer le problème de la coalescence de deux trous noirs en résolvant les équations d'Einstein (6.9) de façon numérique. La « relativité numérique » a fait d'importants progrès ces dernières années, et devrait bientôt améliorer notre connaissance de l'amplitude gravitationnelle émise lors des dernières étapes de l'évolution de systèmes binaires d'objets gravitationnellement condensés (trous noirs ou étoiles à neutrons).



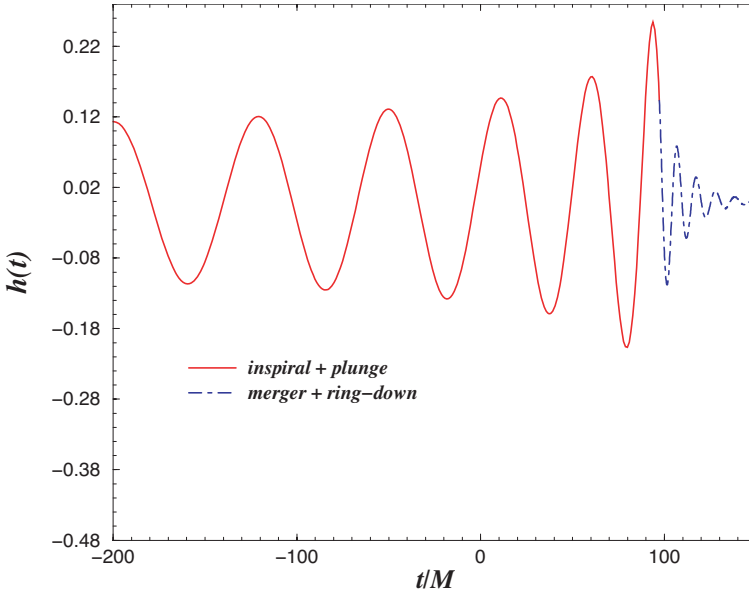


FIG. 6.5 — Amplitude gravitationnelle  $h(t)$  émise par les étapes ultimes de l'évolution d'un système de deux trous noirs de même masse. Le début (à gauche de la figure) du signal (qui est sinusoïdal) correspond à un mouvement en spirale, de rayon décroissant, de deux trous noirs séparés; le milieu correspond à une « chute en spirale » rapide des deux trous noirs l'un vers l'autre; la fin (à droite) correspond aux « oscillations » du trou noir tournant unique final formé par la « fusion » des deux trous noirs initiaux.

## 6.11 Relativité générale et théorie quantique : de la supergravité à la théorie des cordes

Nous n'avons discuté jusqu'ici que la relativité générale *classique*, c'est-à-dire en négligeant tout effet quantique. Que devient la théorie dans le régime quantique? Cette question, d'apparence innocente, ouvre en fait des perspectives très vastes dont l'étude est encore en plein chantier. Nous ne ferons ici que l'effleurer, en indiquant au lecteur quelques chemins sur lesquels le défi d'unifier relativité générale et théorie quantique a entraîné la physique théorique contemporaine.

Rappelons que, dès les débuts de la formulation quasi-définitive de la théorie quantique (1925–1930), les créateurs de la mécanique quantique (Born, Heisenberg, Jordan; Dirac; Pauli; ...) montrèrent comment « quantifier » non plus des systèmes de plusieurs particules (comme un atome), mais des *champs*, c'est-à-dire des systèmes dynamiques continus, dont la description classique implique une distribution continue d'énergie et d'impulsion dans l'espace. Ils montrèrent en particulier comment *quantifier* (c'est-à-dire formuler dans un

cadre compatible avec la théorie quantique) le champ électromagnétique  $A_\mu$ , dont nous avons rappelé plus haut qu'il satisfaisait, au niveau classique, les équations de Maxwell (6.12). Ils rencontrèrent cependant une difficulté due au fait suivant. En théorie quantique, la physique de l'évolution d'un système est essentiellement contenue dans les *amplitudes de transition*  $A(f, i)$  entre un état initial étiqueté par  $i$ , et un état final étiqueté par  $f$ . Ces quantités  $A(f, i)$  sont des nombres complexes. Elles satisfont une propriété de « transativité » du type

$$A(f, i) = \sum_n A(f, n) A(n, i) \quad (6.24)$$

qui contient une somme sur tous les états intermédiaires possibles, étiquetés par  $n$  (cette somme devenant une intégrale quand il y a un continuum d'états intermédiaires possibles). R. Feynman a utilisé l'équation (6.24) comme point de départ d'une nouvelle formulation de la théorie quantique, en l'interprétant comme un analogue du *principe de Huyghens* : si l'on conçoit  $A(f, i)$  comme l'amplitude, « au point  $f$  », d'une « onde » émise « au point  $i$  », l'équation (6.24) exprime que cette amplitude peut se calculer en considérant que l'« onde » émise par  $i$  passe par tous les « points » intermédiaires possibles  $n$  ( $A(n, i)$ ), en réémettant des « ondelettes » à partir de ces points intermédiaires ( $A(f, n)$ ) qui se superposent ensuite pour former l'onde totale arrivant au « point final  $f$  ».

La propriété (6.24) ne pose pas de problème dans la mécanique quantique des systèmes discrets (systèmes de particules). Elle indique simplement que l'amplitude  $A(f, i)$  se comporte comme une onde, et donc doit satisfaire une « équation d'ondes » (ce qui est bien le cas de l'équation de Schrödinger décrivant la dépendance de  $A(f, i)$  par rapport aux paramètres décrivant la configuration finale  $f$ ). En revanche, la propriété (6.24) pose des problèmes redoutables quand on l'applique à la quantification des systèmes dynamiques continus (champs). En effet, pour de tels systèmes l'« espace » des états intermédiaires possibles est infiniment plus vaste que dans le cas de la mécanique des systèmes discrets. Grosso modo, les états intermédiaires possibles pour un champ peuvent être décrits comme contenant  $\ell = 1, 2, 3, \dots$  excitations quantiques du champ, chaque excitation quantique (ou paire de « particules virtuelles ») étant essentiellement décrite par une onde plane,  $\zeta \exp(i k_\mu x^\mu)$ , où  $\zeta$  mesure la polarisation de ces particules virtuelles, et  $k^\mu = \eta^{\mu\nu} k_\nu$ , avec  $k^0 = \omega$ ,  $k^i = \mathbf{k}$ , leur fréquence angulaire et vecteur d'onde, ou (en utilisant les relations de Planck-Einstein-de Broglie  $E = \hbar \omega$ ,  $\mathbf{p} = \hbar \mathbf{k}$ ) leur impulsion-énergie  $p^\mu = \hbar k^\mu$ . La théorie quantique montre que (essentiellement à cause du principe d'incertitude) les quadri-fréquences (et quadri-impulsions)  $p^\mu = \hbar k^\mu$  des états intermédiaires ne peuvent pas être contraints à satisfaire l'équation classique  $\eta_{\mu\nu} p^\mu p^\nu = -m^2$  (c'est-à-dire  $E^2 = \mathbf{p}^2 + m^2$ ; on utilise  $c = 1$  dans cette section). En conséquence, la somme sur les états intermédiaires, pour la théorie quantique d'un champ, contient (entre autres) : (i) quand  $\ell = 1$  (état intermédiaire contenant seulement une paire de particules virtuelles, dit « contribution à une boucle »), une intégrale sur une quadri-impulsion

$p^\mu$ ,  $\int d^4p = \int dE \int d\mathbf{p}$ ; (ii) quand  $\ell = 2$  (deux paires de particules virtuelles; « contribution à deux boucles »), une intégrale sur deux quadri-impulsions  $p_1^\mu$ ,  $p_2^\mu$ ,  $\int d^4p_1 d^4p_2$ ; etc. Le point délicat vient du fait que les énergies-impulsions des états intermédiaires peuvent prendre des valeurs arbitrairement élevées. Cette possibilité est directement reliée (par transformée de Fourier) au fait qu'un champ possède un nombre infini de degrés de liberté, correspondant à des configurations qui varient sur des échelles de temps et d'espace arbitrairement petites.

Les problèmes posés par la nécessité d'intégrer sur le domaine infini des quadri-impulsions des particules virtuelles intermédiaires (c'est-à-dire de tenir compte du fait que les configurations d'un champ peuvent varier sur des échelles arbitrairement petites) apparurent dès les années trente quand on étudia en détail la théorie quantique du champ électromagnétique  $A_\mu$  (dite électrodynamique quantique, ou QED en anglais). Ces problèmes se posaient sous la forme suivante : quand on calculait l'amplitude de transition pour des états initiaux et finals donnés (par exemple la collision de deux particules de lumière, deux photons entrants, en deux photons sortants) en utilisant (6.24), on trouvait un résultat donné sous la forme d'une *intégrale divergente*, à cause de l'intégrale (à l'approximation d'une boucle,  $\ell = 1$ ) sur l'énergie-impulsion arbitrairement grande décrivant les paires électrons-positrons virtuelles apparaissant comme états intermédiaires possibles. Petit à petit, les physiciens théoriciens comprirent que le type d'intégrales divergentes apparaissant en QED était relativement bénin, et, après la seconde guerre mondiale, ils développèrent une méthode (*théorie de la renormalisation*) permettant d'isoler de façon non ambiguë la partie infinie de ces intégrales, et de la soustraire en exprimant les amplitudes  $A(f, i)$  seulement en fonction de quantités directement observables [23] (travaux de J. Schwinger, R. Feynman, F. Dyson etc.)

Les travaux précédents ont conduit à développer des théories quantiques cohérentes non seulement du champ électromagnétique  $A_\mu$  (QED), mais aussi des généralisations de l'électromagnétisme (théorie de Yang-Mills ou théorie de jauge non-abélienne) qui se révélèrent fournir d'excellentes descriptions des nouvelles interactions entre particules élémentaires découvertes au vingtième siècle (théorie électrofaible unifiant partiellement électromagnétisme et interactions faibles, chromodynamique quantique décrivant les interactions fortes). Toutes ces théories ne donnent lieu qu'à des divergences relativement bénignes, qui peuvent être « renormalisées » et conduire ainsi à calculer les amplitudes  $A(f, i)$  correspondant à des processus physiques observables [23] (travaux, notamment, de G. 't Hooft et M. Veltman).

Que se passe-t-il quand on utilise (6.24) pour construire une théorie quantique « perturbative » (c'est-à-dire obtenue en développant selon le nombre  $\ell$  des paires de particules virtuelles apparaissant dans les états intermédiaires) de la relativité générale? La réponse est que les intégrales sur les quadri-impulsions des particules virtuelles intermédiaires ne sont plus du tout du type bénin qui permettait de les renormaliser dans le cas plus simple de l'électromagnétisme. La cause de cette différence n'est pas accidentelle mais est liée

à la physique essentielle de la gravitation relativiste. En effet, les particules virtuelles ont, comme on l'a dit, des énergies  $E$  arbitrairement élevées. À cause des relations de base qui conduisirent Einstein à développer la relativité générale, c'est-à-dire  $E = m_i$  et  $m_i = m_g$ , on en déduit que ces particules virtuelles correspondent à des masses gravitationnelles  $m_g$  arbitrairement grandes. Elles vont donc finir par créer des effets gravitationnels intenses, et de plus en plus intenses quand le nombre  $\ell$  de paires de particules virtuelles augmente. Ces interactions gravitationnelles qui croissent sans limite avec l'énergie et l'impulsion, correspondent (par transformée de Fourier) à des configurations de champ concentrées dans des échelles de temps et d'espace arbitrairement petites. Une façon de voir pourquoi le champ gravitationnel quantique crée des problèmes beaucoup plus violents que le champ électromagnétique quantique est, tout simplement, de recourir à l'analyse dimensionnelle. Des considérations simples montrent en effet que l'amplitude relative (non dimensionnée) à une boucle  $A_1$  doit être proportionnelle au produit  $\hbar G$  et doit contenir une intégrale  $\int d^4k$ . Or, dès 1900, Planck avait remarqué que (en unités où  $c = 1$ ) les dimensions de  $\hbar$  et  $G$  étaient telles que le produit  $\hbar G$  avait les dimensions d'une longueur (ou d'un temps) au carré

$$\ell_P \equiv \sqrt{\frac{\hbar G}{c^3}} \simeq 1,6 \times 10^{-33} \text{ cm}, \quad t_P \equiv \sqrt{\frac{\hbar G}{c^5}} \simeq 5,4 \times 10^{-44} \text{ s} \quad (6.25)$$

On en déduit que l'intégrale  $\int d^4k f(k)$  doit avoir les dimensions d'une fréquence au carré, et donc que  $A_1$  doit (quand  $k \rightarrow \infty$ ) être du type,  $A_1 \sim \hbar G \int d^4k/k^2$ . Une telle intégrale diverge quadratiquement avec la borne supérieure  $\Lambda$  de l'intégrale (fréquence de coupure  $|k| \leq \Lambda$ ), d'où  $A_1 \sim \hbar G \Lambda^2 \sim t_P^2 \Lambda^2$ . L'extension de cette analyse dimensionnelle aux états intermédiaires à plusieurs boucles ( $\ell > 1$ ) fait apparaître des divergences polynomiales encore plus sévères, et d'un type tel que la puissance de  $\Lambda$  qui apparaît croît sans limite avec  $\ell$ .

En conclusion, les caractéristiques physiques essentielles de la gravitation ( $E = m_i = m_g$ , et la dimension de la constante de Newton  $G$ ) impliquent l'impossibilité de généraliser au cas gravitationnel les méthodes qui ont permis un développement satisfaisant d'une théorie quantique des autres interactions (électromagnétiques, faibles et fortes). Plusieurs voies ont été explorées pour sortir de cette impasse. Certaines essaient de quantifier la relativité générale de façon non perturbative, c'est-à-dire sans utiliser un développement en états intermédiaires (6.24) (travaux d'A. Ashtekar, L. Smolin et d'autres). Certaines ont essayé de généraliser la relativité générale en ajoutant au champ gravitationnel einsteinien (bosonique)  $g_{\mu\nu}(x)$  un champ à caractère fermionique, le champ gravitinique  $\psi_\mu(x)$ . Il est en effet remarquable qu'il soit possible de définir une théorie, dite de la *supergravité*, qui généralise de façon profonde l'invariance géométrique de la relativité générale. Après la découverte en 1974 (par J. Wess et B. Zumino) d'une nouvelle symétrie globale possible pour des champs bosoniques et fermioniques en interaction, la *supersymétrie* (qui est une sorte de rotation globale transformant bosons en fermions

et réciproquement), D.Z. Freedman, P. van Nieuwenhuizen et S. Ferrara, et, S. Deser et B. Zumino, montrèrent que l'on pouvait généraliser la supersymétrie globale en une *supersymétrie locale*, c'est-à-dire variable de point en point dans l'espace-temps. La supersymétrie locale est une sorte de généralisation fermionique (c'est-à-dire avec paramètres anti-commutants) de l'invariance géométrique de base de la relativité générale (invariance sous tous les changements de coordonnées). La généralisation de la théorie de la gravitation d'Einstein qui admet une telle supersymétrie locale s'appelle *théorie de la supergravité*. Comme on l'a dit, en quatre dimensions, cette théorie contient en plus du champ bosonique (commutant)  $g_{\mu\nu}(x)$ , un champ fermionique (anti-commutant)  $\psi_\mu(x)$  qui est à la fois un vecteur d'espace-temps (indice  $\mu$ ) et un spineur. (C'est un champ sans masse de spin  $3/2$ , intermédiaire entre un champ sans masse de spin 1 comme  $A_\mu$ , et un champ sans masse de spin 2 comme  $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ .) La supergravité fut étendue à des structures fermioniques plus riches (multi-gravitoniques), et fut formulée dans des espace-temps ayant plus que quatre dimensions. Il est d'ailleurs remarquable qu'il existe une dimension maximale, égale à  $D = 11$ , admettant la formulation d'une théorie de la supergravité (supergravité maximale construite par E. Cremmer, B. Julia et J. Scherk). L'espoir initial sous-tendant la construction de ces théories de la supergravité était qu'elles allaient peut être permettre de donner un sens au calcul perturbatif (6.24) des amplitudes quantiques. En effet, on trouve par exemple qu'à une boucle,  $\ell = 1$ , les contributions venant des états intermédiaires fermioniques ont le signe opposé des contributions bosoniques et (à cause de la supersymétrie bosons  $\leftrightarrow$  fermions) les compensent exactement. Malheureusement, bien que de telles compensations existent pour les premiers ordres d'approximation, on se convainc petit à petit que cela n'allait pas être le cas à tous les ordres. Le fait que la constante d'interaction gravitationnelle  $G$  ait « une mauvaise dimension » reste vrai et crée des divergences non renormalisables à partir d'un certain nombre de boucles  $\ell$ .

Pendant ce temps, une troisième voie pour définir une théorie quantique cohérente de la gravitation s'est développée, sous le nom de *théorie des cordes*. Initialement formulées (notamment par G. Veneziano, M. Virasoro, P. Ramond, A. Neveu et J.H. Schwarz) comme modèles des interactions fortes, les théories des cordes sont fondées sur la quantification de la dynamique relativiste d'un objet étendu selon une dimension spatiale : une « corde ». Cette corde peut être refermée sur elle-même, comme un petit ruban élastique (corde fermée), ou elle peut avoir deux extrémités (corde ouverte). Notons que le point de départ de la théorie des cordes ne comprend que l'espace-temps de Poincaré-Minkowski, c'est-à-dire la métrique  $\eta_{\mu\nu}$  de l'équation (6.2), et la théorie quantique (avec la constante  $\hbar = h/2\pi$ ). En particulier, la seule symétrie manifeste de la dynamique classique d'une corde est le groupe de Poincaré (6.3). Il est cependant remarquable que (comme l'ont montré T. Yoneya, et J. Scherk et J.H. Schwarz, en 1974) l'une des excitations quantiques d'une corde fermée reproduit, quand on prend une certaine limite, toute la structure non-linéaire de la relativité générale (voir ci-dessous). Parmi les

autres propriétés remarquables de la théorie des cordes [24] signalons qu'elle est la première théorie physique à déterminer la dimension  $D$  de l'espace-temps. En effet, cette théorie n'est cohérente que si  $D = 10$ , pour ses versions admettant des excitations fermioniques (la théorie des cordes purement bosonique sélectionne  $D = 26$ ). Le fait que  $10 > 4$  ne signifie pas que cette théorie n'a pas de pertinence au monde réel. En effet, on sait depuis les années vingt (travaux de T. Kaluza et O. Klein) qu'un espace-temps de dimension  $D > 4$  est compatible avec l'expérience si les dimensions (spatiales) supplémentaires se referment sur elle-mêmes (c'est-à-dire sont « compactifiées ») sur des échelles de distance très petites. La physique à basse énergie d'une telle théorie semble se dérouler dans un espace-temps à 4 dimensions, mais elle contient de nouveaux champs (a priori sans masse) liés à la géométrie des dimensions supplémentaires compactifiées.

Mais la propriété la plus remarquable de la théorie des cordes est qu'elle semble éviter d'une manière radicale les problèmes d'intégrales divergentes (non renormalisables) qui grèvent tout essai direct de quantifier la gravitation de façon perturbative. Pour expliquer comment la théorie des cordes arrive à un tel résultat, il nous faut exposer quelques éléments du formalisme de la théorie des cordes.

Rappelons que la dynamique classique de tout système s'obtient en minimisant une fonctionnelle de l'évolution au cours du temps de la configuration du système, appelée *action* (principe de moindre action). Par exemple, l'action pour une particule de masse  $m$ , évoluant dans un espace-temps riemannien (6.6), est proportionnelle à la longueur de la ligne qu'elle trace dans l'espace-temps :  $S = -m \int ds$ . Cette action est minimisée quand la particule suit une géodésique, c'est-à-dire quand son équation du mouvement est donnée par (6.7). D'après Y. Nambu et T. Goto, l'action pour une corde est  $S = -T \iint dA$ , où le paramètre  $T$  (analogue à  $m$  pour la particule) s'appelle la *tension* de la corde, et où  $\iint dA$  est l'aire de la surface (à deux dimensions) tracée par l'évolution de la corde dans l'espace-temps (à  $D$  dimensions) dans lequel elle vit. En théorie quantique, la fonctionnelle d'action sert (comme l'a montré R. Feynman) pour définir l'amplitude de transition (6.24). Essentiellement, quand on considère deux configurations intermédiaires (au sens du membre de droite de (6.24))  $m$  et  $n$  proches l'une de l'autre, l'amplitude  $A(n, m)$  est proportionnelle à  $\exp(iS(n, m)/\hbar)$  où  $S(n, m)$  est l'action classique minimale pour que le système considéré aille de la configuration étiquetée par  $n$  à celle étiquetée par  $m$ . Généralisant la décomposition (6.24) en introduisant une infinité de configurations intermédiaires proches l'une de l'autre, on finit par exprimer (dans une généralisation du principe de Huyghens) l'amplitude  $A(f, i)$  comme une somme multiple sur tous les « chemins » (dans l'espace de configuration du système étudié) faisant passer de l'état initial  $i$  à l'état final  $f$ . Chaque chemin contribue un terme  $e^{i\phi}$  où la phase  $\phi = S/\hbar$  est proportionnelle l'action  $S$  correspondant à ce « chemin », c'est-à-dire à cette évolution possible du système. En théorie des cordes  $\phi = -(T/\hbar) \iint dA$ . La phase étant une quantité non dimensionnée, et  $\iint dA$  ayant la dimension

d'une aire, on voit que la théorie quantique des cordes fait intervenir à un niveau fondamental une quantité,  $\hbar/T$ , ayant les dimensions d'une longueur au carré. Plus précisément on appelle longueur fondamentale de la théorie des cordes,  $\ell_s$ , la quantité définie par

$$\ell_s^2 \equiv \alpha' \equiv \frac{\hbar}{2\pi T} \quad (6.26)$$

Cette longueur fondamentale intervient de façon centrale en théorie des cordes. Grosso modo, elle définit la « taille » des états quantiques d'une corde. Si  $\ell_s$  est beaucoup plus petit que la résolution observationnelle avec laquelle on étudie une corde, et ses interactions, la corde apparaîtra comme une particule ponctuelle, et ses interactions seront décrites par une théorie quantique de particules relativistes, qui est équivalente à une théorie de champs relativistes. C'est précisément en ce sens que la relativité générale émerge comme une limite de la théorie des cordes. Comme il s'agit d'un point conceptuel important pour notre propos, donnons quelques détails sur cette émergence de la relativité générale à partir de la théorie des cordes.

La fonctionnelle d'action qui est utilisée en pratique pour quantifier une corde n'est pas vraiment  $-T \iint dA$  mais plutôt (comme l'a souligné A. Polyakov)

$$\frac{S}{\hbar} = -\frac{1}{4\pi\ell_s^2} \iint d^2\sigma \sqrt{-\gamma} \gamma^{ab} \partial_a X^\mu \partial_b X^\nu \eta_{\mu\nu} + \dots \quad (6.27)$$

où  $\sigma^a$ ,  $a = 0, 1$  sont deux coordonnées servant à repérer un événement sur la surface d'espace-temps tracée par la corde dans l'espace-temps ambiant;  $\gamma_{ab}$  est une métrique auxiliaire ( $d\Sigma^2 = \gamma_{ab}(\sigma) d\sigma^a d\sigma^b$ ) définie sur cette surface ( $\gamma^{ab}$  étant son inverse, et  $\gamma$  son déterminant); et  $X^\mu(\sigma^a)$  définit le plongement de la corde dans l'espace-temps (plat) ambiant. Les points de suspension indiquent des termes supplémentaires, et notamment des termes du type fermionique qui ont été introduits par P. Ramond, par A. Neveu et J.H. Schwarz, et par d'autres. Si l'on sépare les deux coordonnées  $\sigma^a = (\sigma^0, \sigma^1)$  en coordonnée temporelle,  $\tau \equiv \sigma^0$ , et coordonnée spatiale  $\sigma \equiv \sigma^1$ , la configuration à un « moment  $\tau$  » de la corde est décrite par des fonctions  $X^\mu(\tau, \sigma)$ , où l'on peut penser que  $\sigma$  est une abscisse curviligne décrivant l'extension spatiale de la corde. Si l'on considère une corde fermée, c'est-à-dire topologiquement équivalente à un cercle, la fonction  $X^\mu(\tau, \sigma)$  doit être périodique en  $\sigma$ . On démontre que (modulo l'imposition de certaines contraintes) on peut choisir les coordonnées  $\tau$  et  $\sigma$  sur la corde de sorte que  $d\Sigma^2 = -d\tau^2 + d\sigma^2$ . Alors les équations de la dynamique de la corde (obtenues en minimisant l'action (6.27)) se réduisent à l'équation habituelle des ondes sur une corde :  $-\partial^2 X^\mu / \partial \tau^2 + \partial^2 X^\mu / \partial \sigma^2 = 0$ . La solution générale de cette équation décrit une superposition d'ondes se déplaçant le long de la corde dans les deux sens possibles :  $X^\mu = X_L^\mu(\tau + \sigma) + X_R^\mu(\tau - \sigma)$ . Si l'on considère une corde fermée (c'est-à-dire topologiquement équivalente à un cercle), ces deux types d'ondes sont indépendants l'un de l'autre. Pour une corde ouverte (avec certaines

conditions de réflexion aux extrémités de la corde) ces deux types d'ondes sont reliés l'un à l'autre. De plus, comme, dans les deux cas, la corde à une longueur finie, on peut décomposer les ondes progressives  $X_L^\mu(\tau + \sigma)$  ou  $X_R^\mu(\tau - \sigma)$  en série de Fourier. Par exemple, on peut écrire pour une corde fermée

$$X^\mu(\tau, \sigma) = X_0^\mu(\tau) + \frac{i}{\sqrt{2}} \ell_s \sum_{n=1}^{\infty} \left( \frac{a_n^\mu}{\sqrt{n}} e^{-2in(\tau-\sigma)} + \frac{\tilde{a}_n^\mu}{\sqrt{n}} e^{-2in(\tau+\sigma)} \right) + \text{h.c.} \quad (6.28)$$

Ici  $X_0^\mu(\tau) = x^\mu + 2\ell_s^2 p^\mu \tau$  désigne le mouvement du centre de masse de la corde, et le reste désigne la décomposition du mouvement autour du centre de masse en un ensemble discret de modes d'oscillation. Comme toute corde qui vibre, une corde relativiste peut vibrer dans son mode fondamental ( $n = 1$ ) ou dans une « harmonique » du mode fondamental (entier  $n > 1$ ). Dans le cas classique les coefficients complexes  $a_n^\mu, \tilde{a}_n^\mu$  représentent les amplitudes (complexes) de vibration des modes d'oscillation de fréquence  $n$  fois le mode fondamental. ( $a_n^\mu$  correspondant à une onde se déplaçant vers la droite, alors que  $\tilde{a}_n^\mu$  correspond à une onde se déplaçant vers la gauche.) Quand on quantifie la dynamique de la corde la position de la corde  $X^\mu(\tau, \sigma)$  devient un opérateur (agissant dans l'espace des états quantiques du système), et du coup les quantités  $x^\mu, p^\mu, a_n^\mu$  et  $\tilde{a}_n^\mu$  dans (6.28) deviennent des opérateurs. La notation h.c. signifie que l'on doit ajouter les conjugués hermitiens des termes d'oscillations, qui vont contenir les opérateurs  $(a_n^\mu)^\dagger$  et  $(\tilde{a}_n^\mu)^\dagger$ . (La notation  $\dagger$  indique une conjugaison hermitienne, c'est à dire l'analogie, au niveau des opérateurs, de la conjugaison complexe.) On trouve alors que les opérateurs  $x^\mu, p^\mu$  décrivant le mouvement du centre de masse satisfont les relations de commutation habituelles d'une particule relativiste,  $[x^\mu, p^\mu] = i\hbar\eta^{\mu\nu}$ , et que les opérateurs  $a_n^\mu$  et  $\tilde{a}_n^\mu$  deviennent des opérateurs d'annihilation, comme ceux qui apparaissent dans la théorie quantique de tout système vibrant :  $[a_n^\mu, (a_m^\nu)^\dagger] = \hbar\eta^{\mu\nu} \delta_{nm}$ ,  $[\tilde{a}_n^\mu, (\tilde{a}_m^\nu)^\dagger] = \hbar\eta^{\mu\nu} \delta_{mn}$ . Dans le cas d'une corde ouverte, on a seulement un jeu d'oscillateurs, disons  $a_n^\mu$ . La discussion jusqu'ici a négligé de préciser que les amplitudes d'oscillations  $a_n^\mu, \tilde{a}_n^\mu$  devaient satisfaire un nombre infini de contraintes (liées à l'équation obtenue en minimisant (6.27) par rapport à la métrique auxiliaire  $\gamma_{ab}$ ). On peut les satisfaire en exprimant deux des composantes d'espace-temps des oscillateurs  $a_n^\mu, \tilde{a}_n^\mu$  (pour chaque  $n$ ) en fonction des autres. De coup, les états physiques de la corde sont décrits par des oscillateurs  $a_n^i, \tilde{a}_n^i$  où l'indice  $i$  ne prend que  $D - 2$  valeurs dans un espace-temps de dimension  $D$ . Oubliant pour le moment cette subtilité (qui est cependant physiquement cruciale), concluons cette discussion en résumant le « spectre » d'une corde quantique, c'est-à-dire l'ensemble des états quantiques de mouvement d'une corde.

Pour une corde ouverte, l'ensemble des états quantiques décrit les états de mouvement (impulsion  $p^\mu$ ) d'un ensemble infini de particules relativistes, ayant des masses carrées  $M^2 = -\eta_{\mu\nu} p^\mu p^\nu$  égales à  $(N - 1) m_s^2$ , où  $N$  est un nombre entier, et où  $m_s \equiv \hbar/\ell_s$  est la masse fondamentale de la théorie des cordes associée à la longueur fondamentale  $\ell_s$ . Pour une corde fermée,



on trouve une autre « tour infinie » de particules de plus en plus massives, avec cette fois  $M^2 = 4(N - 1)m_s^2$ . Dans les deux cas l'entier  $N$  est donné, en fonction des amplitudes d'oscillations (se déplaçant à droite) de la corde, par

$$N = \sum_{n=1}^{\infty} n \eta_{\mu\nu} (a_n^\mu)^\dagger a_n^\nu \quad (6.29)$$

Dans le cas d'une corde fermée on doit aussi satisfaire la contrainte  $N = \tilde{N}$  où  $\tilde{N}$  est l'opérateur obtenu en remplaçant  $a_n^\mu$  par  $\tilde{a}_n^\mu$  dans (6.29).

Le résultat précédent dit essentiellement que l'énergie interne (quantifiée) d'oscillation d'une corde définit la masse carrée de la particule associée. La présence du terme additionnel  $-1$  dans les formules données ci-dessus pour  $M^2$  signifie que l'état quantique d'énergie interne minimale d'une corde, c'est-à-dire l'état de « vide »  $|0\rangle$  où tous les oscillateurs sont dans leur état fondamental,  $a_n^\mu |0\rangle = 0$ , correspond à une masse carrée négative ( $M^2 = -m_s^2$  pour la corde ouverte et  $M^2 = -4m_s^2$  pour la corde fermée). Cet état quantique inhabituel (un « tachyon ») correspond à une instabilité de la théorie des cordes bosoniques. Il est absent des versions plus sophistiquées des théories des cordes (« supercordes ») dues à F. Gliozzi, J. Scherk et D. Olive, à M. Green et J.H. Schwarz, et à D. Gross et al. Concentrons-nous sur les autres états (qui sont les seuls à avoir des correspondants dans les théories de supercordes). On trouve alors que le premier état quantique physique possible ( $N = 1$ , avec la contrainte  $N = \tilde{N}$ ) décrit une particule de masse nulle. En théorie quantique relativiste on sait que toute particule est l'excitation d'un champ. Donc les particules de masse nulle qui apparaissent en théorie des cordes doivent correspondre à des champs à longue portée. Pour savoir quels sont les champs qui apparaissent ainsi il faut regarder de plus près quelles combinaisons possibles d'excitations des oscillateurs  $a_1^\mu, a_2^\mu, a_3^\mu, \dots$ , apparaissant dans la formule (6.29), peuvent aboutir à  $N = 1$ . À cause du facteur  $n$  multipliant dans (6.29) la contribution de l'harmonique d'ordre  $n$  à la masse carrée, seuls les oscillateurs du mode fondamental  $n = 1$  peuvent donner  $N = 1$ . On en déduit que les états quantiques internes des particules de masse nulle apparaissant dans la théorie des *cordes ouvertes* sont décrits par un état d'oscillation de la corde du type

$$\zeta_\mu (a_1^\mu)^\dagger |0\rangle \quad (6.30)$$

En revanche, à cause de la contrainte  $N = \tilde{N} = 1$ , les états quantiques internes des particules de masse nulle apparaissant dans la théorie des *cordes fermées* sont décrits par un état d'excitation contenant à la fois une oscillation allant à gauche, et une oscillation allant à droite :

$$\zeta_{\mu\nu} (a_1^\mu)^\dagger (\tilde{a}_1^\nu)^\dagger |0\rangle \quad (6.31)$$

Dans les équations (6.30) et (6.31) l'état  $|0\rangle$  désigne l'état fondamental de tous les oscillateurs ( $a_n^\mu |0\rangle = \tilde{a}_n^\mu |0\rangle = 0$ ).

L'état (6.30) décrit donc une particule de masse nulle (impulsion satisfaisant  $\eta_{\mu\nu} p^\mu p^\nu = 0$ ), possédant une « structure interne » décrite par une polarisation vectorielle  $\zeta_\mu$ . On reconnaît là exactement la description d'un photon, c'est-à-dire l'état quantique associé à une onde  $A_\mu(x) = \zeta_\mu \exp(i k_\lambda x^\lambda)$ , où  $p^\mu = \hbar k^\mu$ . La théorie des cordes ouvertes contient donc la théorie de Maxwell. (On démontre aussi que, à cause des contraintes mentionnées brièvement ci-dessus, la polarisation  $\zeta_\mu$  doit être transverse,  $k^\mu \zeta_\mu = 0$ , et qu'elle n'est définie qu'à une transformation de jauge près :  $\zeta'_\mu = \zeta_\mu + a k_\mu$ .) L'état (6.31), quant à lui, décrit une particule de masse nulle ( $\eta_{\mu\nu} p^\mu p^\nu = 0$ ), possédant une « structure interne » décrite par une polarisation tensorielle  $\zeta_{\mu\nu}$ . L'onde plane associée à une telle particule est donc de la forme  $\bar{h}_{\mu\nu}(x) = \zeta_{\mu\nu} \exp(i k_\lambda x^\lambda)$ , où  $p^\mu = \hbar k^\mu$ . Comme pour le cas de la corde ouverte, on démontre que  $\zeta_{\mu\nu}$  doit être transverse,  $\zeta_{\mu\nu} k^\nu = 0$  et qu'il n'est défini qu'à une transformation de jauge près,  $\zeta'_{\mu\nu} = \zeta_{\mu\nu} + k_\mu a_\nu + k_\nu b_\mu$ . On voit apparaître ici le même genre de structure que pour les ondes planes en relativité générale. Mais on a ici une structure plus riche qu'en relativité générale. En effet, comme l'état (6.31) est obtenu en combinant deux états d'oscillations indépendants,  $(a_1^\mu)^\dagger$  et  $(\tilde{a}_1^\mu)^\dagger$ , le tenseur de polarisation  $\zeta_{\mu\nu}$  n'est pas contraint à être symétrique. De plus, il n'est pas contraint à avoir une trace nulle. Alors, si l'on décompose  $\zeta_{\mu\nu}$  dans ses parties irréductibles possibles (une partie symétrique et sans trace, une partie symétrique avec trace, et une partie antisymétrique) on trouve que le champ  $\bar{h}_{\mu\nu}(x)$  associé aux états à masse nulle d'une corde fermée se décompose : (i) en un champ  $h_{\mu\nu}(x)$  (« graviton ») représentant une onde gravitationnelle faible en relativité générale ; (ii) un champ scalaire  $\Phi(x)$  (appelé le « dilaton ») ; et (iii) un champ tensoriel antisymétrique  $B_{\mu\nu}(x) = -B_{\nu\mu}(x)$  soumis à l'invariance de jauge  $B'_{\mu\nu}(x) = B_{\mu\nu}(x) + \partial_\mu a_\nu(x) - \partial_\nu a_\mu(x)$ . De plus, quand on étudie les interactions non linéaires entre ces divers champs, telles qu'elles sont prédites par les amplitudes de transition  $A(f, i)$  en théorie des cordes, on démontre que le champ  $h_{\mu\nu}(x)$  représente vraiment une déformation de la géométrie plate de l'espace-temps de fond où l'on définit initialement la théorie. Insistons sur ce résultat remarquable. On est parti d'une théorie qui étudiait la dynamique quantique d'une corde dans un espace-temps rigide de fond. Cette théorie prédit que *certaines excitations quantiques d'une corde* (qui se propagent à la vitesse de la lumière) *représentent en fait des ondes de déformation de la géométrie de l'espace-temps*. En termes intuitifs, l'« élasticité » de l'espace-temps postulé par la théorie de la relativité générale apparaît ici comme dû à certaines vibrations internes d'un objet élastique étendu selon une dimension spatiale.

Une autre conséquence suggestive de la théorie des cordes est le lien suggéré par la comparaison entre (6.30) et (6.31). Grosso modo, l'équation (6.31) dit que l'état interne d'une corde fermée correspondant à un graviton est construit en faisant le produit (tensoriel) d'états correspondant à des photons dans la théorie des cordes ouvertes. Ce lien inattendu entre la gravitation einsteinienne ( $g_{\mu\nu}$ ) et la théorie de Maxwell ( $A_\mu$ ) se traduit, quand on regarde les interactions en théories des cordes, par des identités remarquables (dues à

H. Kawai, D.C. Lewellen et S.-H.H. Tye) entre les amplitudes de transition de cordes ouvertes et celles des cordes fermées. Cette affinité entre électromagnétisme, ou plutôt théorie de Yang-Mills, et gravitation a récemment donné lieu à de fascinantes conjectures (dues à A. Polyakov et J. Maldacena) reliant des théories quantiques de Yang-Mills en espace-temps plat à des limites quasi-classiques de la théorie des cordes et de la gravitation en espace-temps courbe. Einstein aurait certainement été intéressé de voir comment la relativité générale classique y est utilisée pour éclairer la limite d'une théorie *quantique* de Yang-Mills.

Ayant expliqué le point de départ de la théorie des cordes, nous pouvons indiquer pour quelle raison intuitive cette théorie évite les problèmes d'intégrales divergentes qui apparaissaient quand on essayait de quantifier directement la gravitation. On a vu que la théorie des cordes contient une tour infinie de particules de masses croissant avec le degré d'excitation des oscillateurs internes de la corde. Le champ gravitationnel apparaît dans la limite où l'on considère les interactions à basse énergie ( $E \ll m_s$ ) entre les états de masse nulle de la théorie. Dans cette limite le graviton (c'est-à-dire la particule associée au champ gravitationnel) est traité comme une particule « ponctuelle ». Quand on considère des processus plus compliqués (à une boucle,  $\ell = 1$ , voir ci-dessus), il peut apparaître des gravitons virtuels élémentaires d'énergie arbitrairement élevée. Ce sont ces gravitons virtuels de haute énergie qui sont responsables des divergences. Mais, en théorie des cordes, quand on considère n'importe quel processus intermédiaire où apparaissent des énergies élevées, il faut se rappeler que cette énergie intermédiaire élevée peut aussi être utilisée pour exciter l'état interne des gravitons virtuels, et révéler ainsi le fait qu'ils sont « faits » d'une corde étendue. Une analyse de ce fait montre que la théorie des cordes introduit une troncation effective du type  $E \lesssim m_s$  sur l'énergie des particules virtuelles échangées. En d'autres termes, le fait qu'il n'y ait pas de particules vraiment « ponctuelles » en théorie des cordes, mais seulement des excitations de cordes ayant une longueur caractéristique  $\sim \ell_s$ , élimine le problème des infinis liés à des échelles d'espace et de temps arbitrairement petites. Du coup on peut calculer en théorie des cordes les amplitudes de transition correspondant à la collision entre deux gravitons, et on trouve que le résultat est donné par une intégrale finie [24].

Nous n'avons considéré jusqu'ici que le point de départ de la théorie des cordes. Cette théorie est complexe et elle est toujours en développement rapide. Indiquons brièvement quelques autres aspects de cette théorie qui sont pertinents pour notre exposé centré autour de la gravitation relativiste. Disons d'abord que les versions plus sophistiquées de la théorie des cordes (« supercordes ») requièrent l'inclusion d'oscillateurs fermioniques  $b_n^\mu, \tilde{b}_n^\mu$ , en plus des oscillateurs bosoniques  $a_n^\mu, \tilde{a}_n^\mu$  introduits ci-dessus. On trouve alors qu'il n'y a pas de particules de masse carrée négative, et que la dimension  $D$  de l'espace-temps doit être égale à 10. On trouve aussi que les états de masse nulle contiennent plus d'états que ceux indiqués ci-dessus. En fait, on trouve que les champs correspondant à ces états décrivent les diverses théories possibles

de la supergravité en  $D = 10$ . Récemment (travaux de J. Polchinski) on a aussi compris que la théorie des cordes contenait non seulement les états d'excitation de cordes (c'est-à-dire des objets étendus spatialement dans une direction), mais aussi les états d'excitation d'objets étendus spatialement dans  $p$  directions, où l'entier  $p$  peut prendre d'autres valeurs que 1. Par exemple,  $p = 2$  correspond à une *membrane*. Il semble même (selon C. Hull et P. Townsend) que l'on doit considérer qu'il y a une sorte de « démocratie » entre plusieurs valeurs de  $p$ . Un objet étendu dans  $p$  directions spatiales s'appelle un «  $p$ -brane ». En général, la masse des états quantiques de ces  $p$ -branes est très élevée, étant paramétriquement plus élevée que la masse caractéristique  $m_s$ . Cependant, on a aussi pu considérer une limite où la masse de certaines  $p$ -branes tend vers zéro. Dans cette limite, les champs associés à ces  $p$ -branes deviennent à longue portée. Un résultat surprenant (E. Witten) est que, dans cette limite, la tour infinie des états de certaines  $p$ -branes (pour  $p = 0$ ) correspond exactement à la tour infinie d'états qui apparaissent quand on considère la supergravité maximale en  $D = 11$  dimensions, avec la onzième dimension (spatiale) compactifiée sur un cercle (c'est-à-dire avec une condition de périodicité sur  $x^{11}$ ). Autrement dit, dans une certaine limite, une théorie des supercordes en  $D = 10$ , se transforme en une théorie qui vit en  $D = 11$  dimensions ! Du coup, beaucoup d'experts de la théorie des cordes pensent que la vraie définition (encore à trouver) de la théorie des cordes doit partir d'une théorie (à définir) en 11 dimensions (« théorie  $M$  »).

On a vu au paragraphe 6.8, qu'un point de contact entre la gravitation relativiste et la théorie quantique était le phénomène d'émission thermique des trous noirs découvert par S.W. Hawking. La théorie des cordes a apporté une lumière nouvelle sur ce phénomène, ainsi que sur le concept d'« entropie » d'un trou noir. La question essentielle que le calcul de S.W. Hawking laissait dans l'ombre est : quel est le sens physique de la quantité  $S$  définie par l'équation (6.19) ? Dans la théorie thermodynamique des corps usuels, l'entropie d'un système est interprétée, depuis Boltzmann, comme le logarithme (naturel) du nombre d'états microscopiques  $N$  ayant les mêmes caractéristiques macroscopiques (énergie, volume, ...) que l'état du système considéré :  $S = \log N$ . Bekenstein avait essayé d'estimer le nombre d'états microscopiques internes d'un trou noir défini macroscopiquement, et avait argumenté pour un résultat tel que  $\log N$  soit de l'ordre de grandeur de  $A/\hbar G$ , mais ses arguments restaient indirects et ne permettaient pas d'attribuer un sens clair à ce comptage d'états microscopiques. Les travaux d'A. Sen, et A. Strominger et C. Vafa, ainsi que de C.G. Callan et J.M. Maldacena ont, pour la première fois, donné des exemples de trous noirs dont la description microscopique en théorie des cordes est suffisamment précise pour permettre d'en calculer (dans certaines limites) le nombre d'états quantiques internes,  $N$ . Il est alors très satisfaisant de trouver un résultat final pour  $N$  dont le logarithme est *précisément* égal à l'expression (6.19). Cependant, il reste des zones d'ombre dans la compréhension de la structure quantique des trous noirs. En particulier, les calculs de théorie des cordes permettant de donner un sens statistique précis à

l'entropie (6.19) concernant des trous noirs très spéciaux (dits « extrémaux », c'est-à-dire qui ont la charge électrique maximale que peut supporter un trou noir pour garder un horizon régulier). Ces trous noirs ont une température de Hawking égale à zéro, et n'émettent donc pas de rayonnement thermique. Ils correspondent à des états *stables* dans la théorie quantique. On aimerait cependant comprendre aussi en détail la structure quantique interne des trous noirs *instables*, comme le trou noir de Schwarzschild (6.17), qui a une température non nulle, et qui donc perd petit à petit de la masse sous forme de rayonnement thermique. Quel est l'état final auquel conduit ce processus graduel d'« évaporation » d'un trou noir ? Est-ce qu'un état quantique pur initial rayonne toute sa masse initiale pour se transformer entièrement en rayonnement thermique incohérent ? Ou, un trou noir de Schwarzschild se transforme-t-il, après avoir atteint une taille minimum, en quelque chose d'autre ? Les réponses à ces questions restent en grande partie ouvertes, même s'il a été argumenté qu'un trou noir de Schwarzschild se transforme en un état quantique très massif d'une corde quand son rayon devient de l'ordre de  $\ell_s$  [25].

Nous avons vu plus haut que la théorie des cordes contient la relativité générale dans une certaine limite. En même temps, la théorie des cordes est (stricto sensu) infiniment plus riche que la gravitation einsteinienne, car le graviton n'est qu'une excitation quantique particulière d'une corde, parmi un nombre infini d'autres. Quelles déviations à la gravitation einsteinienne prédit la théorie des cordes ? Cette question reste ouverte aujourd'hui à cause de notre manque de compréhension du lien entre la théorie des cordes et la réalité observée dans notre environnement habituel (espace-temps à 4 dimensions ; interactions électromagnétiques, faibles et fortes ; spectre observé de particules ; ...). Nous nous contenterons d'indiquer quelques possibilités. D'abord, disons que si l'on considère des processus de collision entre gravitons d'énergie-impulsion  $k$  plus petite que, mais non négligeable par rapport à, la masse caractéristique des cordes  $m_s$ , les calculs d'amplitude de transition en théorie des cordes montrent qu'il faut modifier les équations d'Einstein habituelles (en l'absence de matière)  $R_{\mu\nu} = 0$ , en incluant des corrections d'ordre  $(k/m_s)^2$ . On trouve que ces équations d'Einstein modifiées ont la forme (pour la théorie des cordes bosonique)

$$R_{\mu\nu} + \frac{1}{4} \ell_s^2 R_{\mu\alpha\beta\gamma} R_{\nu}^{\alpha\beta\gamma} + \dots = 0, \quad (6.32)$$

où

$$R_{\nu\alpha\beta}^{\mu} \equiv \partial_{\alpha} \Gamma_{\nu\beta}^{\mu} + \Gamma_{\sigma\alpha}^{\mu} \Gamma_{\nu\beta}^{\sigma} - \partial_{\beta} \Gamma_{\nu\alpha}^{\mu} - \Gamma_{\sigma\beta}^{\mu} \Gamma_{\nu\alpha}^{\sigma} \quad (6.33)$$

désigne le « tenseur de courbure » de la métrique  $g_{\mu\nu}$ . (La quantité  $R_{\mu\nu}$ , définie au paragraphe 6.5 et entrant de façon essentielle dans les équations d'Einstein, est une « trace » de ce tenseur :  $R_{\mu\nu} = R_{\mu\sigma\nu}^{\sigma}$ .) Comme il est indiqué par les points de suspension dans (6.32), les termes écrits ne sont que les deux premiers termes d'une série infinie en puissances croissantes de  $\ell_s^2 \equiv \alpha'$ . L'équation (6.32) montre comment le fait que la corde ne soit pas un point,

mais soit étendue sur une longueur caractéristique  $\sim \ell_s$ , modifie la description einsteinienne de la gravitation. Les corrections à l'équation d'Einstein indiquées dans (6.32) sont, cependant, tout à fait négligeables dans la plupart des applications de la relativité générale. En effet, on s'attend à ce que  $\ell_s$  soit de l'ordre de la longueur de Planck  $\ell_p$  (Éq. (6.25)). Plus précisément, on s'attend à ce que  $\ell_s$  soit de l'ordre de grandeur de  $10^{-32}$  cm. (Cependant, cette question reste ouverte, et il a été suggéré récemment que  $\ell_s$  soit beaucoup plus grand, et peut être de l'ordre de  $10^{-17}$  cm.)

Le seul domaine d'application de la relativité générale où les modifications indiquées dans (6.32) devraient jouer un rôle important est la cosmologie primordiale. En effet, près de la singularité initiale du Big Bang (si elle existe), la « courbure »  $R_{\mu\nu\alpha\beta}$  devient très grande. Quand elle atteint des valeurs comparables à  $\ell_s^{-2}$  la série infinie de corrections dans (6.32) commence à jouer un rôle comparable au premier terme, découvert par Einstein. Une telle situation se trouve aussi à l'intérieur d'un trou noir, quand on s'approche beaucoup de la singularité (voir Fig. 6.3). Malheureusement, dans de telles situations, il faudrait pouvoir tenir compte de la série infinie de termes dans (6.32), c'est-à-dire remplacer la description einsteinienne de la gravitation par une *champ* (qui correspond à une particule (quantique) *ponctuelle*) par sa description cordiste exacte. C'est un problème difficile que l'on ne sait pas vraiment attaquer aujourd'hui.

Mais la théorie des cordes prédit, a priori, des modifications à basse énergie ( $k \ll m_s$ ) plus drastiques de la relativité générale que les corrections indiquées dans (6.32). En effet, on a vu dans l'équation (6.31), que la gravitation einsteinienne  $g_{\mu\nu}(x)$  n'apparaissait pas seule en théorie des cordes. Elle est toujours nécessairement accompagnée par d'autres champs à longue portée, notamment un champ scalaire  $\Phi(x)$ , le « dilaton », et un tenseur antisymétrique  $B_{\mu\nu}(x)$ . Quels rôles ces « partenaires » du graviton jouent-ils dans la réalité observable? Cette question n'a pas de réponse claire aujourd'hui. En outre, si l'on se souvient que la théorie des (super) cordes doit vivre dans un espace-temps de dimension  $D = 10$ , et qu'elle contient la théorie de la supergravité en  $D = 10$  (et éventuellement  $D = 11$ ), il y a beaucoup d'autres champs supplémentaires qui s'ajoutent aux dix composantes du tenseur métrique habituel  $g_{\mu\nu}$  (en  $D = 4$ ). Il est concevable que tous ces champs supplémentaires (qui ont une masse nulle en première approximation de la théorie des cordes) acquièrent des masses assez grandes dans notre univers local pour ne plus propager d'effets observables sur des portées macroscopiques. Il reste cependant possible qu'un, ou plusieurs, de ces champs restent de masse (essentiellement) nulle, et puisse alors propager des effets physiques sur des portées assez grandes pour être observables. Il est donc intéressant de comprendre quels sont les effets physiques impliqués, par exemple, par le dilaton  $\Phi(x)$ , ou par  $B_{\mu\nu}(x)$ . Concernant ce dernier, il est intéressant de noter (comme l'ont souligné A. Connes, M. Douglas et A. Schwartz) que, dans une certaine limite, la présence d'un fond  $B_{\mu\nu}(x)$  a l'effet de déformer de façon « non commutative » la géométrie de l'espace-temps. Cela veut dire que, en

un certain sens, les coordonnées  $x^\mu$  de l'espace-temps cessent d'être de simples nombres réels (commutants) pour devenir des quantités non-commutantes :  $x^\mu x^\nu - x^\nu x^\mu = \varepsilon^{\mu\nu}$  où  $\varepsilon^{\mu\nu} = -\varepsilon^{\nu\mu}$  est lié à un  $B_{\mu\nu}$  (uniforme) de fond. Considérons pour finir, l'autre partenaire obligé du graviton  $g_{\mu\nu}(x)$ , le dilaton  $\Phi(x)$ . Ce champ joue un rôle central en théorie des cordes. En effet, la valeur moyenne (dans le vide) du dilaton détermine la constante de couplage de la théorie des cordes,  $g_s = e^\Phi$ . À son tour, la valeur de  $g_s$  détermine (avec d'autres champs) les constantes de couplage de la physique. Par exemple, la constante de couplage de la gravitation est donnée par une formule du type  $\hbar G = \ell_s^2(g_s^2 + \dots)$  où les points de suspension désignent des termes correctifs (qui peuvent devenir très importants si  $g_s$  n'est pas très petit). De même, la constante de structure fine,  $\alpha = e^2/\hbar c \simeq 1/137$ , qui détermine l'intensité des interactions électromagnétiques est une fonction de  $g_s^2$ . À cause de ces relations entre les constantes de couplages de la physique et  $g_s$  (et donc la valeur du dilaton ;  $g_s = e^\Phi$ ), on voit que si le dilaton est à masse nulle (c'est-à-dire à longue portée), sa valeur  $\Phi(x)$  en un point  $x$  de l'espace-temps va dépendre de la distribution de matière dans l'univers. Par exemple, comme c'est le cas pour le champ gravitationnel (confer  $g_{00}(x) \simeq -1 + 2GM/c^2 r$ ), on s'attendrait à ce que la valeur de  $\Phi(x)$  dépende des masses présentes autour du point  $x$ , et soit différente à la surface de la Terre et à une altitude élevée. On pourrait aussi s'attendre à ce que  $\Phi(x)$  soit sensible à l'expansion de l'univers et varie sur une échelle de temps comparable à l'âge de l'univers. Mais si  $\Phi(x)$  varie dans l'espace et/ou dans le temps, on conclut des relations évoquées ci-dessus entre  $g_s = e^\Phi$  et les constantes de couplage de la physique que celles-ci devraient aussi varier dans l'espace et/ou dans le temps. Donc, par exemple, la valeur de la constante de structure fine  $\alpha$  pourrait être légèrement différente de sa valeur actuelle dans une galaxie très lointaine (et donc observée à un instant très reculé). De tels effets sont accessibles à des observations astronomiques fines et, d'ailleurs, certaines observations récentes ont suggéré que les constantes d'interaction sont différentes dans des galaxies lointaines. Cependant, d'autres données expérimentales (réacteur nucléaire fossile d'Oklo, composition isotopique de météorites terrestres anciens) mettent des limites très sévères sur une variabilité éventuelle des « constantes » de couplage. Indiquons enfin que si la « constante » de structure fine  $\alpha$ , et d'autres « constantes » de couplage, varient avec un champ sans masse comme le dilaton  $\Phi(x)$ , alors ceci implique une violation du postulat de base de la relativité générale : le principe d'équivalence. En particulier, on démontre que l'universalité de la chute libre est nécessairement violée, c'est-à-dire que des corps de composition nucléaire différente doivent tomber, dans un champ gravitationnel extérieur, avec des accélérations différentes. Ceci donne une importante motivation pour tester le principe d'équivalence avec une précision accrue. Par exemple, la mission spatiale MICROSCOPE [26] (du CNES) devrait bientôt tester l'universalité de la chute libre au niveau  $10^{-15}$ , et le projet spatial STEP (Satellite Test of the Equivalence Principle) [27] pourrait atteindre le niveau  $10^{-18}$ .

## 6.12 Conclusion

Pendant longtemps la relativité générale était admirée comme une merveilleuse construction intellectuelle, mais ne jouait qu'un rôle marginal en physique. Typique de l'appréciation de cette théorie est la phrase de Max Born [28] prononcée lors du cinquantième anniversaire de l'*annus mirabilis* : « *Les fondations de la relativité générale m'apparaissaient alors, et encore aujourd'hui, comme le plus grand exploit de la pensée humaine quant à la Nature, la plus stupéfiante association de pénétration philosophique, d'intuition physique et d'habileté mathématique. Mais ses liens à l'expérience étaient ténus. Cela me séduisait comme une grande œuvre d'art que l'on doit apprécier et admirer à distance.* »

Aujourd'hui, cent ans après l'*annus mirabilis*, la situation est très différente. La relativité générale joue un rôle central dans un large domaine de la physique, allant de la cosmologie primordiale à la physique des trous noirs, en passant par l'observation des pulsars binaires et la définition du temps atomique international. Elle a même des applications pratiques quotidiennes, via les systèmes de positionnement par satellite (comme GPS, et bientôt son homologue européen Galileo). Beaucoup de projets expérimentaux ambitieux (et coûteux) visent à la tester (GPB, MICROSCOPE, STEP, ...), ou l'utilisent comme outil pour déchiffrer l'univers lointain (LIGO/VIRGO/GEO, LISA, ...). Loin est donc le temps où ses liens à l'expérience étaient ténus. Cependant, il est remarquable de constater que la fascination, évoquée par Born, pour la structure et les implications physiques de la théorie reste entière. L'une des motivations pour penser que la théorie des cordes (et d'autres objets étendus) contient la clef du problème de l'unification de la physique est son affinité profonde avec la relativité générale. En effet, alors que les essais de « Grande Unification » des années soixante-dix laissaient complètement de côté l'interaction gravitationnelle, la théorie des cordes conduit nécessairement au concept einsteinien fondamental d'un espace-temps dynamique. Il semble d'ailleurs qu'il faudra comprendre plus profondément l'espèce de « géométrie généralisée quantique » engendrée par l'interaction de cordes et de  $p$ -branes pour pouvoir formuler complètement cette théorie, et comprendre ses symétries cachées et ses implications physiques. Einstein apprécierait sans doute de voir le rôle central joué par les principes de symétrie au sein de la physique moderne.

## 6.13 Références

- [1] A. Einstein, *Zur Elektrodynamik bewegter Körper*, *Annalen der Physik* **17**, 891 (1905).
- [2] *Albert Einstein, Œuvres choisies*, sous la dir. de F. Balibar, Le Seuil/CNRS, Paris (1993). Le lecteur français y aura accès aux articles principaux d'Einstein. Voir notamment les tomes 2 (Relativités I) et 3



(Relativités II). Pour un accès en ligne de beaucoup d'articles originaux d'Einstein, et de documents sur lui, voir <http://www.alberteinstein.info/>. Notons aussi que l'essentiel des travaux en cours sur la relativité générale peut être consulté sur les diverses archives de <http://xxx.lanl.gov>, notamment l'archive gr-qc. Des articles de revues sur certains sous-domaines de la relativité générale sont accessibles sur <http://relativity.livingreviews.org>.

- [3] Galilée, *Discours et démonstrations mathématiques concernant deux sciences nouvelles*, Maurice Clavelin éd., Armand Colin, Paris (1970).
- [4] Le lecteur intéressé à connaître les tests expérimentaux récents des théories de la gravitation pourra consulter sur internet, soit la revue très détaillée de C.M. Will dans Living Reviews (<http://relativity.livingreviews.org/Articles/lrr-2001-4>), soit la revue brève de T. Damour dans la Review of Particle Physics (<http://pdg.lbl.gov/>).
- [5] A. Einstein, *Die Feldgleichungen der Gravitation*, Sitz. Preuss. Akad. Wiss. (1915), p. 844.
- [6] Le lecteur désirant étudier en détail le formalisme de la théorie de la relativité générale, et ses applications, pourra consulter, par exemple, les ouvrages suivants :  
 L. Landau, E. Lifchitz, *Théorie des champs*, Mir, Moscou (1989).  
 S. Weinberg, *Gravitation and Cosmology*, Wiley, New York (1972).  
 H.C. Ohanian, R. Ruffini, *Gravitation and Spacetime*, Second Edition, Norton, New York (1994).  
 N. Straumann, *General Relativity, With Applications to Astrophysics*, Springer-Verlag (2004).  
 Signalons aussi les notes de cours détaillées, sur la relativité générale, de S.M. Carroll, disponibles sur internet : <http://pancake.uchicago.edu/~carroll/notes/> ; ou bien sur gr-qc/9712019.  
 Enfin, signalons le livre récent (en français) sur l'histoire de la découverte et de la réception de la relativité générale :  
 J. Eisenstaedt, *Einstein et la relativité générale*, CNRS, Paris (2002).
- [7] B. Bertotti, L. Iess, P. Tortora, *A test of general relativity using radio links with the Cassini spacecraft*, Nature **425**, 374 (2003).
- [8] <http://einstein.stanford.edu>
- [9] W. Israel, *Dark stars: the evolution of an idea*, in *300 Years of Gravitation*, S.W. Hawking et W. Israel édés., Cambridge University Press, Cambridge (1987), pp. 199–276.
- [10] La découverte des pulsars binaires est relatée dans la conférence Nobel de Hulse :  
 R.A. Hulse, Review of Modern Physics **66**, 699 (1994).
- [11] Pour une introduction aux caractéristiques observationnelles des pulsars, et à leur utilisation pour tester la gravitation relativiste, voir la conférence Nobel de Taylor :  
 J.H. Taylor, Review of Modern Physics **66**, 711 (1994).

- [12] Pour une mise à jour des caractéristiques observationnelles des pulsars, et de leur utilisation pour tester la relativité générale, voir la Living Review d'I.H. Stairs, disponible sur <http://relativity.livingreviews.org/Articles/lrr-2003-5/>.
- [13] Pour une mise à jour récente des tests de la gravitation relativiste (et des théories tenseur-scalaires) obtenus par chronométrage de pulsars binaires, voir : G. Esposito-Farèse, gr-qc/0402007 (disponible sur l'archive general relativity and quantum cosmology à l'adresse <http://xxx.lanl.gov>), et T. Damour et G. Esposito-Farèse, en préparation. La figure 6.4 est inspirée de ces références.
- [14] Pour une revue du problème du mouvement de deux corps gravitationnellement condensés en relativité générale, jusqu'au niveau où apparaissent les effets liés à la propagation à vitesse finie de l'interaction gravitationnelle, voir :  
T. Damour, *The problem of motion in Newtonian and Einsteinian gravity*, dans *300 Years of Gravitation*, S.W. Hawking et W. Israel édés., Cambridge University Press, Cambridge (1987), pp. 128–198.
- [15] A. Einstein, *Näherungsweise Integration der Feldgleichungen der Gravitation*, Sitz. Preuss. Akad. Wiss. (1916), p. 688 ; ibidem, *Über Gravitationswellen* (1918), p. 154.
- [16] Pour une introduction très détaillée à ces trois problèmes, voir :  
K.S. Thorne, *Gravitational radiation*, dans *300 Years of Gravitation*, S.W. Hawking et W. Israel édés., Cambridge University Press, Cambridge, 1987, pp. 330–458.
- [17] <http://www.ligo.caltech.edu/>
- [18] <http://www.virgo.infn.it/>
- [19] <http://www/geo600.uni-hanover.de/>
- [20] <http://lisa.jpl.nasa.gov/>
- [21] L. Blanchet et al., gr-qc/0406012 ; voir aussi la Living Review de L. Blanchet, disponible sur <http://relativity.livingreviews.org/Articles>.
- [22] La figure 6.5 est inspirée d'un travail d'A. Buonanno et T. Damour : gr-qc/0001013.
- [23] Pour un exposé particulièrement clair du développement de la théorie quantique des champs, voir (entre autres) le premier chapitre de :  
S. Weinberg, *The Quantum Theory of Fields*, vol. 1, *Foundations*, Cambridge University Press, Cambridge (1995).
- [24] Pour une introduction à la théorie des (super)cordes voir <http://superstringtheory.com/>. Pour une introduction détaillée (et technique) à la théorie voir les livres :  
B. Zwiebach, *A First Course in String Theory*, Cambridge University Press, Cambridge (2004).  
M.B. Green, J.H. Schwarz et E. Witten, *Superstring theory*, 2 vols., Cambridge University Press, Cambridge (1987).  
J. Polchinski, *String Theory*, 2 vols., Cambridge University Press, Cambridge (1998).

Pour lire des articles de revue ou de recherche sur cette théorie en plein développement voir l'archive hep-th sur <http://xxx.lanl.gov>. Pour rechercher des informations sur la littérature en théorie des cordes (et plus généralement en physique des hautes énergies) voir aussi le site <http://www.slac.stanford.edu/spires/find/hep>.

- [25] Pour une introduction détaillée à la physique des trous noirs voir : P.K. Townsend, gr-qc/9707012 ; pour une entrée dans la vaste littérature sur l'entropie des trous noirs, voir, par exemple, T. Damour, hep-th/0401160.
- [26] <http://www.onera.fr/microscope/>
- [27] <http://www.sstd.rl.ac.uk/fundphys/step/>
- [28] M. Born, *Physics and Relativity*, dans *Fünfzig Jahre Relativitätstheorie, Bern, 11-16 Juli 1955, Verhandlungen*, A. Mercier et M. Kervaire édés., Helvetica Physica Acta, Supplement 4, 244–260 (1956).



# Chapitre 7

## Cosmologie

Par *François R. Bouchet*

### Sommaire

<b>7.1</b>	<b>Introduction</b>	<b>323</b>
7.1.1	Objet, principes, méthodes de la cosmologie contemporaine	323
7.1.2	Points de repères astronomiques	325
<b>7.2</b>	<b>Quelques étapes marquantes de la cosmologie contemporaine</b>	<b>332</b>
7.2.1	Définition des enjeux	332
7.2.2	Cosmographie, nucléosynthèse et rayonnement fossile	335
7.2.3	Cosmogonie et établissement du paradigme actuel	340
<b>7.3</b>	<b>Cosmographie</b>	<b>347</b>
7.3.1	Relativité et géométrie	348
7.3.2	Dynamique de l'Univers	352
7.3.3	La matière est souvent sombre	359
7.3.4	L'obscur énergie du vide contemporain	371
<b>7.4</b>	<b>Cosmogonie</b>	<b>379</b>
7.4.1	Évolution récente des structures	380
7.4.2	Premières phases de l'évolution des structures	383
7.4.3	Anisotropies fossiles	389
7.4.4	Observations des anisotropies du RCF	396
7.4.5	Concordance, précision et justesse	401
<b>7.5</b>	<b>Au-delà de la concordance...</b>	<b>406</b>
<b>7.6</b>	<b>Références</b>	<b>409</b>



**François R. Bouchet**

François Bouchet est directeur de recherche au CNRS, chercheur à l'Institut d'Astrophysique de Paris, président du conseil scientifique du programme national de cosmologie. Ses travaux portent sur la caractérisation théorique et observationnelle des grandes structures de l'Univers, sur les fluctuations primordiales ayant initié leur croissance et sur leur évolution dynamique sous l'influence de la gravité. Il participe à la construction d'instruments pour le satellite Planck qui cartographiera en détail les anisotropies du corps noir cosmologique après son lancement en 2007.

# Cosmologie

## 7.1 Introduction

### 7.1.1 Objet, principes, méthodes de la cosmologie contemporaine

La cosmologie a pour objet les propriétés globales de l'Univers. Elle ne s'intéresse pas aux objets de l'Univers en tant que tels, et cherche à ne retenir que ce qui n'appartient à aucun d'entre eux en particulier. C'est une science physique, au sens où on utilise la démarche, les méthodes et les lois de la physique. Mais les investigations cosmologiques doivent utiliser les méthodes de l'astronomie, qui est une science naturelle d'*observation*. Les systèmes observés sont à prendre « comme ils sont », sans pouvoir modifier à volonté leur environnement pour en comprendre les mécanismes et valider les modèles qui les décrivent. La validation scientifique est donc toujours plus délicate, puisqu'elle n'a pas accès à l'expérience directe dans des conditions contrôlées ; d'autant plus que dans le cas de la cosmologie, nous n'avons par définition qu'un seul système, qui de surcroît ne peut être observé que de l'intérieur !

La cosmologie moderne pose les principes suivants :

1. Une cosmologie scientifique a un sens. En d'autres termes, il existe effectivement des propriétés d'ensemble, à découvrir.
2. Les lois de la physique sont bien des lois universelles, qui doivent être valables en tous temps et en tous lieux.
3. L'Univers est globalement identique partout, aucun lieu d'observation n'est privilégié.

*Tous ces postulats seront jugés à leurs fruits.* À commencer par l'intelligibilité de l'Univers en tant que tel. Le deuxième principe nous enjoint d'appliquer à distance les lois validées dans notre environnement. La difficulté est que les lois de la physique sont établies, pour un certain domaine d'échelle, de temps, d'énergie, d'accélération, etc., par des expériences multiples dans ce domaine là. Or la cosmologie contemporaine est précisément une entreprise où

l'on considère des conditions jamais rencontrées auparavant et qui risquent de requérir une modification de nos lois communes dans ces conditions extrêmes. C'est une difficulté, mais c'est aussi une chance de forger une meilleure approximation de la réalité. Le troisième principe, connu sous le nom de *Principe cosmologique*, énonce l'homogénéité et l'isotropie de l'espace. Néanmoins, notre environnement immédiat est à l'évidence très inhomogène. On entend donc par homogénéité de l'espace que les propriétés moyennes d'un morceau d'univers de *suffisamment grand volume* seront identiques à celles de tout autre morceau d'univers. L'isotropie est l'identité des propriétés dans toutes les directions autour de l'observateur<sup>1</sup>.

Ce principe cosmologique représente une évolution considérable des conceptions cosmologiques, depuis notamment les idées de l'antiquité grecque où le monde était hiérarchisé en mondes sublunaire et supralunaire obéissant à des lois différentes. Si l'idée de lois universelles commence à s'imposer en Occident au seizième siècle avec Brahé, Kepler et Galilée, et si le géocentrisme est abandonné avec Copernic qui place le soleil au centre du monde, le grand pas est franchi avec Newton et sa loi de gravitation universelle qui opère dans un espace sans centre ni bord, théâtre impassible des phénomènes. Par le passé, on a aussi considéré un *Principe Cosmologique Parfait*, qui exige que le Cosmos, en plus d'être spatialement homogène, soit aussi identique à lui-même dans le temps. Ce qui implique donc un modèle stationnaire, modèle d'univers sans évolution qui eut son heure de gloire. Néanmoins, les moyens observationnels actuels permettent de constater directement l'évolution temporelle des propriétés des galaxies.

Venons-en à la question de la validation d'un modèle cosmologique. Le modèle physique d'un système a pour objectif de fournir, en fonction de lois physiques préétablies et de paramètres mesurables, une description jamais prise en défaut de ce système, au moins dans un certain domaine d'expérimentation. Pour être acceptable, il faut d'abord qu'un modèle soit cohérent, et que l'interprétation des diverses observations aboutisse à des conclusions sans contradiction. Il faut en particulier que les inférences concernant les paramètres du modèle soient compatibles entre elles et, bien sûr, que tous les paramètres soient contraints. L'Histoire du dernier siècle nous a montré que, quand le système est l'Univers lui-même, cette demande de cohérence exhaustive n'est pas si facile à satisfaire. Pour juger des vertus d'un modèle, on attache évidemment beaucoup plus de poids aux « pré-dictions » qu'aux « post-dictions » qui se contentent de fournir une explication satisfaisante de phénomènes déjà connus ou de mesures préexistantes. En effet, seules des prédictions permettent de juger si le modèle a survécu à des tests qui auraient pu l'invalider. Et comme l'a proposé le philosophe Viennois Karl Popper [1], cette notion de réfutabilité par des éléments empiriques nouveaux (outre l'exigence de reproductibilité) est l'élément crucial qui distingue une théorie scientifique d'une métaphysique. Enfin, pour qu'on puisse considérer qu'un modèle décrit

---

<sup>1</sup>L'isotropie en tous points d'observation implique d'ailleurs l'homogénéité.



bien le comportement du système, et donc qu'il offre une approximation satisfaisante d'une partie de la réalité, il faut encore qu'il y ait nettement plus de contraintes que de degrés de liberté, plus d'observables indépendantes qu'il n'y a de paramètres pour les ajuster ; et il est également nécessaire que les vérifications de cohérence testent bien tous les aspects du modèle.

Comme on le verra, ce n'est le cas en cosmologie que depuis récemment. La satisfaction de cette exigence minimum est néanmoins remarquable si l'on considère l'ambition de l'entreprise et la relative maigreur des indications empiriques dont nous disposons. Qu'on en juge. Aux grandes échelles de distance, la gravité est l'acteur dominant<sup>2</sup>. La théorie actuelle de la gravité est la relativité générale d'Einstein qui, par là, fonde la cosmologie moderne. Nous allons en décrire les avancées. Mais il faut d'abord se souvenir que si la relativité générale est bien testée à l'échelle du système solaire, sur des distances  $\sim 10^{11}$  m, il nous faut faire une extrapolation colossale sur 15 ordres de grandeur pour arriver aux échelles que les observations cosmologiques actuelles sondent. Pourtant, en utilisant cette théorie, nous avons construit un scénario qui décrit l'évolution de l'Univers et comment il s'est structuré. Nous avons aussi mesuré ses propriétés globales de plusieurs façons qui s'accordent entre elles. Cet accord est très important, car c'est lui qui permet de vérifier les liens prédits par la théorie entre de multiples observations. Outre la vérification du scénario lui-même, c'est la vérification de ces diverses prédictions qui nous permet de tester les divers aspects de la théorie. Ce texte vise à présenter au lecteur où nous en sommes aujourd'hui.

### 7.1.2 Points de repères astronomiques

La figure 7.1 montre une carte un peu spéciale de l'Univers. Établie par Richard Gott et al. [2], elle permet de répertorier dans une seule représentation nombre d'objets connus, des plus proches aux plus lointains, récemment détectés. En abscisse l'ascension droite, et en ordonnée le logarithme de l'éloignement. La partie gauche offre un raccourci saisissant en permettant d'embrasser d'un coup d'œil 20 ordres de grandeur de taille, de la surface de la Terre aux confins de l'Univers observable. La partie droite présente un agrandissement du haut de la carte qui permet de mieux détailler les structures cosmologiques les plus éloignées ( $> 10^{21}$  km).

L'ascension droite est l'équivalent céleste de la longitude terrestre ; elle se mesure en général en heures, minutes et secondes. Comme l'axe couvre 24 h, toutes les ascensions droites sont représentées et les extrêmes de l'axe des abscisses correspondent à la même direction. L'angle analogue à la latitude terrestre, la déclinaison, a été soit ignoré, soit choisi proche de zéro. Dans le premier cas, les points sont indiqués quelle que soit la déclinaison de l'objet

---

<sup>2</sup>En effet la force forte qui assure la cohésion des noyaux décroît avec la distance beaucoup plus vite que la force d'attraction gravitationnelle. Quant à la force électromagnétique, elle est proportionnelle à la charge électrique et l'on constate la neutralité électrique des grands volumes.

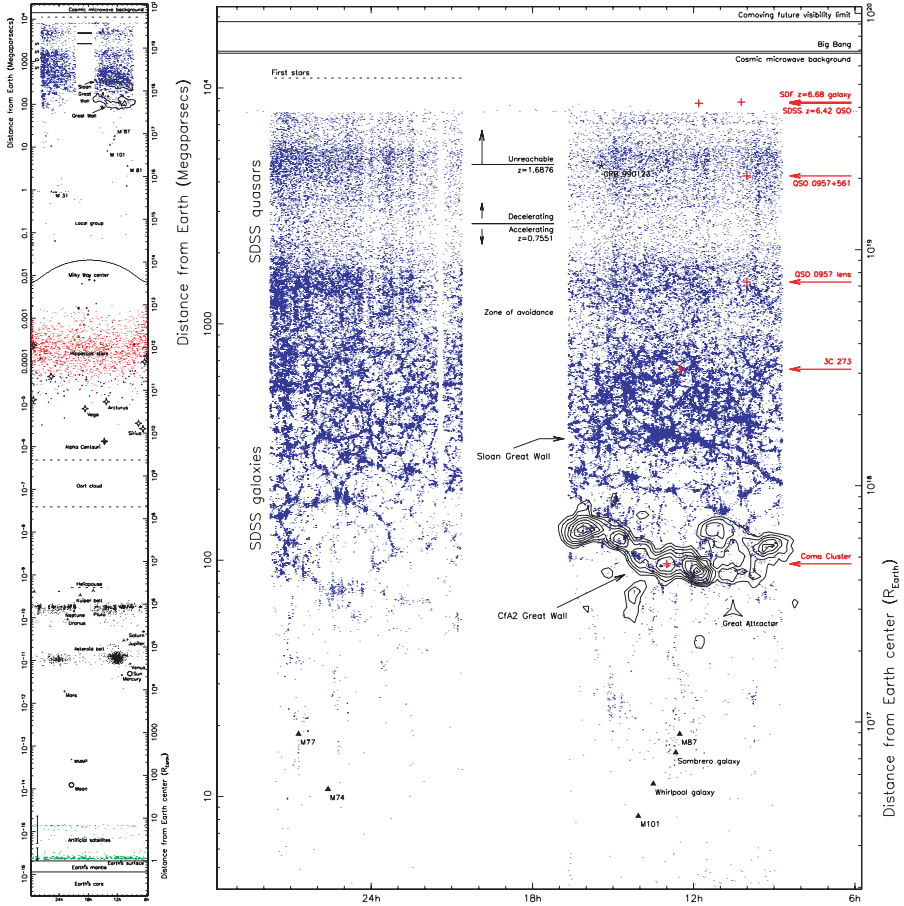


FIG. 7.1 – Carte de l'Univers répertoriant nombre d'objets connus, des plus proches aux plus lointains récemment détectés. La carte de droite est une version agrandie du haut de la carte de gauche. (D'après R. Gott et al. [2], voir aussi le site : [www.astro.princeton.edu/~mjuric/universe.](http://www.astro.princeton.edu/~mjuric/universe.))

correspondant. Dans l'autre cas, ce que l'on voit représente la disposition des objets dans une tranche fine englobant l'équateur terrestre. L'axe de droite exprime l'éloignement en unités de rayon terrestre, soit 6378 km. Dans ce système de représentation, la surface de la Terre est donc indiquée par le trait horizontal à la hauteur unité. Pour avoir une idée de l'échelle, on remarquera que la circonférence de la terre ne mesure que les quelques centimètres de la largeur de la figure, soit une échelle de l'ordre de 1 pour un milliard. Ce facteur d'échelle croît d'un facteur 10 chaque fois qu'on s'éloigne d'un facteur 10 de la surface de la terre. Pour les distances les plus grandes, en haut de la

carte, le facteur de réduction de la représentation par rapport à la taille réelle vaut près de  $10^{29}$  !

En s'éloignant de la surface terrestre, on trouve d'abord les 8420 satellites artificiels de notre planète, là où ils se trouvaient le 12 août 2003, date choisie pour représenter tous les objets de la carte (points verts). On aperçoit ensuite la lune, le satellite WMAP dont nous reparlerons, le soleil qui se trouve par définition à une unité astronomique (une ua, soit 149,6 millions de kilomètres, ou encore  $\sim 8$  minutes-lumière), les planètes proches du système solaire et la ceinture d'astéroïdes. Un peu plus loin, on distingue Uranus et Neptune et les dizaines de milliers de planétésimaux de la ceinture de Kuiper qui se trouve à près d'un million de rayons terrestres. Des marques indiquent ensuite la position des sondes Pioneer 10 et Voyager 1 et 2 approchant l'héliopause (l'extrême limite d'influence du vent solaire). Près de 100 fois plus loin, on trouve une zone appelée le nuage de Oort, du nom de l'astronome néerlandais qui proposa dans les années cinquante que cette région située aux confins du système solaire soit un « réservoir » à comètes. L'approche vers nous d'une comète correspondrait alors à la perturbation gravitationnelle accidentelle de sa lointaine orbite.

Au-delà du système solaire, la carte positionne les dix étoiles les plus brillantes du ciel, dont la plus proche, Alpha du Centaure<sup>3</sup>, est à environ un parsec. Le parsec<sup>4</sup> correspond à 206,265 ua et  $3,086 \times 10^{16}$  m (soit la distance parcourue par la lumière en 3,26 années). C'est une unité de distance fondamentale en astronomie ; elle est définie comme la distance d'une étoile ayant une seconde d'arc de parallaxe. Autrement dit, la position de l'étoile sur le ciel varie d'une seconde d'arc quand l'observateur se déplace – dans une direction perpendiculaire à la ligne de visée – d'une distance égale à la séparation moyenne Terre-Soleil. L'avantage de cette définition est que la distance d'objets relativement proches s'obtient directement, par triangulation, en observant la variation de position angulaire de l'objet à six mois d'intervalle.

Le nuage de points entre 10 et 1000 parsecs (pc en abrégé) montre la position des 3386 étoiles proches du plan de l'équateur céleste (de déclinaison  $\delta$  inférieure à 2 degrés) du catalogue de 118 218 étoiles dont le satellite Hipparcos a récemment mesuré la parallaxe avec précision (points rouges). Plus loin encore le centre de notre Galaxie, à 8 kiloparsecs, et le bord du disque d'étoiles de notre Voie lactée, repéré par une ligne. Cela montre d'ailleurs que nous sommes à peu près à mi-chemin entre le centre et le bord du disque Galactique.

Au-delà s'étend le domaine extragalactique que les astronomes ont commencé à appréhender en tant que tel vers 1926. Les galaxies les plus proches de la nôtre sont le Grand et le Petit Nuage de Magellan qui sont respectivement à 55 et 65 kiloparsecs (kpc). Elles constituent une partie de notre Groupe Local de galaxies, riche de 52 membres, dont la galaxie d'Andromède, M31. Situé à 900 kpc, c'est l'objet le plus lointain qui soit visible à l'œil nu (suivi par le Petit Nuage de Magellan). Andromède était d'ailleurs connue de l'astronome

<sup>3</sup>Alpha du Centaure est en fait un système de trois étoiles.

<sup>4</sup>Parsec vient de la contraction – en anglais – de PARalax SECond angle.

persan Abd-al-Rahman Al-Sufi, comme le « petit nuage » qu'il décrit et représenta en 964 dans son « Livre des étoiles fixes » ! Le lecteur aura peut-être remarqué que j'écris Galaxie quand il s'agit de la nôtre, et galaxie pour les autres. De même, je mettrai par la suite une majuscule à Univers pour signifier le monde où nous vivons tandis que les modèles d'univers n'en auront pas.

En s'éloignant encore, le 81<sup>e</sup> objet du catalogue de Messier, M81, se trouve à environ 8 megaparsecs et c'est la première galaxie hors du Groupe Local dont la distance ait été déterminée par la méthode des céphéides, à l'aide du télescope spatial Hubble. Les céphéides sont une classe particulière d'étoiles variables dont la période de pulsation régulière est reliée à leur luminosité moyenne, c'est-à-dire à la quantité d'énergie lumineuse qu'elles dégagent (en moyenne) par unité de temps. L'observateur perçoit la luminosité apparente de l'étoile, qui diminue comme le carré de sa distance car cette énergie est répartie sur une surface croissante au fur et à mesure de l'éloignement de l'étoile. En supposant que les céphéides lointaines obéissent à la même physique et donc à la même relation Période-Luminosité que localement, la comparaison entre « luminosité apparente mesurée » et « luminosité absolue déduite » permet de calculer la distance où se trouve une céphéide (et la galaxie qui la contient). Comme les céphéides sont des étoiles géantes très brillantes, cette méthode permet de sonder des distances beaucoup plus grandes que celle des parallaxes.

On trouve ensuite M51, une galaxie spirale de face des plus photogéniques, dite galaxie du tourbillon (« whirlpool » en anglais), qui est le 51<sup>e</sup> objet du catalogue de Messier. Ce catalogue contenant initialement 45 objets fut publié en 1744 par l'astronome Français Charles Messier. Il s'intéressait aux comètes et avait donc répertorié au cours de ses recherches tous les objets étendus qui ne se déplacent pas, constituant ainsi une liste de gêneurs qui risquaient d'être confondus avec une de ses chères comètes. Ce catalogue fut ensuite étendu à plusieurs reprises pour finalement comporter 110 objets<sup>5</sup>. M51 est une des découvertes originales de Charles Messier le 13 octobre 1773. C'est la première galaxie dans laquelle la structure spirale a été clairement distinguée, par Lord Rosse, au printemps de 1845. Au registre des célébrités, on trouve encore la galaxie du sombrero dont la forme évoque le chapeau mexicain éponyme et M87, qui est une galaxie elliptique géante située au centre de l'amas de la Vierge, à environ 20 megaparsecs.

Au-delà commencent véritablement les échelles cosmologiques au sens contemporain. Chaque point bleu correspond à un des 126 594 objets proches de l'équateur céleste ( $2^\circ > \delta > -2^\circ$ ) du catalogue SDSS en cours de constitution. Certaines directions ne sont pas couvertes (comme la zone centrale qui est dans la même direction que le centre Galactique – en fait les poussières du disque de notre Galaxie absorbent le rayonnement optique des objets lointains et rendent plus difficile toute cartographie exhaustive). À terme, le SDSS (Sloan Digital Sky Survey) aura répertorié plus de 100 millions d'objets sur

---

<sup>5</sup>Soit 37 galaxies, 29 amas globulaires, 28 amas ouverts, 7 nébuleuses diffuses et 4 nébuleuses planétaires.

un quart du ciel ; il aura mesuré la distance d'un million de galaxies et quasars lointains de l'échantillon. Le relevé est d'ores et déjà quasiment complet pour cette tranche.

Aux distances inférieures à quelques gigaparsecs ( $1 \text{ Gpc} \equiv 10^9 \text{ pc}$ ), les points correspondent principalement à des galaxies. On peut constater sur la figure que la distribution des galaxies est loin d'être uniforme. Elles se répartissent en surdensités reliées entre elles par de longs alignements qui cernent des régions quasiment vides. Comme il s'agit d'une tranche, cette structure évoque une coupe à travers une distribution à trois dimensions en bulle de savon. L'analyse de tranches adjacentes montre que c'est effectivement le cas ; les filaments reliant les amas de galaxies entre eux étant eux-mêmes à l'intersection de feuillets ou nappes de galaxies qui entourent des régions vides à peu près sphériques. Certaines de ces cavités ont un diamètre de près de 50 Mpc. C'est cette distribution remarquable des galaxies, similaire à l'ossature d'une éponge, qu'on appelle les grandes structures de l'Univers. Nous y reviendrons longuement (voir en particulier Fig. 7.14). Au-delà de quelques gigaparsecs, les points du SDSS sont essentiellement des quasars.

Les quasars sont des objets très brillants qui peuvent briller jusqu'à mille fois plus qu'une galaxie entière et qui sont pourtant de très petite taille, même pas un cent millième de la taille d'une galaxie, d'où leur nom qui est une contraction – en anglais – de « quasi stellar object » car ils ressemblent dans un télescope optique à une étoile ordinaire. Sauf qu'ils sont à des distances énormes et qu'ils libèrent donc une énergie lumineuse colossale. On pense aujourd'hui que ces objets sont des trous noirs extrêmement massifs – jusqu'à dix milliards de masse solaire<sup>6</sup> pour le quasar le plus lumineux connu – en train d'attirer la matière environnante. Cette matière tombe en spiralant autour du trou noir avant de s'y engoutir ; pendant la chute, la matière est compressée, cisailée, chauffée, à tel point qu'une partie de l'énergie de masse est transformée en rayonnement qui est émis dans une large gamme de longueur d'onde. Même si ce processus est remarquablement efficace par rapport aux autres mécanismes de conversion d'énergie connus, un trou noir typique doit néanmoins engoutir près de 10 masses solaires par an pour que brille le quasar. Toutes les galaxies, y compris la nôtre, hébergent sans doute en leur centre un tel ogre vorace. Mais dans la plupart des cas, il n'y a aujourd'hui plus de gaz interstellaire à aspirer, rendant le trou noir à son invisibilité, faute d'alimentation. Cependant dans le passé, le gaz était encore en abondance et une fraction beaucoup plus importante des galaxies était active.

C'est la luminosité extrême des quasars qui explique qu'on ne détecte quasiment plus qu'eux au-delà d'une certaine distance. En fait, au-delà de quelques gigaparsecs, ils sont largement majoritaires dans le catalogue SDSS. À ces échelles, la structure remarquable exhibée par la distribution des galaxies a disparu. Ce qui offre une indication tangible de l'homogénéité à grande échelle de l'Univers. Plus précisément, la densité de galaxies ou de quasars peut

---

<sup>6</sup>Une masse solaire,  $M_{\odot}$ , vaut  $1,989 \times 10^{30}$  kg.

s'évaluer en tout point en comptant leur nombre dans un volume centré autour de ce point. Plus le volume retenu sera grand, et plus on va obtenir une image de la distribution lissée sur des échelles importantes. La figure fait bien sentir que si la densité d'objets fluctue beaucoup à l'échelle de quelques megaparsecs, il est néanmoins clair que le contraste de densité diminue rapidement avec l'échelle sur laquelle on prend la moyenne. Aux échelles supérieures à quelques centaines de megaparsecs, les variations du champ de densité sont clairement faibles par rapport à la moyenne de la densité – tout comme la hauteur des vagues n'est qu'une faible perturbation de la profondeur de l'océan. Ce qui montre à tout le moins que postuler l'homogénéité de l'espace à grande échelle n'est pas déraisonnable.

Avant de poursuivre, rappelons nous que la finitude de la vitesse de la lumière est le postulat de base de la relativité. J'ai d'ailleurs indiqué plus haut que la lumière du soleil met environ huit minutes pour nous parvenir, et qu'il faut 3,26 années pour que la lumière parcourt un parsec, distance approximative à l'étoile la plus proche. La galaxie la plus proche est à 55 kpc. Nous la voyons donc telle qu'elle était il y a près de 180 000 ans, temps qu'il a fallu pour que sa lumière nous parvienne. L'objet le plus lointain visible à l'œil nu, la galaxie d'Andromède qui se trouve à 900 kpc, est donc vue telle qu'elle était il y a près de 3 millions d'années. En portant le regard au loin, on remonte donc dans le temps. C'est si vrai que la carte indique une ligne vers 14 gigaparsecs qui mentionne « Big Bang ». Elle marque la distance qu'aurait pu parcourir un photon ayant commencé son voyage vers nous aux tout premiers instants de l'Univers ! Ceci rend tangible le fait qu'aussi grande soit-elle, nous n'avons accès par nos observations qu'à une partie de l'Univers, celle à une distance telle qu'un photon peut nous parvenir en moins de temps que l'âge de l'Univers<sup>7</sup>. Cette distance maximale est notre horizon qui, comme l'horizon terrestre dû à la rotondité de notre planète, borne notre regard.

Une dernière remarque. Comment connaît-on la distance des objets du SDSS ? Leur éloignement est tel qu'il est impossible d'utiliser la méthode des céphéides. Il existe plusieurs techniques, mais la plus commune est l'utilisation du décalage vers le rouge du spectre des objets en fonction de l'éloignement. C'est une quantité qu'il est facile d'obtenir, ce qui permet précisément d'entreprendre la constitution de très grands relevés comme le SDSS, au prix cependant d'une indication qui n'est qu'indirecte. De quoi s'agit-il ?

L'Univers est en expansion, c'est un fait cosmologique majeur qui a été avancé en 1929 par Edwin Hubble et établi toujours plus solidement jusqu'à aujourd'hui. Cette propriété de l'espace fait que plus un objet est loin, plus son mouvement d'éloignement – sa récession – est rapide. Ceci se traduit par un décalage vers le rouge du spectre lumineux de l'objet. Pour des distances et des vitesses pas trop grandes, on peut interpréter ce rougissement comme un effet Doppler dû au déplacement de la source par rapport à l'observateur.

---

<sup>7</sup>Cela peut cependant englober tout l'Univers, si celui-ci est de volume fini et de taille suffisamment petite. Les résultats récents tirés de l'analyse des anisotropies du RCF semblent cependant indiquer que ce n'est pas le cas, cf. paragraphe 7.4.5.

En pratique, les raies caractéristiques de la composition de la source émettrice vont toutes se trouver décalées à des fréquences plus faibles par rapport à leurs valeurs de référence, en laboratoire. Appelant  $z$  ce décalage commun à toutes les raies, on a

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} = \frac{v_{12}}{c} \quad (7.1)$$

si on note  $\nu_1$  la fréquence d'une raie quelconque de la lumière émise,  $\nu_2$  la fréquence observée, et  $v_{12}$  la vitesse relative de la source par rapport à l'observateur (le signe  $\equiv$  dénote une définition). La vitesse de la lumière,  $c$ , vaut par définition  $299\,792\,458 \text{ m.s}^{-1}$ . La loi de Hubble [3] indique, là encore pour des distances et des vitesses pas trop grandes, que la vitesse de récession est proportionnelle à la séparation,  $r_{12}$ , ce qui s'écrit

$$v_{12} = Hr_{12} \quad (7.2)$$

$H$  est appelée la constante de Hubble, même si nous verrons par la suite que c'est en fait une quantité dépendante du temps. Le décalage vers le rouge peut donc bien servir d'indicateur de distance. D'après les mesures actuelles,  $H$  vaut environ  $70 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ . Cependant cette mesure est restée longtemps très incertaine ; il est donc devenu usuel d'écrire cette valeur comme

$$H = 100 h \text{ km.s}^{-1}.\text{Mpc}^{-1} \quad (7.3)$$

où le nombre sans dimension d'ordre un,  $h$ , sert à paramétriser notre ignorance et à la mettre en exergue dans les résultats qui en dépendent. Remarquons que l'inverse de cette constante définit une échelle de temps et de distance,

$$H^{-1} = 9,68 h^{-1} \text{ Gans}; \quad cH^{-1} = 2,998 h^{-1} \text{ Gpc} \quad (7.4)$$

On peut constater qu'à des distances suffisamment grandes ( $\sim 3 \text{ Gpc}$ ), la loi de Hubble (7.2) implique des vitesses de récession approchant ou dépassant la vitesse de la lumière, et des décalages vers le rouge de l'ordre de l'unité ou plus. Nous verrons bientôt que les expressions (7.1) et (7.2) ne sont en fait que des approximations qu'il faudra remplacer par leurs généralisations pleinement relativistes pour relier sans encombre distance, décalage vers le rouge, et temps de parcours quand les vitesses en cause sont du même ordre de grandeur que celle de la lumière. Notons au passage que le fait même que l'on puisse reconnaître dans les spectres d'objets extrêmement lointains des raies caractéristiques avec un arrangement correct et une forme raisonnable, cohérente avec la physique du gaz et des étoiles, nous suggère fortement que la physique est bien la même partout.

## 7.2 Quelques étapes marquantes de la cosmologie contemporaine

### 7.2.1 Définition des enjeux

Au début du vingtième siècle, la vision de l'Univers est très différente de celle révélée par la carte que nous venons de parcourir. L'Univers se réduit alors à notre Galaxie, une île plate et isolée (même si le philosophe Emmanuel Kant avait déjà imaginé à la fin du XVIII<sup>e</sup> siècle la possibilité d'existence d'autres Univers-îles). Les conceptions de l'époque se trouvent par exemple dans le livre d'Arthur Eddington de 1914 intitulé « Mouvements stellaires et structure de l'Univers ». Les comptages d'étoiles en fonction de la direction indiquent en effet une répartition à peu près planaire avec un bord, un disque d'extension finie. La nature et l'éloignement des nébuleuses du catalogue de Messier font cependant l'objet de débats.

Dès 1912, Vesto Slipher a montré que certaines nébuleuses sont animées de vitesses considérables, car il a constaté des décalages importants de leur spectre ; la plupart s'avèrent décalés vers le rouge, et en 1924 il aura dénombré 36 rougissements pour 41 nébuleuses. Mais leurs distances restent inconnues. Jusqu'à ce que Harlow Shapley s'intéresse de près aux Nuages de Magellan au cours de ses travaux sur l'extension de la Galaxie et la position de son centre. Après avoir calibré la relation Période-Luminosité découverte en 1912 par Henrietta Leavitt sur des céphéides de ces mêmes nuages [4], il estime leur distance à environ 75 000 années-lumière (soit 23 kpc), ce qui est une distance énorme pour l'époque. Mais Shapley, peut-être tout à sa volonté de faire reconnaître la taille immense de la Voie Lactée, continue malgré tout de penser que ces nuages font partie de notre Galaxie. D'autres en revanche, comme Heber Curtis ou Knut Lundmark, pensent que les nébuleuses spirales sont en fait des systèmes stellaires extérieurs. Le « grand débat » organisé en 1920 entre Shapley et Curtis par l'Académie des sciences américaine<sup>8</sup> illustre bien l'opposition entre le tenant d'un Univers contenant une seule Galaxie de grande taille, sans position particulière pour notre Soleil, et le tenant d'une Galaxie comparativement petite et quasi héliocentrique.

En quelques années il va devenir clair que si Shapley a raison en ce qui concerne la grande taille de la Galaxie qui n'a rien d'héliocentrique, c'est en revanche le point de vue représenté par Curtis qui va prévaloir en ce qui concerne les Nuages de Magellan. Ainsi Öpik estime en 1922 [5] qu'Andromède est à 450 kpc grâce à un certain nombre d'hypothèses qui vont s'avérer par la suite correctes pour l'essentiel. Il conclut : « *L'ordre de grandeur de la distance trouvée est en accord avec d'autres estimations de la distance à la nébuleuse d'Andromède ; la coïncidence des résultats obtenus par plusieurs méthodes indépendantes augmente la probabilité que cette nébuleuse soit un univers stellaire, comparable à notre Galaxie* ». C'est Edwin Hubble qui apportera en

---

<sup>8</sup>Les surfeurs anglophones pourront consulter le site : [antwrp.gsfc.nasa.gov/debate/debate20.html](http://antwrp.gsfc.nasa.gov/debate/debate20.html).





FIG. 7.2 – La galaxie d'Andromède, M31.

1926 l'argument décisif [6]. En utilisant un échantillon de céphéides d'Andromède qu'il a lui-même assemblé [7], il trouve en effet une distance de 700 000 années-lumière en utilisant la calibration de la relation Période-Luminosité donnée par Shapley en 1925. Cette distance est près de 10 fois plus grande que celle des Nuages de Magellan, et le doute ne semble plus permis. On peut remarquer à ce sujet que la valeur moderne est plutôt  $770 \pm 30$  kpc, (soit environ 2 500 000 années-lumière, près de trois fois la valeur de Hubble) ; cette différence tient à l'existence de *deux types de céphéides* (ayant chacun une relation Période-Luminosité distincte), ce qu'ignoraient les astronomes de l'époque, et qui biaisa leurs estimations. Quoi qu'il en soit, en 1926 le pas est fait : il faut abandonner la vision d'une galaxie héliocentrique de Curtis et celle Galactocentrique de Shapley ; l'Univers est vraiment immense, peuplé de myriades de galaxies, et nous n'occupons aucune position singulière.

C'est dans ce contexte qu'Einstein publie en 1917 sa théorie de la relativité générale et qu'il fait de plus l'hypothèse audacieuse d'un Univers (en moyenne) homogène et isotrope. Cette hypothèse (qu'il fait pour développer un modèle cosmologique qui ne soit pas en contradiction avec les idées de Mach sur l'origine de l'inertie gravitationnelle) est néanmoins surprenante par son accord avec les vues actuelles malgré le manque de support empirique dont elle a pu disposer à ce moment-là. Notons de plus qu'Einstein recherche d'emblée une solution statique, sans expansion. Pour y parvenir, il lui faut introduire dans les équations de la relativité un terme dit de constante cosmologique, traditionnellement noté  $\Lambda$ , qui n'est pas forcément requis. Ce terme permet cependant de contrebalancer l'effet de la matière qui, sinon, implique inéluctablement une évolution temporelle. Or, à cette époque, l'Univers se réduit à notre Galaxie, et les indications observationnelles disponibles sur la vitesse des étoiles ne suggèrent pas d'expansion. La même année, le Néerlandais Willem de Sitter [8] trouve la solution des équations de la relativité correspondant

à un univers vide et en expansion, dont la dynamique est dominée par une constante cosmologique. Cette solution mathématique n'est pas, comme nous le verrons, d'un intérêt uniquement académique; il semble par ailleurs que cette analyse a fait douter Einstein du bien-fondé de son introduction ad hoc de la constante cosmologique. Quoi qu'il en soit, la théorie vient en quelque sorte à point nommé pour accompagner le mouvement de découvertes observationnelles de nature cosmologique qui va perdurer jusqu'à nos jours. En reliant la structure même de l'espace-temps aux propriétés du contenu matériel de l'Univers, la théorie de la relativité générale va permettre de mettre en place un cadre conceptuel cohérent pour développer une vision totalement nouvelle de l'Univers, la nôtre. L'annonce par Eddington en 1919 d'une première vérification expérimentale de la théorie gravitationnelle d'Einstein va rapidement la faire accepter par nombre de théoriciens, sans attendre d'autres confirmations empiriques. La relativité générale va jouer un rôle essentiel pour placer les observations dans un cadre de référence et permettre un aller-retour fructueux entre les avancées observationnelles et théoriques.

Dès 1922, le mathématicien russe Alexandre Friedman [9] trouve la solution mathématique, variable temporellement, qui correspond à un modèle où la distribution de matière est homogène et isotrope; cela implique notamment une relation linéaire entre le décalage vers le rouge,  $z$ , et la distance des galaxies, relation qui ne sera malheureusement découverte observationnellement par Hubble que sept ans plus tard [3], sur la base de mesures de distances de 24 galaxies brillantes proches. D'une certaine manière, Friedman est trop en avance. Cinq ans plus tard, l'abbé belge Georges Lemaître, qui est aussi un physicien théoricien, découvre à son tour [10, 11], et indépendamment, la solution décrivant l'évolution temporelle d'un modèle d'univers relativiste rempli de matière. Et il fait explicitement le lien entre l'expansion de l'espace et les décalages vers le rouge (les vitesses radiales de nébuleuses mesurées par Slipher). Il pose ainsi les fondements de la cosmologie physique. Le modèle, dit de Friedman-Lemaître, est encore aujourd'hui le modèle de référence, même si un modèle d'univers stationnaire, proposé en 1948 par Bondi et Gold [12] et séparément par Hoyle [13], a été pour un temps une alternative possible.

Mentionnons ici un modèle de Friedman-Lemaître particulier, celui proposé conjointement par Einstein et de Sitter en 1932 [14]; il va avoir une influence considérable par sa simplicité et son élégance, même s'il est presque certainement exclu aujourd'hui. C'est un modèle minimal, dans lequel la matière est en quantité juste suffisante pour que la géométrie de l'espace soit la géométrie euclidienne usuelle, plate (et donc sans un rayon de courbure ayant une valeur particulière). Les distances entre points de l'espace croissent toutes comme le temps à la puissance deux tiers sous l'effet de l'expansion. Cette simple loi de puissance valable à tous temps ne permet pas de définir d'époque particulière où l'évolution temporelle changerait de régime. Il n'y a donc pas de temps caractéristique dont il faut rendre compte. Cette simplicité du modèle d'Einstein-de Sitter en fait l'attrait.

Je vais maintenant relater quelques étapes clés des développements ultérieurs qui ont conduit à la vision du monde que nous avons aujourd'hui. Cette cosmologie peut se comprendre comme la réunion de la cosmographie et de la cosmogonie, l'alliance du théâtre et de la pièce jouée, du substrat et de la création. À la cosmographie le soin de préciser la structure géométrique de l'espace-temps, le modèle d'ensemble, indépendamment du détail de son contenu et de sa distribution. À la cosmogonie celui de rendre compte de l'émergence des grandes structures de l'Univers, de la sculpture progressive de l'espace. Paradoxalement, la confiance d'aujourd'hui dans notre modèle cosmographique s'enracine dans les succès récents de la cosmogonie.

Pour s'y retrouver dans cette histoire un peu longue et complexe, même réduite à quelques étapes significatives, on pourra sans doute consulter utilement la table synoptique, page 336, qui permet de mieux apprécier le développement concomitant de la cosmographie et de la cosmogonie. Un mot encore ; l'exposé linéaire qui suit est trompeur, au sens où il ne fait pas justice aux nombreux débats, souvent acharnés, qui ont eu lieu au cours de ces soixante années. Il n'offre pas un panorama exhaustif de l'histoire des idées, c'est plutôt une reconstitution a posteriori de certaines avancées qui se sont avérées importantes pour comprendre où nous en sommes<sup>9</sup>.

### 7.2.2 Cosmographie, nucléosynthèse et rayonnement fossile

L'année même de la proposition d'univers stationnaire, George Gamow [17] fait une prédiction dont la vérification détaillée scellera le destin de cette alternative. L'expansion de l'Univers que nous constatons (et qui se traduit par une récession des galaxies d'autant plus rapide qu'elles sont plus éloignées) implique<sup>10</sup> en effet que, plus tôt dans l'histoire de l'Univers, son contenu était plus dense et plus chaud, tout comme s'échauffe l'air qu'on comprime dans une pompe de bicyclette. En remontant très loin dans le passé arrive un moment où les atomes de la matière sont tellement proches et rapides (chauds) que leurs collisions sont suffisamment énergétiques pour les ioniser, c'est-à-dire séparer de manière durable le noyau de son cortège électronique. En remontant plus encore dans le temps, ce sont les noyaux eux-mêmes qui ne résistent pas. C'est précisément en étudiant la synthèse progressive des éléments à partir de protons, par captures neutroniques successives au cours des premières secondes de l'Univers, que George Gamow déduisit en 1948 qu'il doit y avoir encore aujourd'hui un fond résiduel de photons bien particulier, fossile témoignant de la fournaise primordiale.

La recherche d'une origine pour les divers éléments chimiques conduit les chercheurs dans les années quarante à rechercher des phases chaudes où

---

<sup>9</sup>On pourra avec profit se reporter au livre de Jean-Pierre Luminet [15] pour en savoir plus, ainsi qu'à la première partie du livre plus technique (et uniquement disponible en anglais) de James Peebles [16], l'un des protagonistes les plus éminents de cette histoire.

<sup>10</sup>À condition toutefois que la matière soit conservée.

TAB. 7.1 – Table synoptique d'événements marquants de la cosmologie moderne.

1917	<i>Einstein</i> publie la théorie de la relativité générale et fait l'hypothèse audacieuse d'un Univers (en moyenne) homogène et isotrope pour proposer un modèle cosmologique statique. <i>De Sitter</i> trouve la solution correspond à un Univers vide dont la dynamique est dominée par une constante cosmologique.
1922	<i>Friedman</i> trouve la solution mathématique décrivant la modification temporelle d'un univers homogène en évolution (qui implique notamment une relation linéaire entre le décalage vers le rouge, $z$ , et la distance d'un objet).
1926	<i>Hubble</i> situe Andromède très en dehors de notre Galaxie. Les nébuleuses spirales deviennent des galaxies. Il montre ensuite que la distribution des galaxies est approximativement uniforme, sans indication de bord.
1927	<i>Lemaître</i> pose les fondements de la cosmologie physique. Il fait une connexion explicite entre les observations et la théorie, dans le contexte des modèles de Friedman-Lemaître.
1929	<i>Hubble</i> découvre observationnellement une relation linéaire entre vitesse de récession et distance ( $v = Hr$ ).
1932	<i>Einstein</i> et <i>de Sitter</i> proposent le modèle qui porte leurs noms.
1931	<i>Lemaître</i> se rend compte qu'un univers en expansion est gravitationnellement instable. La recherche cosmogonique d'une théorie de formation des grandes structures par cette instabilité peut commencer. . .
1937	<i>Zwicky</i> remarque que la vitesse des galaxies dans l'amas de Coma est beaucoup trop grande pour que le système soit lié. C'est la première indication de l'existence de matière sombre.
1946	<i>Lifshitz</i> fait une analyse complète de l'évolution (linéaire) de petites perturbations dans le cadre des modèles de Friedman-Lemaître. Mais il conclut trop hâtivement que l'instabilité gravitationnelle <i>n'est pas</i> à l'origine de la condensation de la matière en nébuleuses distinctes.
1948	<i>Gamow</i> propose une synthèse dynamique des éléments par captures neutroniques successives. Cela le conduit à prédire l'existence d'un rayonnement fossile ayant une température aujourd'hui $T_0 \sim 10$ K.
1948	<i>Bondi et Gold</i> d'une part et <i>Hoyle</i> d'autre part proposent le modèle d'univers stationnaire.
1964	<i>Novikov</i> réalise que la conclusion de Lifshitz en 1946 est erronée : l'instabilité gravitationnelle <i>peut</i> engendrer les grandes structures.
1965	<i>Penzias et Wilson</i> annoncent la première détection du Rayonnement Cosmologique de Fond.
1970	Introduction de conditions initiales « naturelles », correspondant à des fluctuations du contraste de densité invariants d'échelle (spectre primordial dit de Harrison-Peebles-Zeldovich).
1975	Au cours des années soixante-dix, la notion de matière sombre dans et autour des galaxies devient progressivement acceptée à cause des courbes de rotation des spirales qui sont plates à grande distance.
1981	<i>Guth</i> propose le modèle d'inflation dont les défauts initiaux seront amendés dès 1982 (par d'autres auteurs).

TAB. 7.1 – Suite.

1982	<i>Peebles</i> propose le modèle dit de matière sombre Froide « CDM ».
1986	<i>De Lapparent, Geller, Huchra</i> montrent que les galaxies semblent réparties à la surface de bulles de diamètre compris entre 25 et $50 h^{-1}$ Mpc.
1990	<i>Maddox et al.</i> trouvent que l'amplitude des fluctuations à grande échelle du relevé APM outrepassé nettement les prédictions du modèle CDM standard.
1990	<i>Mather et al.</i> établissent précisément la distribution de corps noir du RCF et mesurent sa température, $T_0 = 2,735 \pm 0,06$ K, avec l'instrument FIRAS à bord du satellite COBE.
1992	<i>Smoot et al.</i> détectent les premières anisotropies du RCF avec l'instrument DMR du satellite COBE.
1995	Le modèle de concordance $\Lambda$ CDM émerge comme nouveau standard de travail possible.
1998	Les mesures de supernovae de type Ia indiquent une accélération actuelle de l'expansion, ce qui contribue à un changement du modèle cosmologique de référence ( $\Lambda$ CDM), même s'il n'est pas encore standard.
2000	Les mesures d'anisotropies du RCF par les expériences ballon Boomerang et MAXIMA indiquent précisément une géométrie cosmologique à courbure spatiale faible ou nulle.
2003	Les anisotropies du RCF mesurées par le satellite WMAP en conjonction avec les nouveaux résultats du relevé de galaxies Sloan et des supernovae détectées avec le HST fournissent des indications indépendantes précises d'un même modèle de concordance $\Lambda$ CDM.

suffisamment d'énergie serait disponible pour permettre la construction des éléments par transformation progressive. La question est de comprendre si des interactions nucléaires entre protons et neutrons ont pu graduellement générer l'ensemble des éléments connus dans les proportions observées. L'intérieur des étoiles ou la phase primitive chaude que nous venons d'évoquer sont deux candidats possibles pour réunir les conditions requises. On comprend aujourd'hui que les interactions dans l'Univers primordial permettent de transformer les protons en noyaux de deutérium constitués d'un proton et d'un neutron puis en hélium 3 (deux neutrons et un proton) et hélium 4 (deux protons et deux neutrons) par captures successives de neutrons. Le processus s'arrête là, quelques minutes après le Big Bang, et tout le reste des éléments que nous connaissons a été synthétisé dans les étoiles. Si seules les étoiles avaient été à l'œuvre, la fraction d'hélium serait beaucoup plus proche de celle des éléments lourds, ce qui est totalement exclu par les observations. Cette théorie offre ainsi une explication satisfaisante à l'origine de l'hélium. La concordance entre les prédictions de ce modèle de nucléosynthèse primordiale et les observations permet de considérer cet accord quantitatif comme un des piliers du modèle standard.

Remontons encore le temps ; à un certain stade, les énergies sont telles que les collisions furieuses et incessantes entre photons et électrons imposent une distribution en énergie des photons très spécifique, que l'on appelle une distribution de Planck. Cette distribution – dite aussi « de corps noir » – est la distribution qu'adopte tout corps en équilibre. Elle se caractérise par une seule quantité, la température, qui est une mesure de l'énergie disponible. Il suffit de connaître la dite température pour déterminer complètement l'ensemble de la distribution. Il s'avère qu'une fois établie, cette distribution reste une distribution de corps noir au cours de l'évolution, à une variation près de la température, *même si plus tard l'équilibre est rompu*. D'ailleurs aujourd'hui, à cause de l'extrême raréfaction de la matière cosmique, des étoiles chaudes, des nuages de plasma brûlant ou des nébuleuses glaciales coexistent sans que leurs températures s'équilibrent. La préservation hors équilibre de la forme de la distribution de photons explique d'ailleurs pourquoi on peut attribuer une température à chaque instant de l'histoire de l'Univers. Cette température décroît proportionnellement au facteur d'expansion (qui quantifie la façon dont croissent les séparations moyennes entre particules) ; elle est souvent plus commode à utiliser que le temps pour repérer les différentes époques de l'évolution.

En supposant qu'il y ait eu à peu près autant de protons que de neutrons dans l'Univers primitif, Gamow déduisit que la création du deutérium par agrégation d'un proton plus un neutron et libération d'un photon a dû se produire quand la température de l'Univers était d'environ 1 milliard de degrés. À des températures plus élevées, c'est en effet la réaction inverse qui est dominante, c'est-à-dire que les photons sont alors suffisamment énergétiques pour dissocier en proton et neutron tout atome de deutérium qui aurait pu se former. Cela fixe l'époque de cette création à environ 200 secondes après le Big Bang. Sachant de plus que le deutérium est bien moins abondant dans l'Univers que l'hydrogène, Gamow en déduisit l'ordre de grandeur de la densité de particules à cette époque. Et en comparant avec une estimation de la densité de particules aujourd'hui, il put en déduire un facteur d'expansion de l'ordre de cent millions entre l'époque de la nucléosynthèse du deutérium et aujourd'hui. Ce qui lui permit donc de prédire l'existence d'un corps noir avec une température aujourd'hui de l'ordre de dix degrés kelvin. Cette estimation fut raffinée la même année (1948) par Ralph Alpher et Robert Herman [18] qui prédirent une température de l'ordre de 5 K (toujours sous l'hypothèse a priori d'une égale quantité de protons et de neutrons primitifs).

Ce n'est que 17 ans plus tard, en 1965, que deux ingénieurs des laboratoires Bell, Arno Penzias et Robert Wilson, découvrent par hasard les preuves directes du passé torride de l'Univers. À l'occasion de la calibration minutieuse d'une antenne destinée à un système de communication, Penzias et Wilson ne peuvent trouver l'origine d'un bruit de fond, isotrope, à une longueur d'onde de 7 cm. C'est pourquoi l'article qu'ils publient s'intitule sobrement « Bruit instrumental en excès anormal à 4,3 GHz » [19]. Une interprétation cosmologique est proposée dans l'article suivant du même journal [20] par un groupe

de chercheurs de Princeton dirigé par Robert Dicke, à quelques dizaines de kilomètres de la Bell. Les circonstances méritent d'être relatées.

Il se trouve en effet que l'équipe de Dicke travaillait à cette époque sur la disparition des éléments lourds qui, dans le contexte d'un univers cyclique, doit avoir lieu en fin de cycle pour éviter leur accumulation de cycle en cycle. Des calculs de James Peebles (menés sans connaissance des travaux de Gamow et al. qui ne sont donc, ironie de l'histoire, pas cités dans l'article) avaient montré que l'Univers doit passer par une phase chaude d'au moins  $10^{10}$  K pour permettre la destruction de ces éléments lourds. D'où un fond résiduel, lui aussi de l'ordre de 10 K, et la construction en cours d'une antenne pour détecter ce fond par deux autres membres de l'équipe, Peter Roll et David Wilkinson. C'est par le simple bouche à oreille que Penzias et Wilson auront vent d'un séminaire sur le sujet que venait de faire Peebles. Remarquons au passage que le détecteur utilisé par Penzias et Wilson est un radiomètre, un appareil inventé par Dicke lui-même, au cours de recherches qu'il avait menées à la Bell au cours de la seconde guerre mondiale. Quel incroyable enchevêtrement de la trame du destin ! D'ailleurs Penzias et Wilson ne sont pas les premiers à avoir détecté le rayonnement fossile, mais c'est leur mesure qui sera interprétée en tant que tel la première.

Cette découverte, quoique à une seule fréquence, d'un fond à  $T_0 = 3,5 \pm 1$  K va commencer à faire pencher nettement la balance du côté des tenants du modèle de Friedman-Lemaître. Nombre d'autres expériences s'ensuivent pour vérifier si la variation de la densité d'énergie du rayonnement fossile en fonction de la fréquence est bien celle prévue par la loi de corps noir. Grâce à l'instrument FIRAS à bord du satellite COBE, cette quête s'achève en 1990 par la mesure précise du spectre énergétique du fond ; elle se décrit précisément par une distribution de corps noir à une température aujourd'hui de  $T_0 = 2,735 \pm 0,06$  K [19] (et [21]) pour l'analyse finale). Ce spectre est précisément isotrope, c'est-à-dire identique dans toutes les directions, à l'exception d'une importante composante dipolaire, d'amplitude égale à 3 mK environ, qui résulte de notre déplacement. John Mather et ses collègues ont ainsi établi que le rayonnement fossile est extrêmement proche d'un corps noir idéal. Par là même, ils ont fini de convaincre la communauté dès lors quasi unanime des astrophysiciens que le modèle de Friedman-Lemaître de Big Bang capture une part de la réalité physique. Les recherches faites montrent en effet que si cette prédiction est naturelle pour ce modèle, il en est tout autrement pour les modèles concurrents.

Si l'expérience FIRAS permet de fixer la contribution du rayonnement à la densité d'énergie de l'Univers, elle ne dit cependant rien sur la valeur des autres contributions. Or la relativité générale lie précisément les caractéristiques de la géométrie de l'espace-temps aux propriétés de son contenu. D'autres tests cosmologiques permettent cependant de vérifier d'autres aspects du modèle et de fixer la valeur d'autres paramètres importants. Nous verrons bientôt (paragraphe 7.3.4) que la mesure de l'éclat des supernovae en fonction de leur éloignement va fortement contribuer à faire accepter à la

communauté un modèle de Friedman-Lemaître particulier, voire singulier. Ce modèle, repéré par l'acronyme quelque peu cryptique  $\Lambda$ CDM, a déjà commencé à émerger vers 1995, mais sur une base essentiellement cosmogonique, c'est-à-dire en recherchant quel cadre permet la construction d'une théorie satisfaisante de la formation et de l'évolution des grandes structures de l'Univers. Ce sont d'ailleurs des mesures concernant l'empreinte primordiale des grandes structures qui achèveront d'emporter, vers l'an 2000, l'adhésion de la communauté à ce modèle. On ne peut donc sérieusement discuter de l'état actuel de la cosmologie sans une appréciation assez approfondie de son volet cosmogonique.

### 7.2.3 Cosmogonie et établissement du paradigme actuel

Lemaître a posé avec Friedman les fondements de la cosmologie physique fondée sur la théorie d'Einstein. Il se rend de plus compte, en 1933 [22], qu'un univers en expansion est gravitationnellement instable. Il initie ainsi la recherche d'une cosmogonie physique, d'une théorie de formation des grandes structures par l'action de cette instabilité sur des condensations primordiales, en acceptant toutefois d'admettre comme acquis l'existence préalable de ces condensations. Ce programme de recherche particulièrement ambitieux est aujourd'hui considéré comme achevé par l'immense majorité des astrophysiciens. Mais cet accord est très récent, et n'inclut pas le détail de la formation des galaxies elles-mêmes. D'ailleurs pendant longtemps, disons de 1935 à 1965, la grande question est plutôt celle du modèle d'univers à retenir, Friedman-Lemaître ou stationnaire, et la cosmogonie semble piétiner. Il se produit cependant pendant ces trente années une avancée majeure qui est le calcul relativiste complet des premières phases de l'évolution de petites fluctuations de la densité que publiera Evgueni Lifshitz [23] en 1946. Il ne se rend malheureusement pas compte de l'existence d'un régime particulier important, ce qui l'amène à conclure à tort que l'instabilité gravitationnelle n'est apparemment pas la source du rassemblement de la matière sous forme de nébuleuses. C'est Igor Novikov [24] qui trouvera en 1964 la faille du raisonnement de Lifshitz.

Autre grande avancée, plusieurs groupes réalisent en 1970 qu'il existe des conditions initiales *naturelles* pour initier la croissance gravitationnelle ; les postuler constitue donc une bonne hypothèse de travail. En l'absence d'une théorie fondamentale, il est en effet raisonnable de ne considérer que des fluctuations primordiales Gaussiennes dont l'amplitude typique (au sens de l'écart quadratique moyen) ne singularise pas d'échelle particulière, ce qui amène à supposer une dépendance en loi de puissance de l'échelle. L'homogénéité à grande échelle et la causalité restreignent les exposants possibles. Le cas particulier dit invariant d'échelle correspond à un exposant tel que les perturbations (de la géométrie) ne dépendent que logarithmiquement de l'échelle, aussi bien pour les très grandes que pour les très petites échelles. Cette divergence aux deux extrêmes étant très lente, les deux échelles caractéristiques où il doit y avoir écart à la loi de puissance peuvent donc être



très éloignées de celles qui nous intéressent en cosmologie, disons entre le parsec et le gigaparsec. Ce qui rend le résultat effectivement indépendant de toutes échelles injustifiées. Il s'avère que ce type de conditions initiales, dites de Harrison [25] Peebles [26] Zeldovich [27], est capable d'engendrer de façon relativement simple les grandes structures observées. Un grand nombre de ceux qui le démontreront ont lu deux véritables classiques de ce domaine, « Physical Cosmology » publié par James Peebles en 1971, et « Gravitation and Cosmology » publié par Stephen Weinberg en 1972.

Au cours des années soixante-dix, la notion de matière sombre dans et autour des galaxies est progressivement acceptée à cause, notamment, des courbes de vitesse de rotation des spirales qui restent plates jusqu'à grande distance du centre (cf. Fig. 7.5, paragraphe 7.3.3). Une simple application de la gravité Newtonienne, a priori valable à cette échelle somme toute relativement modeste, montre en effet qu'il doit y avoir beaucoup plus de matière que ne le laisse présager la distribution stellaire. C'est au fond le même argument que celui de Zwicky qui, 40 ans plus tôt [28,29], avait le premier remarqué que la vitesse des galaxies dans l'amas de Coma est beaucoup trop grande pour que le système soit lié. Pour que l'amas ne vole pas en éclats, il estime qu'il doit y avoir entre 100 et 500 fois plus de masse que ce que l'on peut estimer sur la base du nombre de galaxies dans l'amas (et de leur masse évaluée en supposant un rapport masse/luminosité constant). Sous quelle forme se cache cette matière qui ne brille pas et qu'on prendra donc l'habitude d'appeler la *matière sombre*, faute de mieux ? Diverses études vont exclure un certain nombre de possibilités naturelles à base de matière atomique habituelle, celle dont nous sommes faits. Plus précisément, on parle de matière baryonique, c'est-à-dire constituée essentiellement de protons et de neutrons (ainsi qu'un certain nombre de particules instables plus lourdes), le terme baryonique dérivant d'un mot grec qui signifie pesant.

Certains chercheurs commencent alors à considérer que la matière sombre pourrait être constituée de neutrinos, des particules non-baryoniques interagissant très faiblement avec le reste de la matière, dans l'hypothèse où ces particules seraient dotées d'une masse faible (mais non nulle comme cela avait été supposé jusqu'alors). Ainsi Cowsik et McClelland suggèrent en 1973 [30] que les neutrinos, de symbole  $\nu$ , pourraient bien constituer la matière sombre des amas de galaxies, pourvu qu'ils aient une masse de quelques  $\text{eV}/c^2$ , valeur très inférieure aux limites expérimentales disponibles alors ( $m_{\nu_e} < 60 \text{ eV}$ ,  $m_{\nu_\mu} < 0,8 \text{ MeV}$ ). Dès 1976, les Hongrois Alexander Szalay et George Marx [31] commencent l'étude détaillée de la croissance gravitationnelle de perturbations primordiales (à l'échelle des amas) dans un univers où les neutrinos constituent l'essentiel de la matière (il s'agit en fait d'une partie de la thèse de doctorat de Szalay, sous la direction de Marx et soutenue en 1975, en poursuivant les idées de Dolgov et Zeldovich). Le modèle « HDM » prend rapidement corps. HDM est l'acronyme anglais de *Hot Dark Matter*, ce qui veut dire matière sombre chaude. Le terme chaud, dans ce contexte, fait référence à la dispersion des vitesses de ces particules qui est suffisamment

grande pour littéralement faire disparaître, au cours des premiers stades de l'évolution, toutes les perturbations de masse inférieure à celle d'un amas de galaxie environ.

Très tôt dans l'histoire des modèles de Friedman-Lemaître, les conditions énergétiques sont telles que toutes les particules sont en équilibre thermique, avec une énergie de masse ( $mc^2$ ) négligeable devant l'énergie thermique ( $kT$ ). Mais les vitesses de particules diminuent au fur et à mesure de l'expansion (sauf bien sûr celles des particules de masse nulle qui restent constantes). À un moment donné, la vitesse devient plus petite que  $c$ , la particule est devenue non relativiste. La distance parcourue par une particule vaut approximativement la distance qu'elle a parcourue tant qu'elle était relativiste (tant que  $kT > mc^2$ ). Comme une particule plus lourde devient non relativiste plus tôt, il s'ensuit qu'elle parcourt une distance plus faible qu'une particule plus légère. Imaginons maintenant une fluctuation primordiale *adiabatique*<sup>11</sup>, c'est-à-dire conjointe de tous les types de particules sans distinction, photons, baryons, neutrinos, etc., sans que soient modifiées les proportions relatives des différentes espèces. Considérons par exemple une vague de densité sinusoïdale de ce fluide un peu particulier, succession de régions sur- et sous-denses. Si les neutrinos allaient tous dans la même direction, on verrait simplement se propager une onde plane. Mais pour une distribution isotrope des vitesses initiales de neutrinos, si les neutrinos ont pu parcourir une distance nettement supérieure à la longueur d'onde de la vague, cette fluctuation de masse est effectivement effacée (car on suppose qu'ils dominent largement la densité de matière). À un instant donné, toutes les fluctuations dont la longueur caractéristique est plus courte que la distance parcourue par les neutrinos jusqu'à cet instant sont donc gommées par cette diffusion libre. Le calcul précis de cette longueur de diffusion montre que, pour une masse de l'ordre de 1 keV, les seules fluctuations primordiales qui ne sont pas amorties par ce mécanisme ont la masse d'une galaxie typique. Une telle particule serait en fait classée comme de la matière tiède. On parle plutôt de matière chaude pour des masses de l'ordre de la dizaine d'eV, quand les fluctuations qui sont préservées correspondent à l'échelle des grandes structures ou superamas. Et pour compléter la terminologie, on appelle matière froide une matière dont la dispersion de vitesse n'amortit aucune échelle cosmologique significative.

Pour de la matière sombre chaude, la formation des grandes structures et des galaxies ne peut procéder qu'en cascasant des grandes échelles aux petites, puisque seules les grandes échelles ont pu échapper à la diffusion des particules dont nous venons de parler. Dès 1970, Zeldovich [32] a montré que les premières structures à apparaître dans un milieu sans pression (comme un fluide de neutrinos) et en expansion doivent alors être des feuilletés, ou crêpes.

---

<sup>11</sup>On peut imaginer d'autres types de fluctuations, où l'on fait varier l'abondance relative des différentes espèces. Si ces variations se compensent exactement, de façon à ce qu'il n'y ait pas de perturbation de la densité d'énergie totale, on parle de fluctuations *isocourbes*, c'est-à-dire qu'elles n'affectent pas la courbure de l'espace, seule l'équation d'état varie spatialement.

Leur séparation et leur dimension caractéristique de l'ordre de  $40 h^{-2}$  Mpc rappellent fortement les grandes structures du relevé SDSS. La matière rassemblée dans ces feuillets, dans les filaments, et amas à leurs intersections, peut alors se fragmenter et donner naissance aux galaxies. Cette « théorie des crêpes » a donc pour elle d'avoir prédit l'existence de feuillets, à une époque où une telle distribution était, au mieux, peu manifeste dans les observations (il faudra attendre le relevé CfA2 de 1986 pour qu'apparaisse un grand mur dans la distribution des galaxies proches). Ce scénario suscite alors beaucoup d'intérêt, d'autant qu'une équipe soviétique [33] annonce en 1980 avoir justement mesuré une masse du  $\nu_e$  de quelques dizaines d'eV. Mais cette détection ne sera pas confirmée par la suite, et les simulations numériques de ce scénario (à partir de 1983 [34, 35]) confirment que, pour des conditions initiales naturelles, les amas des simulations sont beaucoup trop vieux par rapport aux galaxies (ils se forment avant dans la cascade), en désaccord avec ce que les observations commencent à indiquer.

Un changement de paradigme se produit alors entre 1980 et 1985. J'utilise ici le mot paradigme au sens original donné à ce mot par l'historien des sciences et épistémologue Thomas Kuhn dans son ouvrage sur la « Structure des révolutions scientifiques » [36]. Un paradigme fait donc référence à un ensemble de convictions qui sont partagées par la communauté scientifique mondiale à un moment donné. Dans le cas du scénario HDM, l'approche avait consisté à considérer les implications cosmologiques au cas où une particule connue s'avérerait avoir une masse non-nulle. Cette démarche avait permis dans un premier temps d'utiliser l'Univers comme un laboratoire, en l'espèce, pour poser une limite supérieure d'origine cosmologique sur la masse des neutrinos, comme le font Marx et Szalay dès 1972 [37] (en requérant qu'ils ne fournissent pas plus de matière que ne le permettent les mesures astronomiques). Ceci assuré, cette hypothèse permettait alors de résoudre un problème proprement cosmologique, celui de la masse manquante. Il est évidemment naturel d'utiliser le laboratoire-univers pour contraindre aussi les propriétés d'une particule hypothétique et guider ainsi les développements de la théorie. C'est exactement ce que font par exemple Pagels et Primack en 1982 en appliquant la même contrainte sur la densité de l'Univers, mais cette fois dans le cas du gravitino, la particule partenaire hypothétique de spin  $3/2$  du graviton dans une théorie supersymétrique des forces fondamentales. Ils montrent ainsi que le gravitino ne doit pas excéder une masse de 1 keV [38]. Très peu de temps après, Pagels et Primack, en compagnie de l'astronome Blumenthal, réalisent qu'un univers dont la densité serait dominée par les gravitinos (matière sombre tiède, donc) pourrait former directement des galaxies par instabilité gravitationnelle [39], une réflexion que se font simultanément Bond, Szalay et Turner [40].

Le changement de paradigme qui s'est produit alors a consisté à accepter de rechercher une solution à un problème cosmologique, celui de la masse manquante, en supposant l'existence de particules (dont il existe des candidats théoriques) ayant les propriétés nécessaires pour résoudre ce problème. Ce changement d'attitude a été sans doute grandement facilité par la prise de

conscience d'un problème nouveau, et sa résolution potentielle. Dès la découverte de Penzias et Wilson, les astrophysiciens commencent à calculer les conséquences de l'existence des fluctuations primordiales sur les propriétés de ce fond cosmologique. Ils montrent qu'on attend des petites fluctuations de température en fonction de la direction d'observation, des anisotropies minuscules qui sont l'empreinte des perturbations. Dès 1970, les grandes lignes de la théorie des anisotropies du rayonnement fossile sont bien développées [26, 41–45]. Mais ce que montrent Silk et Wilson en 1981 [46], c'est que les anisotropies attendues sont supérieures aux limites observationnelles de l'époque ( $\delta T/T \lesssim 1 \times 10^{-4}$ ) pour des conditions initiales naturelles dans un univers dont la masse est dominée par les baryons. En d'autres termes, il faut, d'une manière ou d'une autre, accepter d'être non conventionnel. James Peebles propose alors le modèle CDM, de l'acronyme anglais *Cold Dark Matter*, ce qui veut dire matière sombre froide [47]. Ce scénario permet de contourner le problème des anisotropies excessives du rayonnement fossile tout en restant particulièrement simple, en évitant d'introduire des échelles caractéristiques injustifiées. Il s'agit d'un modèle d'univers sans courbure (donc pas de rayon de courbure à expliquer), dominé par de la matière sombre froide, dont la masse précise est sans importance (il n'y a donc pas d'échelle caractéristique de diffusion en-deçà de laquelle les fluctuations sont amorties) et on suppose les conditions initiales naturelles dont nous avons parlé précédemment (là aussi donc sans échelle particulière). À part l'hypothèse sur l'existence même d'une particule inconnue, il s'agit vraiment là d'un modèle minimal.

La croissance des fluctuations y est hiérarchique, les échelles les plus petites formant en premier des objets liés. Ces objets s'attirent, se rassemblent, s'assemblent et fusionnent pour former un nouveau niveau de la hiérarchie. Contrairement au scénario HDM des crêpes, on forme d'abord des protogalaxies, puis des galaxies, des amas puis des superamas, ces derniers étant aujourd'hui en cours de formation. Les premiers succès vont rapidement s'accumuler. Dès 1984, Blumenthal et al. [48] montrent qu'un certain nombre de propriétés des galaxies, celles reliées aux propriétés des baryons, s'expliquent naturellement dans ce contexte, tandis que les premières simulations numériques de l'évolution de la matière noire effectuées par Davis, Efstathiou, Frenck et White [49] confirment le potentiel de ce scénario. Un mouvement d'intérêt durable vient d'être lancé, même si, bien sûr, d'autres solutions sont explorées en parallèle par une partie de la communauté. En tout état de cause, il est devenu acceptable, voire naturel, pour beaucoup de postuler que l'essentiel de la matière dans l'Univers est sous une forme inconnue (dans le laboratoire, pas dans les articles des théoriciens). C'est une démarche qui n'est pas sans précédent ; ainsi Dirac n'a pas hésité en 1928 à postuler l'existence du positron, pour assurer la cohérence de sa théorie relativiste de l'électron. Mais une autre solution pourrait consister à modifier la théorie elle-même (outre évidemment l'autre possibilité qui consiste à supposer que les observations du moment sont erronées ou entachées d'erreurs systématiques).

C'est au même moment qu'est introduite une autre idée d'influence durable, celle de l'inflation [50–52]. Il s'agit d'une phase d'expansion exponentielle qui aurait eu lieu très tôt dans l'histoire de l'Univers. L'attrait essentiel de cette supposition est que cela permet de proposer un mécanisme physique de génération du spectre primordial de fluctuations. En bref, des fluctuations d'un champ quantique seraient étirées par l'expansion jusqu'à des échelles véritablement cosmologiques, avec des amplitudes approximativement invariantes d'échelle. Outre d'offrir un mécanisme plausible de génération dynamique de conditions initiales « naturelles », cette théorie permet de comprendre l'absence constatée de monopoles magnétiques. Ces derniers sont en effet engendrés naturellement dès qu'on replace les tentatives théoriques de description unifiée des forces fondamentales dans le contexte cosmologique du modèle standard de Big Bang chaud. L'introduction de cette théorie a aussi consacré auprès d'une communauté grandissante le lien entre physique microscopique et cosmos macroscopique, en montrant que la cosmologie permet de sonder des échelles d'énergie à tout jamais inaccessibles directement, si ce n'est dans l'Univers.

Au cours de la décennie suivante, entre 1985 et 1995, les cosmologues vont étudier en détail le scénario CDM « standard » (c'est-à-dire tel que je l'ai décrit ci-dessus) et explorer vigoureusement d'autres possibilités, notamment les conséquences de l'existence potentielle d'autres objets spéculatifs, comme les cordes cosmiques [53] ou les textures. Ceux-ci, comme les monopoles, peuvent apparaître quand on considère les tentatives théoriques d'unification des forces en cosmologie. Exprimant en 1988 et 1990 leur point de vue sur l'état du sujet dans le magazine « Nature », Peebles et Silk considèrent ainsi, sans les départager observationnellement, cinq scénarios concurrents [54, 55]. Tous présentent des difficultés sur un point ou un autre. La principale difficulté est bien sûr de faire le tri entre de multiples observations, potentiellement entachées d'erreurs systématiques et partiellement contradictoires, pour décider du problème décisif à résoudre. Certains résultats observationnels se détachent cependant. La révélation des grandes structures de l'Univers telles qu'on les entend aujourd'hui (c'est-à-dire que les galaxies semblent dessiner l'ossature d'une éponge), a frappé les imaginations lorsque Valérie de Lapparent, Margaret Geller et John Huchra ont publié en 1986 [56] les premiers résultats de l'extension du relevé de galaxies dit CfA2 du Centre d'Astrophysique de Harvard. Dans les termes mêmes des auteurs : « *La distribution des galaxies de l'échantillon, qui contient 1100 galaxies dans une tranche de 6 deg × 117 deg passant par l'amas de Coma, ressemble à une coupe au travers des bulles de savon d'un évier de cuisine* » ! Plus le scénario CDM est étudié, et plus certaines contradictions deviennent apparentes. D'abord, en dépit de nombreuses recherches, la densité de matière à grande échelle semble insuffisante par rapport à ce que le modèle postule. Les vitesses de la matière semblent aussi beaucoup plus corrélées à grande échelle que ne le prédit le scénario [57, 58]. Enfin, Maddox et collaborateurs trouvent en 1990 que l'amplitude des fluctuations à grande échelle du relevé de galaxies APM outrepassent nettement les prédictions du modèle

CDM standard. Les difficultés sont devenues patentes, car la conclusion est observationnellement solide et le lien avec la théorie relativement directe aux échelles considérées.

L'étape décisive sera franchie quand George Smoot et ses collaborateurs de l'instrument DMR à bord du satellite COBE annoncent en 1992 la première mesure indiscutable des anisotropies du rayonnement fossile [59, 60] qui attestent de l'empreinte des fluctuations primordiales. Le niveau détecté (fluctuations d'écart quadratique moyen  $\delta T/T \simeq 10^{-5}$  aux échelles angulaires supérieures à 10 deg) est très inférieur à ce qui est nécessaire pour engendrer les structures dans un univers baryonique naturel. En revanche, il diffère d'un facteur inférieur à deux de celui prédit par le scénario CDM. Une telle concordance approchée est sans doute un succès. Mais c'est une victoire à la Pyrrhus. En effet, les incertitudes théoriques et observationnelles sont déjà devenues suffisamment faibles pour que cette différence soit largement significative ; elle signe en fait la fin du scénario CDM dans sa version « standard ». Le scénario semble pourtant à même de capturer une part essentielle de l'histoire du développement des structures ; mais alors, quelle partie modifier ?

Tout cela prépare un deuxième changement de paradigme, qui me semble avoir eu lieu approximativement entre 1995 et 2000. Il s'agit d'abord d'entériner le fait que la quasi totalité des indications observationnelles concernant la densité de matière pointent vers une valeur qui n'est environ que le tiers de la valeur critique. La densité critique est celle qui conduit à la géométrie spatialement plane du modèle d'univers de type Einstein-de Sitter. Le scénario CDM suppose que la densité de matière est critique, au moins dans sa version minimale, que nous dénoterons dorénavant par SCDM (S pour Standard, pour la distinguer d'autres variantes). Une modification possible de la théorie minimale est de diminuer la valeur de la quantité de matière que le modèle suppose, pour la mettre plus en conformité avec ce qu'indiquent maintenant plus clairement les observations. Comme nous le verrons, cela conduit à un modèle d'univers dit ouvert, où la géométrie spatiale n'est plus euclidienne ; elle est de type hyperbolique (modèle dit OCDM, pour *Open CDM*). Une autre option consiste à préserver la nature plate de la géométrie spatiale, en acceptant la valeur de la densité de matière indiquée par les observations. Ceci n'est possible que si un autre terme, de type « constante cosmologique »,  $\Lambda$ , est non nul dans les équations d'Einstein et possède une valeur précise (dont l'explication est un défi très difficile). Nous y reviendrons longuement. Quoi qu'il en soit, ce modèle dit  $\Lambda$ CDM va rapidement devenir le modèle de référence, comme le proposent par exemple Ostriker et Steinhardt en 1995 [61]. Ce n'est certes pas la première fois qu'est évoquée la possibilité d'une constante cosmologique non nulle comme solution possible d'une difficulté en cosmologie. À titre d'exemples relativement récents, Peebles en 1984 [62], Vittorio et Silk en 1985, Efstathiou et al. en 1990 rappellent que le modèle  $\Lambda$ CDM permet de rapprocher les prédictions fondées sur l'inflation de diverses observations (entre autres la faiblesse relative des anisotropies du rayonnement fossile, et des fluctuations à grande échelle du relevé APM). Cette hypothèse

est cependant loin d'être la seule façon possible de remédier au problème. Simplement, la communauté se convainc majoritairement que c'est la meilleure hypothèse de travail du moment.

Deux faits observationnels majeurs ont aidé puissamment au basculement d'opinion. Le premier est la mesure de la façon dont diminue l'éclat de supernovae en fonction de leur éloignement (paragraphe 7.3.4), mesure qui est sensible à la géométrie de l'espace-temps. Les résultats de mesure publiés à partir de 1998 [63, 64] s'interprètent naturellement dans le cadre d'un modèle  $\Lambda$ CDM et tendent à exclure les modèles OCDM ou SCDM. Par ailleurs, de nouvelles mesures des anisotropies du rayonnement fossile, à des échelles plus petites que celles faites par DMR, apportent, à partir de 2000, un fort support à l'hypothèse d'univers spatialement plat [65–70], tout en confirmant une prédiction cruciale et générique des modèles considérés. Pris conjointement, ces résultats indiquent de manière convaincante que nous avons maintenant un modèle d'univers qui peut sembler bizarre, mais qui « marche ». Depuis l'an 2000, nous assistons à une *phase de consolidation* des résultats. En particulier, de nouvelles mesures sur les supernovae lointaines ont considérablement affiné le résultat de 1998 (paragraphe 7.3.4) et levé un certain nombre de doutes [71, 72]. Pour couronner le tout, l'équipe du satellite WMAP a dévoilé en 2003 [73] ses résultats concernant les anisotropies du rayonnement fossile ; ils confirment entièrement, tout en les affinant, les résultats antérieurs. La surprise majeure est sans doute qu'étant donné le gain de précision des mesures, il n'y a essentiellement pas eu de surprise, et qu'aucune incohérence n'est apparue, alors que la possibilité en était bien réelle.

Ces résultats récents indiquent un Univers à la composition étrange, où la matière atomique ordinaire ne constitue que quelques pour cent de la densité moyenne ; celle-ci apparaît dominée par une matière sombre inusuelle, non-baryonique, qui reste à détecter en laboratoire. L'effet sur la dynamique de l'ensemble de cette matière sombre semble à son tour presque marginal devant l'effet d'une constante cosmologique, ou d'une énergie du vide selon certaines interprétations, dont la magnitude reste à expliquer. Les fluctuations initiales en revanche semblent relativement naturelles, de statistique gaussienne, adiabatiques, pratiquement invariantes d'échelle, en bref telles qu'une théorie comme l'inflation sait en rendre compte. Sans vouloir exclure la possibilité d'une révision en profondeur de nos conceptions, ce modèle surprenant est néanmoins très convaincant au vu des informations disponibles. L'ambition de la suite de ce texte est de fournir des clés plus précises pour apprécier la solidité des indications observationnelles qui le supportent, et de pointer les difficultés annonciatrices de progrès ou de révisions à venir.

## 7.3 Cosmographie

Les éléments de relativité qui suivent sont destinés à fournir au lecteur une base pour mieux comprendre la nature des tests observationnels qui fondent le

modèle actuel. Il s'agit donc d'exposer les rudiments essentiels à une première approche de la cosmologie, sans même exposer les raisons qui ont amené Einstein à cette théorie. Il s'avère que des mathématiques élémentaires suffisent pour saisir la plupart des notions qui nous importent.

### 7.3.1 Relativité et géométrie

La relativité générale nous propose de passer d'une conception du monde où l'espace est absolu et rigide, comme chez Newton, à un espace qui a sa dynamique propre. Cette théorie est indispensable à une description cosmologique, car elle permet de traiter des phénomènes gravitationnels à l'échelle du cosmos, là précisément où la gravité domine. Elle offre une description relativiste, au sens de la relativité restreinte qu'elle englobe, de la gravitation.

La relativité restreinte s'ensuit de la finitude de la vitesse de la lumière,  $c$ , qui est postulée constante dans le vide pour tout observateur, quel que soit son mouvement. Si l'on note  $x$  une des coordonnées d'espace, et  $t$  la coordonnée temporelle, la seule relation linéaire possible reliant des écarts infinitésimaux entre ces coordonnées (mesurés par des observateurs en déplacement relatif), est la transformation de Lorentz,

$$dx' = \gamma \left( dx - \frac{v}{c} cdt \right) \quad (7.5)$$

$$cdt' = \gamma \left( cdt - \frac{v}{c} dx \right) \quad (7.6)$$

pour une vitesse  $v$  de déplacement des observateurs (et  $\gamma = (1 - v^2/c^2)^{-1/2}$ ). Cette transformation rend explicite que l'espace et le temps sont et doivent être traités sur un pied d'égalité. Elle manifeste qu'aucune des séparations de l'un ou l'autre n'est préservée séparément (sauf précisément dans la limite non-relativiste  $v \ll c$ , où l'on retrouve la conservation séparée des intervalles de distance et de durée). En conséquence, il va falloir remplacer les points de l'espace Newtonien par des événements de l'espace-temps, et la quantité conservée dans une transformation sera un intervalle mixte. Cet intervalle, qu'on appelle la métrique de l'espace-temps, est ce qui permet de calculer des intervalles entre le « là-bas autrefois » et le « ici maintenant ». Nous avons d'ailleurs déjà rencontré de tels intervalles spatio-temporels quand nous avons remarqué que la distance de l'observateur à un objet éloigné peut s'exprimer sous la forme d'un temps de parcours de la lumière pour nous parvenir, ce qui matérialise qu'un objet éloigné est en fait vu tel qu'il était au moment de l'émission de la lumière qui nous parvient. Nous enregistrons donc bien un événement, ou une série d'événements, de l'espace-temps.

La forme de l'intervalle entre deux événements révèle la structure géométrique de l'espace-temps sous-jacent, ses propriétés métriques intrinsèques. Dans le cas de la relativité restreinte, le carré de l'intervalle infinitésimal de



l'espace-temps, sa métrique  $ds^2$ , est donné<sup>12</sup> par

$$ds^2 = c^2 dt^2 - d\ell^2 = c^2 dt^2 - (dx^2 + dy^2 + dz^2) \quad (7.7)$$

qui fonde en un objet unique la définition des intervalles infinitésimaux de distance  $d\ell$  et de durée  $dt$ . Cette métrique, dite de Minkowski, est identique partout, la définition est la même en tout point. Cela rend manifeste la géométrie figée de l'espace-temps de la relativité restreinte. La spécificité de la relativité générale, c'est que cette géométrie peut être plus générale, et qu'elle peut évoluer dans le temps : l'Univers possède une structure et une histoire, et c'est le contenu énergétique de l'Univers qui les détermine. C'est précisément ce dernier point qui permet de décrire les effets de la gravité.

L'espace de la relativité restreinte est plat, euclidien, ce qui se traduit par le fait qu'une séparation entre deux points (calculée au même instant,  $dt = 0$ ) est donnée par la métrique euclidienne

$$d\ell^2 = dx^2 + dy^2 + dz^2 \quad (7.8)$$

qui généralise à trois dimensions la relation familière de Pythagore sur la valeur de l'hypoténuse d'un triangle. En coordonnées sphériques, on peut écrire de manière équivalente

$$d\ell^2 = dr^2 + r^2 d\Omega^2 \quad (7.9)$$

si l'on pose  $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ , la distance calculée entre deux points étant bien sûr indépendante du système de coordonnées utilisé. L'espace peut néanmoins être courbe, non-euclidien, et cela d'une infinité de façons. Cependant s'il est homogène et isotrope, il n'existe en fait que trois types différents de métrique possible, repérés par le paramètre de courbure  $k$ ,

$$d\ell^2 = d\chi^2 + S_k^2(\chi) d\Omega^2 \quad (7.10)$$

avec  $S_{-1}(\chi) = \sinh \chi$ ,  $S_0(\chi) = \chi$ , et  $S_{+1}(\chi) = \sin \chi$ , et où  $\chi$  joue explicitement le rôle de la coordonnée radiale  $r$  en géométrie plate. On peut facilement se convaincre que c'est bien là la généralisation du cas familier à deux dimensions de la géométrie sur la sphère dans le cas  $k = +1$ , et sur une surface hyperbolique dans le cas  $k = -1$  (avec  $k = 0$  pour le cas plan, bien sûr). Notons au passage que la métrique spatiale permet aussi de calculer surfaces et volumes, là aussi par intégration des infinitésimaux. À titre d'exemple, l'élément de volume  $d^3V = dx dy dz = dr r^2 d\theta \sin \theta d\phi = r^2 dr d\Omega^2$  en espace euclidien devient ainsi  $d^3V = S_k^2(\chi) d\chi d\Omega^2$  en espace courbe.

L'espace dont nous venons de parler n'est qu'une section spatiale, prise à  $t = \text{constante}$ , de l'espace-temps. L'espace-temps pseudo-euclidien de la relativité restreinte, n'est qu'un cas particulier appelé espace-temps de Minkowski.

---

<sup>12</sup>On peut également choisir une autre convention, selon laquelle la métrique est définie par  $ds^2 = d\ell^2 - c^2 dt^2$ , comme dans le texte de T. Damour.

Dans le cas de la relativité générale, la seule métrique compatible avec l'hypothèse d'homogénéité et d'isotropie, et sans perte de généralité, est la métrique de Robertson-Walker [74, 75],

$$ds^2 = c^2 dt^2 - a^2(t) d\ell^2 \quad (7.11)$$

où  $a(t)$  est une fonction arbitraire du temps appelée facteur d'échelle (et où le  $d\ell^2$  de la partie spatiale de la métrique est toujours donnée par l'équation (7.10)). La séparation spatiale dite *propre* de deux points quelconques,  $d_{propre}$ , s'obtient en intégrant l'élément métrique  $ds$  des coordonnées de l'un à celles de l'autre, en les considérant au même instant ( $dt = 0$ ). Cette séparation croît donc comme  $a$ . Si  $a$  croît avec le temps, cet univers sera en expansion, car en l'absence de tout mouvement propre des points, toutes les séparations physiques vont croître d'un même facteur global. Un volume va quant à lui croître comme  $a^3$ . Remarquons de plus que le temps propre,  $\tau$ , celui mesurant la durée d'un phénomène là où il a lieu ( $d\tau^2 = ds^2 = cdt^2$ ), peut être identifié à la coordonnée temporelle  $t$  du système de coordonnées de notre monde homogène et isotrope.

Nous pouvons maintenant revenir à une difficulté fondamentale que nous avons ignorée quand, dans la première partie, nous parcourions les chemins de l'univers. Que veut dire exactement la distance dans un monde en expansion ? Il s'avère particulièrement commode d'introduire la notion de position comobile d'un objet, qui est la position qu'aurait cet objet aujourd'hui, *s'il avait suivi l'expansion cosmique générale*. On définit alors une distance comobile correspondante,  $d_{comob}$  qui est reliée à la distance propre par

$$d_{comob} = \frac{a_0}{a(t)} d_{propre}(t) \quad (7.12)$$

où comme dans toute la suite l'indice 0 fait référence à une quantité prise aujourd'hui (et donc  $a_0 \equiv a(t_0)$ ). Un des avantages de la distance comobile, c'est qu'elle est constante, alors que la distance propre augmente avec le temps. Les quantités comobiles dérivées (volume, densité, etc.) seront donc conservées par l'expansion.

Les photons suivent des géodésiques, d'équation  $ds^2 = 0$ , ce qui donne  $ad\ell = c dt$ , qui est bien l'équation usuelle hormis la présence du facteur d'échelle. Pour une distance radiale comme celle de la figure 7.1 ( $\theta = \text{constante}$ ,  $\psi = \text{constante}$ , et donc  $d\Omega = 0$ ), on a alors  $a d\chi = c dt$ , ce qui montre que  $a_0\chi = c \int a_0/a dt$  est bien la distance radiale comobile. Les objets de la figure 7.1 sont donc représentés à la position qu'ils occupent aujourd'hui s'ils n'ont pas de mouvement propre se superposant au mouvement d'expansion d'ensemble. De tels mouvements existent, mais leur magnitude est le plus souvent négligeable par rapport au mouvement d'expansion cosmologique.

Le décalage vers le rouge des spectres est une conséquence directe de l'expansion. Ainsi un observateur local (2) qui reçoit la lumière d'un objet fixe (1) situé en  $\chi = c \int_{t_1}^{t_2} dt/a$  va observer le rayonnement émis une période plus

tard, à  $t_1 + T_1$ , avec une période différente  $T_2$  telle que

$$\int_{t_1}^{t_2} dt/a = \int_{t_1+T_1}^{t_2+T_2} dt/a \iff \int_{t_1}^{t_1+T_1} dt/a = \int_{t_2}^{t_2+T_2} dt/a \quad (7.13)$$

ce qui donne  $T_1/a_1 = T_2/a_2$ , compte tenu du fait qu'on peut certainement négliger la variation du facteur d'échelle sur une période. La fréquence  $\nu = 1/T$  du rayonnement diminue donc au fur et à mesure de l'expansion et nous obtenons donc que le décalage de fréquence  $z \equiv (\nu_1 - \nu_2)/\nu_2$  que nous avons défini à l'équation (7.1), satisfait la relation simple

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{a_2}{a_1} \quad (7.14)$$

Remarquons au passage que pour des séparations pas trop grandes, on a  $r \simeq a \delta\ell$ , et on retrouve alors la loi de Hubble (7.2)

$$v \equiv \frac{dr}{dt} \simeq \frac{1}{a} \frac{da}{dt} a \delta\ell = Hr \quad (7.15)$$

si l'on définit  $H$  comme la dérivée temporelle du logarithmique du facteur d'échelle.

Notons que pour obtenir cette loi, nous avons dû considérer des séparations suffisamment faibles pour qu'on puisse négliger la variation du facteur d'échelle pendant le temps de parcours de la lumière. De même, pour des décalages spectraux faibles  $z \ll 1$ , on a  $z = \delta a/a = 1/a \delta a/\delta t \delta t$ , ce qui donne  $z = Hr/c$ . On voit bien qu'avec  $z$  donnant la vitesse de récession d'un objet en unité de la fraction de la vitesse de la lumière, l'interprétation du décalage vers le rouge comme un effet Doppler est, tout comme la loi de Hubble, une approximation qui ne tient que pour l'Univers local. Pour des distances plus lointaines, il faudra prendre explicitement en compte la variation temporelle du facteur d'échelle, ce qui implique de supposer un modèle particulier. C'est pourquoi la plupart du temps les astronomes analysent les grandes structures en utilisant directement le  $z$  observé comme mesure d'éloignement.

L'expansion joue d'autres tours aux astronomes. À propos de l'utilisation de la relation Période-Luminosité des céphéides, j'avais indiqué au paragraphe 7.1.2 que l'observateur perçoit la luminosité apparente de l'étoile, qui diminue comme le carré de sa distance, car la quantité d'énergie qu'elle émet sous forme de lumière est répartie sur une surface croissante au fur et à mesure de l'éloignement de l'étoile. Mais de quelle distance s'agit-il? En utilisant la métrique, on trouve que l'éclat apparent  $F$  observé à la fréquence  $\nu_2$  a été émis à  $\nu_1 = (1 + z)\nu_2$ , ce qui donne

$$F(\nu_2) = \frac{L((1+z)\nu_2)}{4\pi a_0^2 S_k^2(\chi)(1+z)} \quad (7.16)$$

où on a tenu compte du fait que la densité de flux reçu est la luminosité absolue  $L(\nu_1)$  répartie sur une surface  $4\pi a_0^2 S_k^2$ , et diminuée d'un facteur  $(1+z)$  dû

au changement de fréquence des photons. On ne retrouve la forme usuelle,  $F = L/(4\pi d_L^2)$ , que si l'on pose que  $d_L$  est une distance-luminosité *définie* par

$$d_L = a_0 S_k(\chi)(1+z) \quad (7.17)$$

Même en espace plat, quand  $S_0(\chi) = \chi$ , cette distance n'est égale à la distance comobile  $a_0\chi$  que quand  $z$  est négligeable. Similairement, on définit la distance angulaire,  $d_A$ , comme l'expression qui, multipliée par l'angle  $d\psi$  sous-tendu par un objet, donne la taille transverse propre de l'objet,  $d\ell = a S_k(\chi)d\psi = a_0/(1+z)S_k(\chi)d\psi$ . On déduit alors que

$$d_A = a_0 S_k(\chi)/(1+z) \quad (7.18)$$

qui diffère de  $d_L$  par un facteur  $(1+z)^2$ . Il est donc clair que dès que l'on considère des objets lointains, il faut prendre soin de préciser de quelle distance il s'agit, par quel moyen elle a été obtenue. En retour, la comparaison de distances sur une large plage offre en principe la possibilité de contraindre l'histoire de l'expansion, et donc le modèle cosmologique.

### 7.3.2 Dynamique de l'Univers

Pour décrire l'effet de la gravité, Einstein va la promouvoir du statut d'acteur de l'action à celui de metteur en scène, qui agit sur le théâtre de l'action, c'est-à-dire sur les propriétés géométriques de l'espace-temps. Les équations d'Einstein relient les caractéristiques géométriques, déterminées par les coefficients de la métrique, au contenu matériel de la scène. Le contenu transforme donc l'environnement, ce qui peut le transformer à son tour. En cosmologie, on peut décrire le contenu très simplement comme un fluide homogène, muni d'une densité d'énergie,  $\rho$ , et d'une pression,  $p$ , qui sont identiques en tous lieux puisqu'on ne considère ici que des modèles homogènes et isotropes. Dans ce cas, les équations d'Einstein se ramènent à deux équations simples qui déterminent l'évolution du facteur d'échelle de la métrique de Robertson-Walker.

La première équation donne l'accélération  $\ddot{a}$ ,

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{4\pi G}{3}(\rho + 3p) \quad (7.19)$$

où on a utilisé un système d'unités où  $c = 1$ , et la dérivée temporelle a été repérée par un point. On voit qu'une constante cosmologique positive,  $\Lambda > 0$ , entraîne une accélération, alors qu'en revanche densité et pression tendent à décélérer la variation. La seconde équation est *l'équation de Friedman*

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho_{tot}}{3} + \frac{\Lambda}{3} - \frac{k}{a^2} \quad (7.20)$$

L'indice *tot* de la densité rappelle que dans le cas de plusieurs fluides, il convient de prendre en compte toutes les contributions,  $\rho_{tot} = \sum_i \rho_i$ .

En combinant ces deux équations, on déduit l'équation de conservation de l'énergie (qui traduit l'invariance par translation)

$$\frac{d}{da}(\rho a^3) = -3pa^2 \quad (7.21)$$

En effet, cette équation indique que la variation d'énergie dans un volume comobile est égale au produit de la pression par la variation de volume comobile  $d(a^3)/da$ , ce qui généralise la première loi de la thermodynamique. Enfin il manque encore, pour spécifier complètement le modèle, une *équation d'état* reliant la pression à la densité :

$$p = f(\rho) \quad (7.22)$$

On appelle modèles de Friedman-Lemaître les modèles d'univers qui sont décrits par une métrique de Robertson-Walker et dont l'évolution est donnée par les équations ci-dessus.

Une fois spécifiée l'équation d'état (7.22), l'équation (7.21) permet de calculer  $\rho$  en fonction de  $a$ , ce qu'on peut alors substituer dans (7.20) pour obtenir l'évolution temporelle du facteur d'expansion lui-même. Mais nous pouvons déjà remarquer que, si  $\rho + 3p > 0$ , la première équation impose qu'en l'absence de  $\Lambda$ , l'accélération  $\ddot{a}/a$  soit négative. Comme le facteur d'échelle est par définition positif ( $a > 0$ ), et comme nous observons plutôt des décalages vers le rouge que vers le bleu ( $\dot{a}/a > 0$ ), nous en concluons que  $a(t)$  est une fonction convexe du temps et qu'il existe un temps  $t_i$  tel que  $a(t_i) = 0$  (qu'on convient de prendre  $t_i = 0$ ). Ceci implique donc l'existence d'une singularité passée, et une limite supérieure sur l'âge de l'Univers, quelle que soit l'équation d'état :  $t < H^{-1}$ . Nous venons de retrouver une des grandes déductions de la cosmologie relativiste : l'Univers a un âge fini, et il existe en principe dans le passé un point singulier où le facteur d'échelle s'annule. Bien sûr cette conclusion ne tient que si la théorie utilisée pour l'obtenir reste valable dans les conditions extrêmes qui prévalaient aux premiers instants du « Big Bang ».

Remarquons qu'il est possible d'éclairer le sens des équations (7.19) et (7.20) par une dérivation dans le cadre classique et plat. Soit  $r(t) = a(t)x$ , où  $x$  est une coordonnée comobile. Écrivant l'équation fondamentale de la dynamique et le théorème de Gauss pour une distribution sphérique de matière, on a  $\ddot{r} = -GM/r^2$ , d'où  $\ddot{a} = -4\pi G(\rho a^3)/3a^2$ , ce qui donne (7.20) par intégration. Et l'équation de conservation de l'énergie s'écrit  $d(\rho V) = -pdV$ , ce qui redonne bien (7.21). Cet argument peut être raffiné pour le rendre rigoureux. On notera toutefois qu'outre la présence de la constante cosmologique, l'approche relativiste diffère de l'approche classique par le fait que *la pression intervient comme source de l'accélération*, sur le même plan que la densité d'énergie, ce qui est un concept relativiste sans équivalent classique.

Pour nous familiariser avec ces équations, commençons par considérer le cas où  $\Lambda$  est négligeable. L'équation de Friedman montre qu'il existe une

densité critique

$$\rho_c \equiv \frac{3H^2}{8\pi G}, \quad \text{avec } \rho_{c0} = 1,88 \times 10^{-26} h^2 \text{ kg} \cdot \text{m}^{-3} = 10,5 h^2 \text{ eV} \cdot \text{cm}^{-3} \quad (7.23)$$

pour laquelle l'équation est satisfaite quand  $k = 0$ , c'est-à-dire dans le cas d'une métrique spatialement plate (cf. Éq. (7.10)). Il est alors commode de définir le paramètre de densité cosmologique  $\Omega$  comme la densité d'énergie relative à la densité critique,

$$\Omega = \frac{\rho}{\rho_c} \quad (7.24)$$

On peut alors écrire l'équation de Friedman sous la forme simple

$$\frac{k}{a^2} - \frac{\Lambda}{3} = H^2(\Omega - 1) \quad (7.25)$$

Introduisons maintenant différents types de contributions qui toutes, nous allons le voir, permettent de décrire une des phases d'évolution de l'Univers tel que nous le comprenons aujourd'hui. Soit une équation d'état

$$p = w\rho \quad (7.26)$$

Si le paramètre  $w$  de l'équation d'état est constant, l'équation de conservation (7.21) prend la forme  $\dot{\rho} = -3(1+w)\rho\dot{a}/a$ , ce qui donne  $\rho \propto a^{-3(1+w)}$ . Et si cette composante domine toutes les autres formes de densité d'énergie ( $\rho_{tot} \simeq \rho_w$ ), au moins pendant une certaine durée, l'équation de Friedman implique alors dans le cas plat<sup>13</sup>,  $k = 0$ , une expansion en loi de puissance du temps,

$$a(t) \propto t^{\frac{2}{3(1+w)}} \quad (7.27)$$

si  $w \neq -1$ . Dans le cas de matière sombre froide sans pression, on a  $w = 0$ , et donc  $\rho_m \propto a^{-3}$ , qui traduit que la densité d'énergie diminue en raison inverse du volume (qui lui, croît comme  $a^3$ ). On a donc  $\Omega_m = \Omega_{m0}(1+z)^3$  (et  $a \propto t^{2/3}$ ). Ceci définit le comportement du modèle Einstein-de Sitter de 1932. Dans le cas du rayonnement, comme le rayonnement fossile (ou tout autre forme de matière relativiste), on a plutôt  $p = \rho c^2/3$  ( $w = 1/3$ , car nous utilisons  $c = 1$ ) et donc  $\rho_r \propto a^{-4}$  et  $\Omega_r = \Omega_{r0}(1+z)^4$ . On retrouve là l'effet du décalage vers le rouge de la fréquence du rayonnement, qui ajoute une décroissance supplémentaire en  $a$  à l'effet de volume. Dans le cas où ce serait plutôt le terme de constante cosmologique qui serait dominant ( $w = -1$ ), l'équation de Friedman se réduit à  $H^2 = \Lambda/3$ . On retrouve alors la phase de croissance exponentielle du facteur d'échelle du modèle introduit par de Sitter dès 1917 ( $a \propto \exp(Ht)$ ).

---

<sup>13</sup>Le cas plat est toujours une bonne approximation aux temps suffisamment reculés car dans l'équation de Friedman (7.20), le terme de courbure  $k/a^2$  devient négligeable devant le terme  $\propto \rho_{tot}$ , qui augmente au moins comme  $a^3$ .

Depuis d'ailleurs la même époque, les physiciens discutent du niveau d'énergie fondamental des champs physiques comme le champ électromagnétique, puis plus tard les champs quantiques. Dans le laboratoire où on mesure toujours des différences de niveau d'énergie, les valeurs très élevées, voire divergentes, obtenues n'ont pas trop d'importance. Mais en relativité générale, toute énergie doit a priori être prise en compte dans le recensement du contenu de l'Univers; l'énergie du fondamental, souvent appelée énergie du vide (de particules stables) va donc influencer sur la géométrie de l'espace-temps. Et Lemaître [76] remarque qu'il faut associer une pression  $p_v = -\rho_v c^2$  à cet état de vide pour ne pas définir un référentiel absolu dans lequel on pourrait repérer un mouvement par rapport au vide. On a donc un paramètre de l'équation d'état  $w = -1$  et on retrouve le cas de Sitter. On voit donc qu'une constante cosmologique et l'énergie du vide ont a priori un effet cosmologique identique, même si leurs origines physiques n'ont rien à voir : la constante cosmologique est un terme géométrique qui fait partie intégrante des équations d'Einstein; il spécifie la façon dont le contenu agit sur la géométrie. L'énergie du vide est un terme de contenu qui, en outre, est susceptible d'évolution au cours de l'histoire de l'Univers. Pour simplifier la discussion, nous noterons la contribution jointe de la constante cosmologique et de l'énergie du vide par

$$\Omega_\Lambda = \frac{\rho_v}{\rho_c} + \frac{\Lambda}{3H^2} = \frac{8\pi G\rho_v + \Lambda}{3H^2} \quad (7.28)$$

et nous y ferons référence globalement en parlant d'énergie noire. L'équation de Friedman s'écrit alors  $k/(aH)^2 = \Omega_m + \Omega_r + \Omega_\Lambda - 1$ , qui traduit l'absence de courbure ( $k \neq 0$ ) quand  $\Omega_{tot} = \Omega_m + \Omega_r + \Omega_\Lambda = 1$ .

Il est aussi commode de définir  $\Omega_k = -k/(aH)^2$  afin d'obtenir alors une règle de somme facile à mémoriser,  $\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = \sum_i \Omega_i = 1$ . Cependant tous ces termes ont une dépendance temporelle différente, et se rappelant (7.14) ( $1 + z = a/a_0$ ), il est préférable en pratique d'écrire le taux d'expansion  $H(z) = \dot{a}/a$  explicitement en fonction des paramètres de densité actuels

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 [\Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + (1 - \Omega_{r0} - \Omega_{m0} - \Omega_{\Lambda0})(1+z)^2 + \Omega_{\Lambda0}] \quad (7.29)$$

Cette forme de l'équation de Friedman permet de faire plusieurs observations. Suffisamment tôt dans l'histoire de l'Univers, quand  $z \gg 1$ , la densité d'énergie doit être dominée par le rayonnement, dont le fond à 2,73 K est un fossile, même si aujourd'hui on a seulement<sup>14</sup>

$$\Omega_{r0} = 4,3 \times 10^{-5} h^{-2} \quad (7.30)$$

<sup>14</sup>La densité d'énergie du rayonnement cosmologique fossile est donnée par  $\rho_{\gamma,0} = a_S T^4$ , où  $a_S = \pi^2 k^4/[15c^2(\hbar c)^3]$  est la constante de Stefan, ce qui donne  $\Omega_{\gamma,0} = 2,6 \times 10^{-5} h^{-2}$ . La valeur de l'équation (7.30) correspond à la prise en compte de la contribution additionnelle de trois espèces de neutrinos et leurs antineutrinos, à une température de 1,95 K.

Cependant, il arrive un moment au cours de l'expansion où matière et rayonnement vont contribuer de manière équivalente, ce qui se produit à

$$1 + z_{eq} = \Omega_{m0}/\Omega_{r0} \simeq 2,3 \times 10^4 \Omega_{m0} h^2 \quad (7.31)$$

quand la température du rayonnement est  $T = T_0(1 + z_{eq}) \simeq 8,7 \times 10^3$  K (pour  $\Omega_{m0} h^2 = 0,14$ ). Commence alors l'ère de matière, suivie éventuellement par une expansion dominée par la courbure, avant que ne débute une phase de Sitter exponentielle, si  $\Omega_\Lambda \neq 0$ .

La figure 7.3 permet de comparer l'évolution temporelle du facteur d'échelle et de quantités dérivées comme  $H$ , le paramètre de décélération  $q \equiv -\ddot{a}a/\dot{a}^2$ ,  $z$ ,  $\Omega_m$  et  $\Omega_\Lambda$ , dans trois cas de figure ayant chacun  $H_0 = 71 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ . La courbe verte correspond à un univers Einstein-de Sitter [14], qui ne contient que de la matière à la densité critique  $\Omega_m = \Omega_{m0} = 1$ . C'est le cas  $w = 1/3$  discuté plus haut, qui décrit bien à  $z < z_{eq}$  l'évolution du modèle SCDM populaire au cours des années 80. Pour des  $z$  plus grands, il faut explicitement tenir compte du rayonnement, mais la différence est imperceptible dans un tel diagramme linéaire. L'âge de l'Univers aujourd'hui, quand  $a = a_0$ , est repéré par la ligne verticale verte pointillée à 9,2 milliards d'années (dans ce cas simple, on a  $a \propto t^{2/3}$  et donc  $t_0 = 2/3H_0$ ; cf. (7.27) et (7.4)). Les courbes en bleu correspondent au cas ouvert  $\Omega_{m0} = 0,27$  ( $k = -1$ ), avec un âge de 11,5 milliards d'années. Voilà qui est plus raisonnable par rapport à l'âge des plus vieilles étoiles que nous indiquent les modèles d'évolution stellaire. On utilisera l'acronyme OCDM (*Open CDM*) pour y faire référence dans la suite. Enfin la courbe rouge correspond à un univers sans courbure spatiale ( $k = 0$ ) où  $\Omega_{m0} = 0,27$  et  $\Omega_{\Lambda 0} = 1 - \Omega_{m0}$ , dont l'âge est de 13,8 milliards d'années. Ce modèle décrit bien à  $z < z_{eq}$  le modèle  $\Lambda$ CDM qui est la référence aujourd'hui. Dans ce dernier cas, on perçoit clairement l'inflexion de la courbe  $a(t)$  au passage très récent d'une phase décélérée à une phase d'accélération, quand  $q = \Omega_m/2 - \Omega_\Lambda = 0$  (avec  $q_0 \sim 0,6$ ). Ceci se produit à  $1 + z_\Lambda = (2\Omega_{\Lambda 0}/\Omega_{m0})^{1/3} = 1,76$ , vers  $t \sim 7$  Gans. Ce cas  $\Lambda$ CDM est précisément celui qui a été utilisé pour calculer les distances comobiles de la carte de l'Univers de la figure 7.1. Ces distances comobiles se calculent en remarquant que  $ad\chi = cdt = cda/\dot{a} = cda/(aH)$  et donc que

$$a_0 d\chi = c \frac{dz}{H(z)}. \quad (7.32)$$

Cette relation fondamentale permet de calculer l'éloignement comobile  $a_0\chi(z)$  d'un objet à un décalage vers le rouge  $z$ , en fonction des paramètres cosmologiques qui déterminent  $H(z)$  par l'équation (7.29). Cette dépendance est la raison pour laquelle les astronomes utilisent en général l'observable directe,  $z$ , y compris pour faire des cartes<sup>15</sup>, plutôt que la distance qui est une *reconstruction* susceptible de varier si le modèle cosmologique évolue.

<sup>15</sup>La dépendance temporelle du  $z$  d'un objet est en effet suffisamment faible pour être couramment négligée. Le modèle Einstein-de Sitter nous permet d'avoir aisément l'ordre de grandeur, puisque  $a \propto t^{2/3} \implies \delta z/z = 2/3 \delta t/t$ , et donc  $\delta z/z = 2\delta t/3t_0 \sim 5 \times 10^{-9}$  sur un siècle.



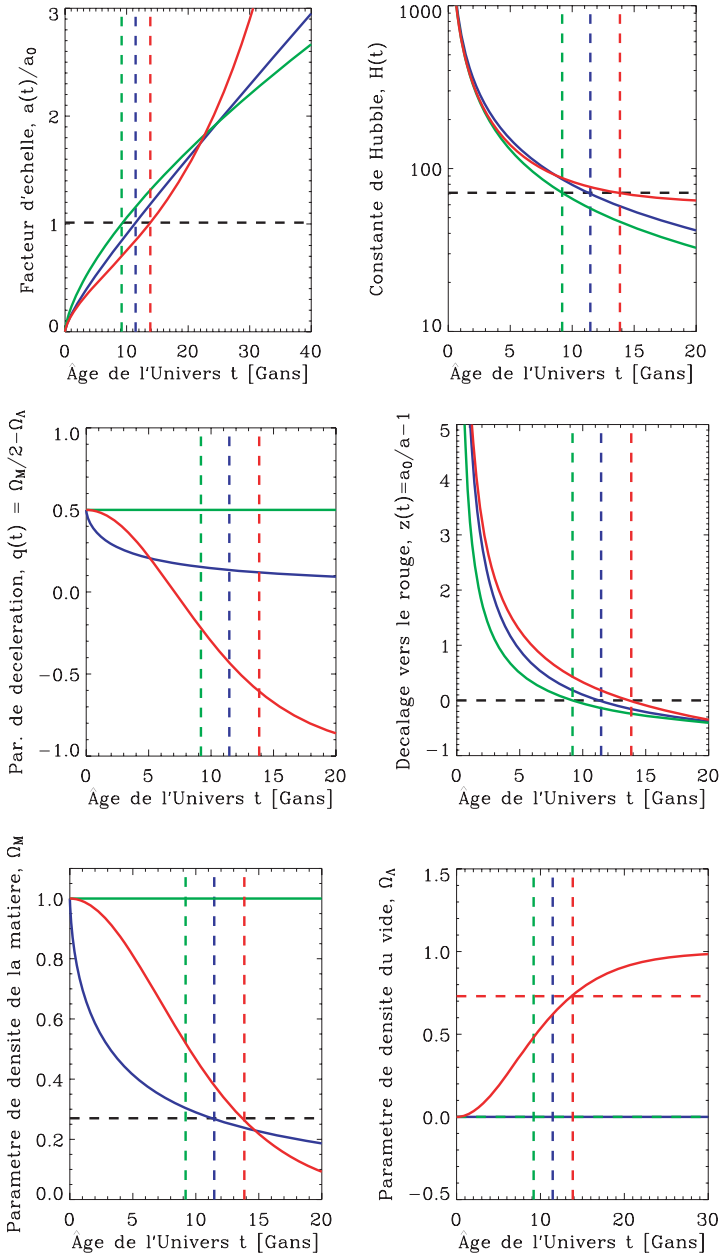


FIG. 7.3 – Évolution temporelle de paramètres cosmologiques dans trois modèles cosmologiques avec  $H_0 = 71 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ , correspondant à divers valeurs du couple  $(\Omega_m, \Omega_\Lambda)$  (courbes vertes pour  $(1, 0)$ , bleues pour  $(0, 27, 0)$  et rouges pour  $(0, 27, 0, 63)$ ). Les lignes pointillées repèrent l'âge actuel.

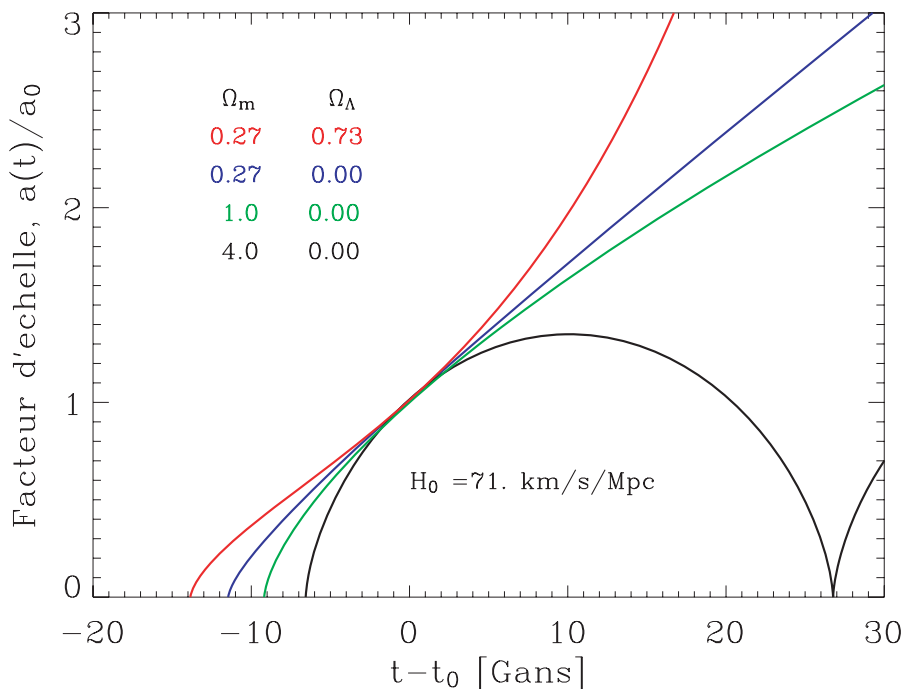


FIG. 7.4 – Évolution temporelle du facteur d'échelle dans divers modèles cosmologiques repérés par leur valeur de  $\Omega_m$  et  $\Omega_\Lambda$  aujourd'hui.

Après l'acceptation dans les années 1970 de l'existence d'une grande quantité de matière sombre, une partie importante des études de ces 30 dernières années a cherché à déterminer si la densité de matière sombre atteint ou non la densité critique, et si la géométrie spatiale est plate ou non. Plus précisément, on a cherché à savoir si  $\Omega_m \sim 0,3$  ou 1, et corrélativement si  $k = -1$  ou 0, tout en croisant ces indications directes avec d'autres observables cosmologiques, comme l'âge des plus vieilles étoiles. C'est pourquoi les trois types d'univers présentés furent les références incontournables de toutes les discussions sur le modèle d'univers. La figure 7.4 permet de comparer l'évolution du facteur d'échelle dans ces trois modèles, du point de vue de l'observateur, en prenant l'époque actuelle comme référence temporelle. J'ai aussi tracé le cas d'un modèle fermé  $\Omega_{m0} = 4$ , qui, bien qu'irréaliste, illustre bien la nature attractive de la matière dans l'équation d'accélération (7.19). Dans ce cas, le facteur d'échelle atteint une valeur maximale avant de diminuer pour aboutir à un « Big Crunch » en un temps fini. La courbe illustre de plus ce que pourrait être un univers cyclique, qui « rebondit » à proximité de la singularité future pour commencer une nouvelle ère. C'est d'ailleurs un modèle de ce type qu'étudiait l'équipe de Dicke à Princeton en 1965 quand le rayonnement relique fut découvert par Penzias et Wilson.

Un dernier point. Nous avons vu que la distance radiale comobile parcourue par un rayon lumineux est donnée par  $a_0\chi = c \int dt a_0/a = c \int dz/H$ . Remarquons que pour une expansion en loi de puissance,  $a(t) \propto t^\alpha$ , aucune information ne peut être propagée entre les temps  $t_i$  et  $t_0 \gg t_i$  sur une échelle comobile plus grande que

$$a_0\chi_H(t_0) = \frac{ct_0}{(1-\alpha)} \quad (7.33)$$

expression qui permet de voir comment l'expansion modifie la distance  $ct_0$  parcourue par la lumière dans un espace plat sans expansion. Comme le parcours a essentiellement lieu dans l'ère de matière, quand  $\alpha = 2/3$ , cette distance est environ trois fois plus grande qu'en ignorant l'expansion. En tout état de cause, le point important est qu'il existe un horizon causal, dont la taille croît avec le temps, mais qui n'en est pas moins fini. Le calcul exact dans le cas  $\Lambda$ CDM de la figure 7.1 donne  $a_0\chi_{H_0} = 14,28$  Gpc que visualise la ligne étiquetée « Big Bang » (à comparer à  $ct_0 = c \times 13,8$  Gans =  $13,8/3,26$  Gpc, ce qu'on peut écrire comme  $14,28/3,37$  Gpc). Cette figure 7.1 montre d'ailleurs aussi la limite comobile de visibilité future, la plus grande distance que nous pourrions jamais observer dans ce modèle; cette limite quand  $t$  tend vers l'infini de  $a_0\chi$  vaut 19,03 Gpc. Outre la distance correspondant à  $z_\Lambda = 0,75$  quand l'univers entame sa phase d'accélération, la figure 7.1 indique aussi le décalage vers le rouge remarquablement bas,  $z = 1,69$ , qui correspond à la plus grande distance ( $t \rightarrow \infty$ ) que pourra parcourir un rayon lumineux envoyé aujourd'hui, soit  $19,03 - 14,28 = 4,75$  Gpc. Cette distance est aussi faible parce que nous entrons dans une phase d'expansion exponentielle où les distances vont augmenter extrêmement vite avec le temps. Voilà une excellente illustration du supplice de Tantale de l'astronome, voir mais ne pas toucher, et en fait, plus radicalement, voir sans aucun espoir d'influencer jamais, si  $z > 0,75$ !

### 7.3.3 La matière est souvent sombre

Comme l'a montré la discussion précédente, la quantité de matière est une donnée cosmologique cruciale. Nous aimerions savoir en particulier s'il y a assez de matière pour que  $\Omega_m = 1$ . Le résultat de la confrontation des calculs de nucléosynthèse primordiale et de l'abondance observée de Deutérium indique une abondance baryonique  $\Omega_{b0} \simeq 0,04$ . Les analyses de l'absorption de la lumière de quasars lointains par la matière le long de la ligne de visée permettent d'inférer la présence d'une quantité similaire de baryons à  $z \sim 3 - 4$  [77]. En revanche, seulement un quart de ces baryons se retrouve localement, à  $z \sim 0$ , sous forme stellaire,  $\Omega_{*0} \simeq 0,01$ , le reste est invisible. Les baryons sont donc aujourd'hui en partie « noirs » et de nombreuses hypothèses sur différentes façons de les dissimuler (sous forme de petits trous noirs par exemple) ont progressivement été éliminées. En tout état de cause, les baryons ne peuvent constituer à eux seuls qu'une petite fraction de la densité cosmologique critique.

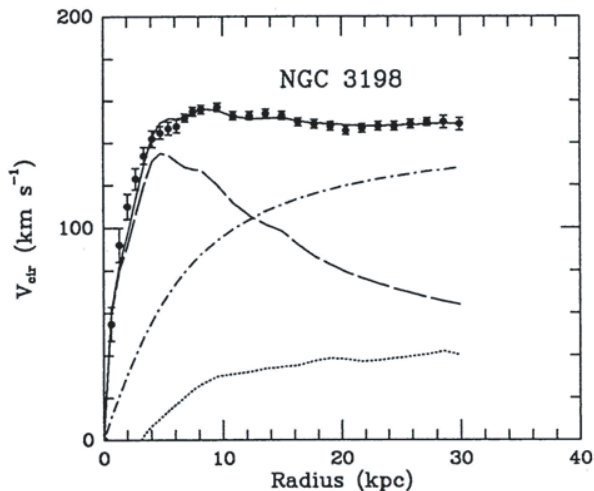


FIG. 7.5 – Une courbe de rotation plate, et sa décomposition en diverses contributions. La courbe tiretée  $\approx \sqrt{r}$  correspond à la contribution attendue des étoiles, celle en pointillés correspond à celle du gaz, tandis que la courbe en tirets-points indique la contribution (déduite) du halo de matière sombre, qui domine largement à grande distance. (D’après [78].)

Mais nous disposons aussi de nombreuses indications de l’existence de grandes quantités de matière qui ne manifeste sa présence que par l’entremise de ses effets gravitationnels. L’argument est en général le suivant : si un système est à l’équilibre, son énergie potentielle gravitationnelle doit être comparable à l’énergie cinétique de ses constituants. Ce qui donne  $v^2 = kGM(R)/R$ , où  $k$  dépend du détail de la distribution de matière. Dans cette formule,  $v$  est, selon le système considéré, la dispersion de vitesse du constituant (étoiles dans une galaxie elliptique, galaxies dans un amas), ou alternativement la vitesse de rotation s’il s’agit des étoiles et du gaz du disque d’une galaxie spirale. C’est le théorème « du viriel », d’origine Newtonienne, mais qui reste valable en relativité aux échelles ( $\ll cH^{-1}$ ) et aux vitesses ( $\ll c$ ) considérées. C’est en l’appliquant que Zwicky remarque en 1933 que la vitesse des galaxies dans l’amas de Coma est beaucoup trop grande pour que le système soit gravitationnellement lié. Afin que l’amas ne vole pas en éclats, il estime alors qu’il devrait y avoir 400 fois plus de masse que ne le laisse penser une évaluation sur la base du nombre d’étoiles dans l’amas.

La figure 7.5 montre un exemple de courbe de rotation en fonction de la distance au centre de la Galaxie NGC 3198. Les points indiquent la vitesse du gaz du disque  $\sim 150 \text{ km.s}^{-1}$  qu’on obtient par observation radiométrique du décalage Doppler de la raie à 21 cm de l’hydrogène neutre. S’il n’y avait que ce gaz, on devrait voir la ligne en pointillé. Pour évaluer la contribution des étoiles, on rassemble des données de photométrie de surface, qui indiquent

la densité surfacique de lumière dans le disque optique brillant. En supposant un rapport masse (associée aux étoiles)-luminosité,  $M/L$ , constant en moyenne, on peut calculer la vitesse de rotation attendue. On obtient alors la courbe tiretée à la décroissance en  $\sqrt{r}$  qui est prédite à grande distance si la distribution de masse est dominée par le bulbe central d'étoiles. On constate alors que, même pour un disque « maximal », c'est-à-dire dont on a choisi le  $M/L$  aussi grand que possible (sans toutefois que les prédictions n'excèdent les vitesses observées), il est clair qu'au-delà de quelques kiloparsecs, il y a tellement de masse manquante qu'elle domine complètement la dynamique. La courbe en tirets-points indique la contribution déduite du halo de matière sombre. L'indication est si directe que, outre l'introduction de vastes quantités de manière sombre, la seule échappatoire est d'en venir à modifier de manière ad hoc la gravité elle-même. C'est d'ailleurs ce que propose la théorie phénoménologique MOND (« MODified Newtonian Dynamics ») qui consiste à modifier la loi Newtonienne habituelle à l'échelle des galaxies (voir en particulier Milgrom 1983 [79]), la loi de la gravitation devenant alors en  $1/r$  dans la limite des accélérations faibles. Si le modèle de halo de matière noire décrit bien les courbes de rotation observées, la théorie MOND aussi, avec l'avantage qu'elle n'a qu'un paramètre libre dans l'ajustement des données. En dépit de ce succès, l'immense majorité des cosmologues a préféré accepter l'existence de matière sombre en grande quantité plutôt que d'abandonner un cadre conceptuel qui a fait ses preuves, au profit d'une modification ad hoc. Il pourrait s'avérer que MOND soit une théorie effective, limite de basse énergie d'une théorie plus fondamentale, mais cela nous laisse alors sans cadre théorique bien défini, en dépit de nombreuses tentatives (cf. [80] pour une revue récente). Il devient alors difficile de progresser, puisqu'il est dès lors impossible de vérifier si l'alternative est capable de reproduire les succès bien réels de la cosmologie relativiste, même s'ils sont obtenus au prix de révisions de nos idées sur ce qu'est le contenu « naturel » de l'Univers. En bref, même sans trop d'a priori, il semble que, pour aller de l'avant, le plus prometteur soit de chercher d'autres indications de la matière sombre, en ignorant cette tentation de MOND.

Un des résultats des simulations numériques effectuées dans le cadre d'études de la formation des structures est que les halos de matière sombre froide du scénario CDM (et notamment sa variante  $\Lambda$ CDM) devraient posséder un profil universel (NFW ou similaire, [81]), qui permet d'expliquer correctement les courbes de rotation à grande distance. En revanche, ce profil NFW décrit une distribution de densité piquée à courte distance,  $\rho \propto r^{-\alpha}$ , avec  $-1 > \alpha > -1,2$ . Or les courbes de rotation indiquent plutôt l'existence d'un cœur de densité pratiquement constante. Pour vérifier l'existence de ce cœur, de Block et al. [82] ont notamment pris garde d'analyser des galaxies à faible brillance de surface, pour lesquelles la contribution de la matière sombre domine celle de la matière visible, même dans le disque optique brillant. On peut alors considérer la matière visible comme un simple témoin dont la dynamique propre ne risque pas trop d'entacher le résultat de la comparaison. La prise en

compte de mouvements non circulaires, d'asymétries et décalages entre centres optiques et dynamiques ne suffit pas à réconcilier les prédictions CDM avec les observations. D'après eux, il faut un exposant  $\alpha = -0,2 \pm 0,2$  dans la partie interne du profil de densité de masse. Difficulté de l'analyse observationnelle, imprécisions des prédictions théoriques, ou indication précieuse d'insuffisance du modèle cosmologique actuel, la réponse n'est pas encore claire.

Il existe bien d'autres indications de la présence de matière sombre à l'échelle galactique, quand on utilise diverses contraintes tirées de la dynamique des systèmes examinés ; que ce soit en considérant la stabilité de disques froids qu'un halo stabilise, ou l'émission diffuse en rayons X des galaxies elliptiques qui confirme la profondeur du puits de potentiel, ou encore la distribution des débris de l'interaction entre galaxies elliptiques qui trahit les caractéristiques des halos associés. Les groupes compacts de galaxies offrent aussi un indice intéressant, car on constate qu'ils devraient rapidement fusionner et disparaître, sauf si les différents membres du groupe sont immergés dans un halo commun. Cependant ce sont les amas de galaxies qui vont nous permettre d'étudier la distribution de la matière à des échelles plus grandes.

Prenons le cas d'un amas particulier, Abell 2029, choisi pour l'absence de signe d'interaction avec d'autres concentrations de masse qui pourraient perturber la mesure de son profil. Le satellite Chandra, dont les détecteurs sont sensibles dans la gamme des rayons X, a permis d'obtenir le spectre d'émission entre 0,7 et 8 keV de 7 anneaux circulaires concentriques. Lewis, Buote et Stocke [83] en ont alors tiré la densité électronique, la densité et la température du gaz, ainsi que l'abondance du fer à sept distances du centre. Pour cela ils ont ajusté des modèles d'émission du gaz ionisé à haute température qui emplissent l'espace entre les galaxies. À partir de là ils ont pu déduire la distribution tridimensionnelle par une méthode de déprojections successives en supposant une structure en pelures d'oignon. Cette procédure leur fournit donc la densité de gaz,  $\rho_g(r)$ , et sa température,  $T_g(r)$ , entre 1 et 200 kpc.

La masse totale incluse dans un rayon donné,  $M_T(r)$ , peut alors être obtenue en faisant l'hypothèse d'un équilibre hydrostatique (et de la symétrie sphérique déjà supposée à l'étape de déprojection), qui impose la relation  $dP/dr = -\rho g$  entre la pression,  $P = nkT$ , la densité  $\rho = n\mu$  et l'accélération gravitationnelle locale,  $g = GM(< r)/r^2$ , ce qui donne

$$M_T(r) = -\frac{rkT_g}{G\mu m_p} \left[ \frac{d \ln \rho_g}{d \ln r} + \frac{d \ln T_g}{d \ln r} \right] \quad (7.34)$$

si on note  $k$  la constante de Boltzmann,  $G$  celle de la gravitation,  $\mu$  le poids atomique moyen du gaz (pris égal à 0,62), et  $m_p$  l'unité de masse atomique. Les points encadrés en noir de la figure 7.6a montrent le résultat obtenu. Ce profil est très bien décrit par le profil universel NFW (courbe pleine en bleu), tout comme d'ailleurs par une loi de puissance  $r^{-\alpha}$ ,  $\alpha = 1,2$  (tirets rouges). Contrairement à des études antérieures, le profil observé est donc en plein accord avec le profil NFW du scénario CDM, ce qui est sans doute dû au fait que Abell 2029 est exceptionnellement relaxé, sans signe aucun d'interaction

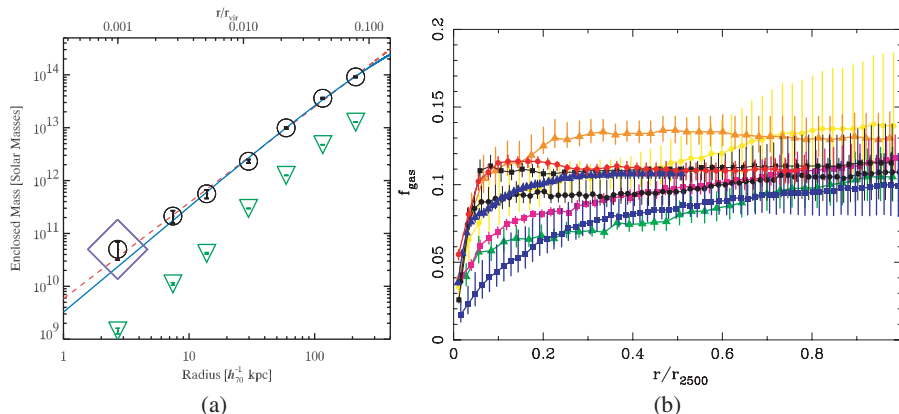


FIG. 7.6 – (a) Profil de masse totale déduite de l'émission X (points noirs) comparé à celui du gaz (points verts). (b) Fraction de la masse sous forme de gaz,  $f_{gas} = \rho_g/\rho_M$ , en fonction de la distance au centre dans 10 amas; elle tend vers une valeur universelle à grande distance. (D'après [83–85].)

récente, car son profil est partout extrêmement lisse. En fait la plupart des études récentes convergent vers un accord avec les prédictions NFW, que ce soit en utilisant Chandra ou sa contrepartie européenne, le satellite XMM-Newton.

La masse de gaz incluse dans  $r$ , qui est l'intégrale de  $\rho_g$ , est indiquée sur la figure par les points verts entourés de triangles. Elle s'élève de moins de 3 % au centre pour atteindre 14 % au dernier point. Plus systématiquement, la figure 7.6b montre la fraction de gaz  $f_{gaz} = M_{gaz}/M_T$  en fonction du rayon. Elle permet de constater qu'à suffisamment grande distance du centre, tous les amas ont la même fraction gazeuse (à  $r_{2500}$  qui est le rayon où la densité de l'amas est 2500 fois la densité moyenne à l'époque indiquée par le  $z$  de l'amas). La moyenne pondérée donne  $f_{gas} = (0,065 \pm 0,002) h^{-1,5}$  ce qui correspond<sup>16</sup> à  $f_{gas} = 0,108 \pm 0,003$  pour  $h = 0,71$ . On sait par ailleurs que la fraction de masse stellaire associée aux galaxies de l'amas est donnée par  $f_{\star gal} = M_{\star gal}/M_T = 0,010^{+0,005}_{-0,004} h^{-1}$  [77], ce qui correspond à  $f_{\star gal} = 0,014 \pm 0,007$  pour  $h = 0,71$ . Il n'existe pas d'autre forme de masse baryonique connue/détectée. On aboutit alors à

$$f_{baryon} = f_{gas} + f_{\star gal} = 0,065 \pm 0,002 h^{-1,5} + 0,010^{+0,005}_{-0,004} h^{-1} \quad (7.35)$$

et il nous faut admettre qu'on n'est capable d'identifier, sous forme baryonique, qu'environ 12 % de la masse totale détectée!

<sup>16</sup>Cette valeur est remarquablement en accord avec celle déterminée par Myers et al.,  $M_{SZ}/M_T = 0,061 \pm 0,011 h^{-1}$ , en utilisant l'effet Sunyaev-Zeldovich [86]. Cet effet induit une légère variation du flux du rayonnement fossile dans la direction de l'amas, qui permet d'en déduire la quantité de gaz chaud, si sa température est connue.

Les amas sont de grandes structures et cette fraction devrait être assez représentative de la fraction baryonique moyenne<sup>17</sup> du modèle d'univers, qui est  $\Omega_b/\Omega_m$  [87] ; si on admet la valeur  $\Omega_{b0} \simeq 0,04$  qu'indique la nucléosynthèse primordiale, on en déduit que  $\Omega_{m0} \sim 0,3$ . Nous verrons que cela est en plein accord avec le résultat d'autres méthodes. Il existe beaucoup d'autres analyses utilisant les observations en rayons X pour obtenir des contraintes sur la densité de matière, mais elles sont toutes de nature cosmogonique : elles infèrent quelles conditions doit satisfaire un modèle de formation des structures pour permettre de rendre compte du nombre d'amas observés en fonction de leur luminosité en X, ou de leur température, éventuellement en fonction de  $z$ . Et de ce point de vue, ce ne sont pas ces analyses-là qui sont les plus directes. La relativité nous offre cependant une autre voie, directissime, pour confirmer ce qu'indique l'émission X (et l'effet Sunyaev-Zeldovich).

Là encore, Fred Zwicky a vu loin. Il écrit en 1937 un article prémonitoire [88], suite à un article d'Einstein de l'année précédente. Après un calcul de l'effet de lentille gravitationnelle d'une étoile sur la ligne de visée d'une étoile d'arrière plan, Einstein conclut à la faible probabilité que cela se produise [89]. Zwicky fait en effet remarquer qu'une nébuleuse extragalactique serait une lentille bien plus prometteuse par sa plus grande concentration de masse, et aussi par sa taille qui permettrait de la résoudre jusqu'à de grandes distances. Il cite alors trois raisons pour lesquelles l'observation d'images de nébuleuses déformées par leur passage dans le champ gravitationnel d'une lentille serait intéressante. La première est qu'il s'agirait d'un test supplémentaire de la relativité. La deuxième est que cela permettrait de détecter et résoudre des objets à des distances inatteignables autrement. La troisième est que « *l'observation de la déflexion de la lumière autour de nébuleuses pourrait fournir la détermination la plus directe des masses de nébuleuses* », et résoudre le désaccord qu'il avait relevé quatre ans auparavant entre la masse des étoiles et la masse nécessaire pour que l'amas de Coma soit stable.

Depuis, la détection de l'effet de lentille gravitationnelle à l'échelle cosmologique est effectivement devenue une des manières convaincantes de tester la relativité à des échelles très supérieures à celle du système solaire. Sa deuxième « vision », le télescope à lentille gravitationnelle, est aujourd'hui un outil presque courant pour détecter et compter les objets les plus lointains ainsi que pour spectrographier différentes parties de galaxies extrêmement éloignées afin de contraindre leur dynamique (voir par exemple Soifer et al. 1998). De plus, nous utilisons les déformations systématiques d'objets d'arrière plan pour cartographier la distribution de matière sombre autour d'amas, et de grand relevés dédiés sont en cours pour appliquer cette technique à l'échelle des grandes structures de l'Univers.

---

<sup>17</sup>En fait les simulations numériques de formation des amas montrent que  $f_{baryon}$  n'est pas tout à fait représentatif de la valeur moyenne à cause de l'expulsion d'une fraction des baryons au cours de l'évolution. D'après Eke, Navarro et Frenk, on a plutôt  $f_{baryon} = (0,83 \pm 0,04) \Omega_b/\Omega_m$ , ce qui donne alors  $\Omega_{m0} = 0,27 \pm 0,03$ .



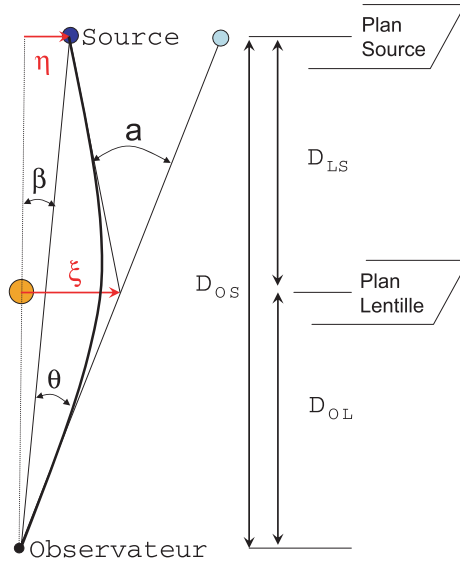


FIG. 7.7 – *Déflexion gravitationnelle.*

La relativité prédit un angle de déflexion par une masse ponctuelle  $M$ ,

$$a = \frac{4GM}{c^2\xi} \tag{7.36}$$

quand le paramètre d'impact, c'est-à-dire la distance de plus grande proximité rayon-défecteur, vaut  $\xi$ . Cette valeur est double de la valeur newtonienne obtenue en considérant la lumière comme un flot de particules massives. L'expédition montée en 1919 par Eddington pour profiter d'une éclipse solaire a vérifié que l'image d'une étoile est déplacée de 1,75 arcsec quand ses rayons lumineux sont en incidence rasante par rapport au soleil, et non de 0,87 arcsec. Dans les cas qui nous intéressent, les déviations sont toujours peu importantes (au plus une minute d'arc), ce qui correspond à des perturbations de la métrique toujours très faibles. L'effet d'une distribution de masse se calcule alors simplement par superposition de l'action de masses infinitésimales. Le calcul se simplifie considérablement en utilisant l'approximation eikonale : la courbure des rayons lumineux est très faible à l'échelle de la concentration de masse, ce qui permet d'ignorer le détail de la répartition de masse dans la direction  $r_3$  parallèle au rayon. Si l'on se repère par le vecteur  $\vec{\xi} = (\xi_1, \xi_2)$  dans le plan perpendiculaire, on peut définir la densité surfacique,  $\Sigma$ , du plan de la lentille,

$$\Sigma(\vec{\xi}) = \int dr_3 \rho(\xi_1, \xi_2, r_3) \tag{7.37}$$

et on trouve l'angle de déflexion en fonction du paramètre d'impact,

$$a(\vec{\xi}) = \frac{4G}{c^2} \int d^2\xi' \Sigma(\xi') \frac{\vec{\xi} - \vec{\xi}'}{|\xi - \xi'|^2} \quad (7.38)$$

expression qui généralise l'équation (7.36). Comme par ailleurs la déflexion a lieu sur une distance faible par rapport au déplacement le long de la ligne de visée, on peut représenter les choses comme si la déflexion avait lieu instantanément au passage du plan-lentille, c'est ce qu'on nomme l'approximation de lentille mince.

On voit d'après la figure 7.7 que

$$\vec{\eta} = \frac{D_{OS}}{D_{OL}} \vec{\xi} - D_{LS} a(\vec{\xi}), \quad \vec{\eta} = D_{OS} \vec{\beta}, \quad \text{et} \quad \vec{\xi} = D_{OL} \vec{\theta} \quad (7.39)$$

Toutes les distances sont des distances angulaires (définies par (7.18), et calculables grâce aux équations (7.32) et (7.29)). La position angulaire vraie,  $\vec{\beta}$  est donc reliée à la position apparente,  $\vec{\theta}$ , par

$$\vec{\beta} = \vec{\theta} - a \frac{D_{LS}}{D_{OS}} \equiv \vec{\theta} - \vec{\alpha} \quad (7.40)$$

Il s'avère commode de repérer les densités surfaciques par rapport à une densité surfacique critique,  $\Sigma_c$ , qui ne dépend que de la position relative des plans source et lentille,

$$\kappa = \frac{\Sigma(D_{OL}\vec{\theta})}{\Sigma_c}, \quad \text{avec} \quad \Sigma_c = \frac{c^2}{4\pi G} \frac{D_{OS}}{D_{OL}D_{LS}} \quad (7.41)$$

On peut alors montrer que le champ de déflexion  $\vec{\alpha}$  dérive d'un potentiel bidimensionnel  $\Psi$  obéissant à une équation de Poisson où  $\kappa$  est le terme source :

$$\vec{\alpha} = \nabla\Psi, \quad \nabla^2\Psi = 2\kappa(\theta) \quad (7.42)$$

Ce champ  $\vec{\alpha}$  est l'analogue exact du champ d'accélération gravitationnel newtonien tridimensionnel (qui est lié à la densité de masse par une même équation de Poisson,  $\vec{g} = \nabla\phi$ ,  $\nabla^2\phi = \nabla\vec{g} \propto \rho$ ).

On peut distinguer trois régimes de déformations : le régime des déflexions gravitationnelles fortes, ce qui rend possible l'existence de plusieurs chemins entre une source et l'observateur. Il y a alors des images multiples d'un même objet, certaines pouvant parfois être très déformées. Dans le cas limite d'une source exactement dans l'axe d'une lentille parfaitement symétrique, son image peut prendre une forme annulaire appelé anneau d'Einstein. Si ces conditions draconiennes ne sont pas tout à fait réalisées, l'anneau est brisé et la source peut alors avoir pour image deux grands arcs. Le deuxième régime est celui des petits arcs, où les images d'objets d'arrière-plan sont significativement distordues, allongées mais uniques. La figure 7.8 offre de nombreux

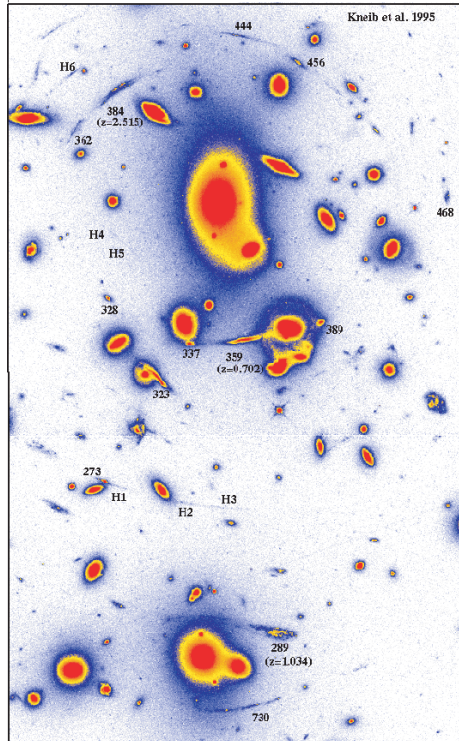


FIG. 7.8 – Image de l’amas Abell 2218, prise par le télescope spatial Hubble (après réparation), où l’on peut distinguer facilement de nombreux arcs et arceaux. (D’après [90].)

exemples de grands et petits arcs. Le dernier régime est celui des déflexions faibles, si faibles que l’image des objets n’est qu’imperceptiblement modifiée et qu’il faut retrouver le signal de manière statistique. Le premier régime est pertinent quand il existe des directions pour lesquelles la densité surfacique excède la densité critique, i.e. quand  $\kappa > 1$ . Sinon, les images sont déplacées et déformées mais restent uniques.

L’astronomie gravitationnelle est longtemps restée théorique. Ce n’est qu’en 1979 qu’a été découverte la première source à images multiples, le quasar 0957+561 [91]. Les premiers grands arcs seront quant à eux découverts en 1986 et 1987 [92, 93]. L’interprétation en terme d’effet de lentille gravitationnelle a été unanimement acceptée dès 1988, quand le groupe toulousain de Bernard Fort a démontré que l’arc se trouve à  $z = 0,724$ . Beaucoup plus loin, donc, que les galaxies de l’amas qui se trouvent à  $z = 0,37$ . Le premier anneau d’Einstein autour d’une galaxie a été découvert la même année [94]. L’intérêt des fortes déflexions est qu’elle permettent des reconstructions simples et robustes. Ainsi, dans le cas idéal d’un anneau d’Einstein, la mesure de son rayon

$R_E = \sqrt{4GM/c^2 D_{LS}/(D_{OS}D_{OL})}$  permet de déduire directement la masse en fonction des seules distances angulaires.

Cette sorte de tomographie par inversion de la déformation d'images utilisant grands et petits arcs a fait l'objet de très nombreux travaux. La démonstration la plus spectaculaire est sans doute la reconstruction de la distribution de masse dans l'amas CL 0024+1654 que Tyson a effectuée en 1998 avec ses collaborateurs [95]. Dans ce cas particulièrement propice, une fois soustraite l'émission des galaxies de l'amas, il y a huit sous-images bien résolues de la même galaxie d'arrière-plan, qui est aisément reconnaissable par sa couleur et sa morphologie. On peut alors procéder à une inversion paramétrique du modèle qu'on se donne, qui a ici 512 paramètres. Après deux millions d'essais répartis sur 12 mois de calculs, les auteurs ont abouti à l'image de la figure 7.9a, où les concentrations de masse associées aux galaxies sont indiquées par du bleu, et où le rouge montre la répartition de la masse sombre. Cette dernière est à 98 % décrite par deux larges distributions diffuses, toutes deux centrées près de la galaxie la plus brillante de l'image. Les contours délimitent les régions de densité surfacique  $\Sigma$  supérieure à 0,5, 1 (plus épais) et 1,5 fois la densité critique. La taille de l'image est de  $336 h^{-1}$  kpc. La figure 7.9b montre les profils radiaux correspondants, la ligne mince indiquant la masse associée aux galaxies, et la ligne épaisse la masse totale. La figure donne aussi le profil d'intensité lumineuse visible (« V light »), associée aux galaxies (tirets), et totale (courbe pleine). Ceci permet d'une part de visualiser que le rapport masse sur luminosité visuelle,  $M/L_V$ , des galaxies augmente avec le rayon (matière sombre des halos galactiques), et d'autre part que l'essentiel de la masse est dans un halo sombre commun, qui représente 83 % de la masse totale dans  $107 h^{-1}$  kpc,  $M_{107} = 1,662 \pm 0,002 \times 10^{14} h^{-1} d_{0,57}^{-1} M_\odot$  (le rapport des distances angulaires  $d_{0,57} = D_{LS}/D_{OS}/0,57 = 1 \pm 0,15$  rend compte de l'incertitude sur le  $z$  de la source). Ce halo est bien décrit par un profil NFW à grande distance du centre (pointillés), mais il existe un cœur à  $35 h^{-1}$  kpc que ne prédit pas les simulations actuelles (et que nous n'avions pas davantage vu en X dans le cas très relaxé présenté). L'absence de barres d'erreur empêche toutefois d'évaluer la portée du désaccord à petite échelle. Quoi qu'il en soit, l'étude confirme indubitablement la grande quantité de matière sombre non associée aux galaxies (la quantité détectée correspond à un  $M/L_V$  qui ramené à aujourd'hui, vaut  $M/L_V = 390 h d_{0,57}^{-1} (M/L_V)_\odot$ ).

Les premières tentatives d'utilisations de l'effet de lentille faible ont très rapidement suivi la première détection d'images multiples, avant même qu'aient été détectés les premiers grands arcs. Il est vrai cependant que ces travaux pionniers de 1983 par Tony Tyson et collaborateurs n'ont alors abouti qu'à des limites supérieures<sup>18</sup> (la détection du champ de cisaillement cosmique

<sup>18</sup>Valdes, Jarvis et Tyson [96] ont limité l'inhomogénéité de la distribution de masse à grande échelle par l'absence d'ellipticité cohérente détectable dans chacun des 35 champs où résidaient 44 462 galaxies choisies au hasard. La dispersion (de champ à champ) de l'ellipticité étant inférieure à 0,03, cela a permis aux auteurs de fixer une limite supérieure (à  $3\sigma$ ) de 0,1 pour le contraste de densité de masse sur une échelle du gigaparsec.

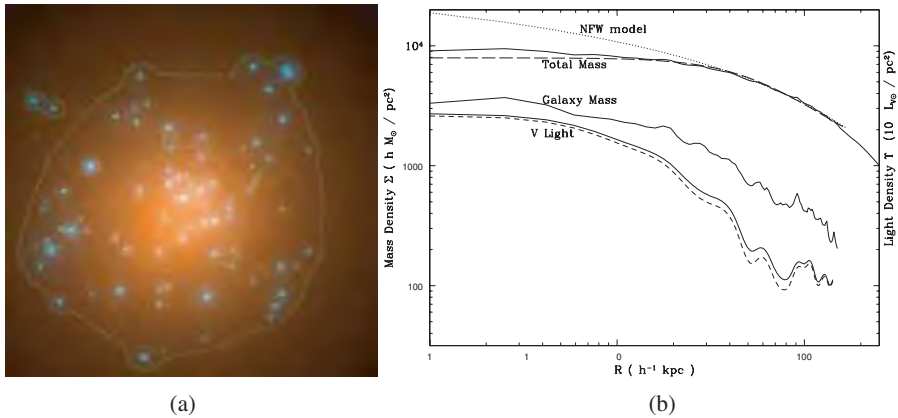


FIG. 7.9 – Analyse de la distribution de masse dans l’amas CL 0024+1654. (a) Distribution déduite par inversion paramétrique. En orange, la masse sombre, à 98 % dans une composante extrêmement lisse; en bleu, la masse associée aux galaxies. (b) Profils radiaux de masse. (D’après [95].)

n’aura en fait lieu qu’en 2000!). L’année suivante, Tyson et ses collaborateurs proposent de mesurer statistiquement la distortion moyenne de 46 954 galaxies d’arrière-plan pour contraindre la masse de lentilles d’avant-plan (en co-additionnant de nombreuses images remises à l’échelle et centrées sur chacune des lentilles [97]). Ils concluent à juste titre que la masse associée aux galaxies individuelles n’est pas suffisante pour atteindre  $\Omega_{m0} = 1$ . C’est là encore un résultat négatif, basé sur une absence de détection. Mais six ans plus tard, en 1990, Tyson, Valdes et Wenk démontrent que les galaxies derrière des amas riches ont une distortion mesurable [98].

Il devient dès lors possible de cartographier le champ de déformation gravitationnelle. La figure 7.10a montre ce champ à la périphérie de l’amas CL 0024+1654 (dont l’analyse de la partie centrale vient d’être présentée). Cette image a été obtenue en 1994 par Bonnet, Mellier et Fort [99] en moyennant dans des macro-pixels l’ellipticité mesurée des galaxies d’arrière-plan. Le bâtonnet (ou pseudo-vecteur) dans chaque macro-pixel offre ainsi une visualisation directe de la direction et de l’intensité de la déformation. Cette moyenne statistique par pixel revient à moyenner le champ aléatoire fort mais incohérent dû à l’ellipticité propre des galaxies; elle fait apparaître le champ de cisaillement gravitationnel créé par la lentille. Ce sont Nick Kaiser et Gordon Squire [100] qui posent en 1993 les bases algorithmiques rigoureuses de l’utilisation de ce champ de cisaillement (en fait de la partie sans trace).

Abell 2390 est l’un des amas les mieux étudiés et il permet des comparaisons de méthodes. La figure 7.10b montre ainsi le profil obtenu en 1996 par

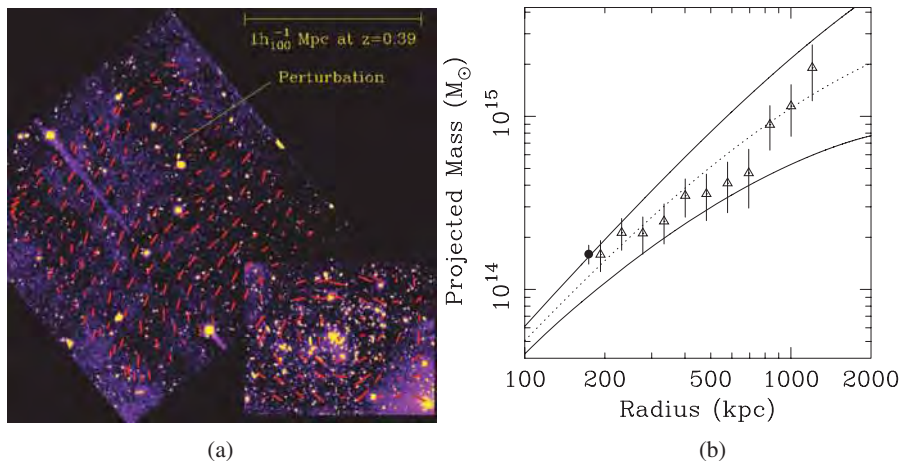


FIG. 7.10 – (a) Le champ de cisaillement à la périphérie de l’amas CL 0024+1654. (b) Comparaison des profils radiaux dans Abell 2390 tirés des analyses en X (pointillés et  $\pm 1\sigma$  en lignes pleines) et des effets de lentille fort (cercle plein noir) et faible (triangles évidés). (D’après [84, 99].)

Squires et al. [101] par analyse du champ de cisaillement (triangles évidés), la valeur déduite de la présence d’un grand arc par Pierre et al. la même année (cercle plein, [102]), ainsi que le profil déduit d’observations en X avec Chandra par Allen, Ettori et Fabian (pointillés et valeurs à  $\pm 1\sigma$  en lignes pleines [84]). Ce diagramme nous permet ainsi de vérifier directement un aspect de la relativité à l’échelle des amas ( $\sim 1$  Mpc) tout en validant l’hypothèse d’équilibre hydrostatique utilisée pour mettre en œuvre la physique usuelle. Il existe depuis peu des travaux équivalents dans de multiples amas. Cet accord quantitatif entre méthodes est relativement récent, grâce à des progrès dans les méthodes d’analyse (comme l’exclusion de secteurs de l’amas qui présente des signes d’interaction avec une autre concentration de masse), ou la sélection d’amas particulièrement relaxés, ou encore grâce à l’abandon d’hypothèses de symétrie trop fortes. La présence de vastes quantités de matière sombre est donc bien établie même si, comme nous l’avons vu, des divergences demeurent, à petite échelle, concernant l’existence ou non d’un cœur.

Les analyses présentées montrent que la quantité de matière sombre détectée augmente avec l’échelle. Jusqu’où se poursuit ce mouvement ? Un modèle simple Einstein-de Sitter ( $\Omega_m = 1$ ) offre-t-il une bonne description de l’Univers aujourd’hui ? Ces questions ont été ardemment débattues, en particulier entre 1975 et 1995, entre le moment où l’existence de matière sombre a été généralement acceptée jusqu’à l’acceptation collective de la présence d’énergie sombre.

### 7.3.4 L'obscur énergie du vide contemporain

Un des objectifs essentiels de la cosmographie est de recenser les diverses composantes de l'Univers et, en particulier, de déterminer le bilan énergétique des diverses contributions en fonction de leur équation d'état. Nous avons vu que le rayonnement ( $w = 1/3$ ) a une contribution extrêmement faible aujourd'hui,  $\Omega_{r,0} \sim 10^{-5}$ , alors que pour la matière (de pression négligeable,  $w = 0$ ), on a  $\Omega_{m0} \sim 0,3$ . Est-ce tout ? En près d'un siècle, de très nombreux tests ont été inventés et mis en œuvre en supposant que diverses catégories d'objets astrophysiques sont des marqueurs aux propriétés invariables, ou calculables, ou à tout le moins calibrables observationnellement. Si c'est le cas, on peut utiliser ces propriétés pour mesurer l'évolution de diverses quantités comme la distance ou le volume. Il existe ainsi une immense littérature sur les comptages d'objets en fonction de leur flux, car la mise en œuvre en est particulièrement simple. Ce test a d'ailleurs permis au début du vingtième siècle de démontrer que notre Galaxie a un bord, par comptage d'étoiles en fonction de leur flux apparent. En revanche, les très nombreuses études de comptages de galaxies n'ont pas permis de contraindre sans ambiguïté les paramètres cosmologiques qui déterminent la variation avec l'éloignement du flux et du volume sous-jacent. La raison en est simple, il faut accéder à de très grandes distances pour que les modélisations auxquelles on compare les quantités observables varient notablement selon la valeur des paramètres cosmologiques adoptés. Ces distances cosmologiques correspondent à l'exploration d'une plage de temps comparable à l'âge de l'Univers lui-même. La difficulté consiste à comparer des objets tels qu'ils étaient au début de l'Univers avec ceux d'aujourd'hui. Ainsi à  $z = 1$  l'Univers a selon les modèles entre un tiers et la moitié<sup>19</sup> de son âge actuel (cf. Fig. 7.4). Dans le cas des galaxies, les effets d'évolution sont trop complexes et trop importants pour espérer obtenir des contraintes solides sur les paramètres cosmologiques. Dans la même veine, rappelons-nous que la probabilité qu'un quasar ait plusieurs images par effet de lentille sur une galaxie d'avant-plan est d'autant plus grande que le chemin parcouru par la lumière est plus grand (ce chemin augmente beaucoup avec  $\Omega_{\Lambda 0}$  à  $H_0$  fixe). On trouve ainsi  $\Omega_{\Lambda 0} \lesssim 0,75$ . C'est tout à fait raisonnable, mais il est néanmoins bien difficile d'être certain qu'on a évalué suffisamment bien l'abondance des sources et l'effet des lentilles galactiques pour que les contraintes posées soient vraiment solides. . . Les tests de taille angulaire menés jusqu'ici ne se sont pas, non plus, montrés très concluants [103].

Un des tests les plus simples, au moins en principe, consiste à mesurer la diminution de flux lumineux en fonction du  $z$  de « chandelles standards » dont la luminosité intrinsèque est invariable. Ce test a été récemment à l'origine de progrès remarquables et nous allons l'examiner de plus près. La distance

---

<sup>19</sup>Et les objets les plus lointains de la figure 7.1, dont on a la spectroscopie, sont à  $z \sim 6,5$  quand l'Univers avait moins de 10 % de son âge actuel.

de luminosité a une forme analytique un peu complexe<sup>20</sup> qu'on doit en général calculer numériquement, mais un développement limité à petit  $z$  permet d'éclairer ce que sonde ce test :

$$d_L = \frac{c}{H_0} \left[ z + \frac{1 - q_0}{2} z^2 + \mathcal{O}(z^3) \right] \quad (7.44)$$

Si l'on se restreint au premier terme, approximation certainement suffisante à  $z \lesssim 0,1$ , on retrouve simplement la loi de Hubble. À plus grand  $z$ , le premier terme correctif au taux d'expansion fait naturellement apparaître le paramètre de décélération  $q_0 \equiv -(\ddot{a}/\dot{a}^2)_0$  qui est donné par  $q_0 = 1/2 \sum_i \Omega_i(1 + 3w_i)$ . Et dès que  $z \sim 1$ , on est sensible aux termes d'ordre supérieur comme la variation temporelle récente de l'accélération.

Le flux est repéré par les astronomes par des magnitudes, définies telles qu'un facteur 100 en flux soit réparti en 5 intervalles logarithmiques,  $m_2 - m_1 = -2,5 \log(F_2/F_1)$ ; cette convention un peu bizarre vient du fait qu'au deuxième siècle avant J.-C., Hipparcos divisa les étoiles en 6 classes, les plus brillantes étant de première magnitude. Il se trouve que l'œil a une réponse logarithmique et qu'il est capable de distinguer une étoile environ 100 fois moins brillante que l'étoile la plus brillante du ciel! Dans le cas d'une chandelle standard vue à deux distances différentes, le flux apparent reçu variant par définition comme l'inverse du carré de la distance de luminosité, on a  $m_2 - m_1 = 5 \log(d_{L2}/d_{L1})$ . Par ailleurs, on peut définir la magnitude absolue,  $M$ , comme la magnitude apparente qu'aurait un objet placé à une distance de 10 parsecs. (Pour fixer les idées, on note que le soleil a une magnitude absolue visuelle<sup>21</sup>  $M(V)_\odot = 4,83$  dans la bande autour de  $\lambda_V = 550$  nm.) La magnitude apparente, satisfait donc  $m - M = 5 \log(d_L/10pc) = 5 \log d_L/1 \text{ Gpc} + 25$ .

Une mesure de ce qu'on a coutume d'appeler le module de distance  $\mu \equiv m - M$  pour une collection d'objets à différents  $z$  (ayant tous la même magnitude absolue) est donc un échantillonnage de la courbe  $d_L(z)$ . On trace ainsi un « diagramme de Hubble »,  $\mu(z)$ , qui permet d'abord de vérifier la loi de Hubble à  $z \lesssim 0,1$ , puis de jauger l'accélération (positive ou négative) de l'Univers à plus grand  $z$ ... Il ne reste plus qu'à identifier ce Graal astronomique, une véritable chandelle standard. Historiquement, beaucoup de chandelles ont été essayées. Nous avons déjà rencontré les céphéides, dont la

---

<sup>20</sup>La distance de luminosité se calcule à partir de sa définition,  $d_L = a_0 S_k(\chi)(1+z)$  et de celle de la distance radiale comobile,  $a_0 d\chi = cdz/H(z)$ , grâce à l'équation de Friedman qui relie le taux d'expansion au contenu de l'Univers (cf. Eqs. (7.17), (7.32)), (7.29) et (7.27) du paragraphe 7.3.1), ce qui donne

$$d_L(z) = a_0(1+z)S_k \left( \frac{c}{a_0 H_0} \int_0^z dz' \left[ \sum_i \Omega_{i0}(1+z')^{3(1+w_i)} + \Omega_{k0}(1+z')^2 \right]^{-1/2} \right) \quad (7.43)$$

où  $a_0$  se déduit de la définition  $\Omega_{k0} \equiv -k/(a_0 H_0)^2$  (et  $\Omega_{k0} = 1 - \sum_i \Omega_{i0}$ ).

<sup>21</sup>La bande bleue autour de  $\lambda_B = 445$  nm est aussi souvent utilisée, et dans ce cas  $M(B)_\odot = 5,48$ .



connaissance de la luminosité en fonction de la période a permis de reconnaître que la nébuleuse d'Andromède est un « Univers-île » différent de la Voie Lactée. Rappelons-nous aussi que les estimations initiales de Hubble en 1926 étaient fortement biaisées (d'un facteur  $\sim 3$ ), car on ignorait à l'époque l'existence de *deux* classes distinctes de céphéides (ayant chacune une magnitude absolue différente pour la même période de pulsation...). Cette erreur est restée inaperçue jusqu'à la mise en service en 1953 du télescope de 5 mètres (200-inch) du Mont Palomar !

Les astronomes prennent donc grand soin d'isoler au mieux des classes homogènes en étudiant en détail images puis spectres ; les classifications se raffinent au fur et à mesure de la collecte d'échantillons plus importants et de spectres plus résolus (en fréquence), sur une plus large plage. De nombreuses études se sont intéressées aux supernovae dont le potentiel pour la cosmologie a été reconnu dès les années trente. Ces explosions d'étoile peuvent en effet être aussi brillantes qu'une galaxie entière pendant plusieurs semaines et leur luminosité maximum ne semble pas varier. Ainsi Kowal [104] indique en 1968 une dispersion de la luminosité au maximum  $\sigma = 0,6$  mag pour les supernovae de type I, les SNe-I, dont le spectre ne présente pas une certaine raie de l'hydrogène. Au cours des années quatre-vingt ces supernovae déficientes en hydrogène ont été subdivisées à leur tour en sous-types a, b et c. Il s'est alors avéré que les SNe Ia forment une classe exceptionnellement homogène. Dans une revue de 1992, Branch et Tamman [105] concluent que ce sont probablement les meilleures chandelles standard connues, avec une dispersion au maximum inférieure à un quart de magnitude, soit 25 % en flux (aussi bien dans la bande bleue que dans la bande visuelle). La constitution d'un échantillon homogène de 20 supernovae proches ( $z \lesssim 0,1$ ), très bien mesurées, a par la suite permis de vérifier que la séquence temporelle de spectres au cours de l'explosion est effectivement très similaire d'une SNe Ia à une autre (pour la plupart des supernovae). Cependant, cet échantillon a aussi permis d'établir une corrélation entre la rapidité du déclin de la luminosité et la luminosité absolue au maximum. De multiples méthodes ont alors été développées pour réduire la dispersion, en mettant à profit ces corrélations empiriques pour, en quelque sorte, mettre les SNe Ia au même gabarit, tout en éliminant les moins standards des standards. La figure 7.11a permet de juger la réduction de la dispersion des courbes de lumière qu'on peut atteindre par une des méthodes proposées. La figure 7.11b illustre la réduction de dispersion résultante pour un diagramme de Hubble avec une méthode équivalente. Dans ce dernier cas, la dispersion des  $\mu$  de cet échantillon étendu de 50 SNe Ia a été réduite de 0,44 à 0,15 mag (un facteur 3 en log !) pour atteindre une précision de  $\sim 7$  % en distance. La ligne pleine permet de vérifier la très grande linéarité de la relation entre  $d_L$  et  $cz \simeq v$  (la ligne est de pente 5 car  $\mu = 5 \log d_L + 25$ ), et établit la « loi de Hubble » au pour cent près, aussi loin qu'on peut raisonnablement vouloir l'appliquer ( $z \lesssim 0,1$ ). La valeur numérique de la constante de Hubble,  $H_0$ , doit cependant faire appel à un calibrateur extérieur, en général les céphéides.

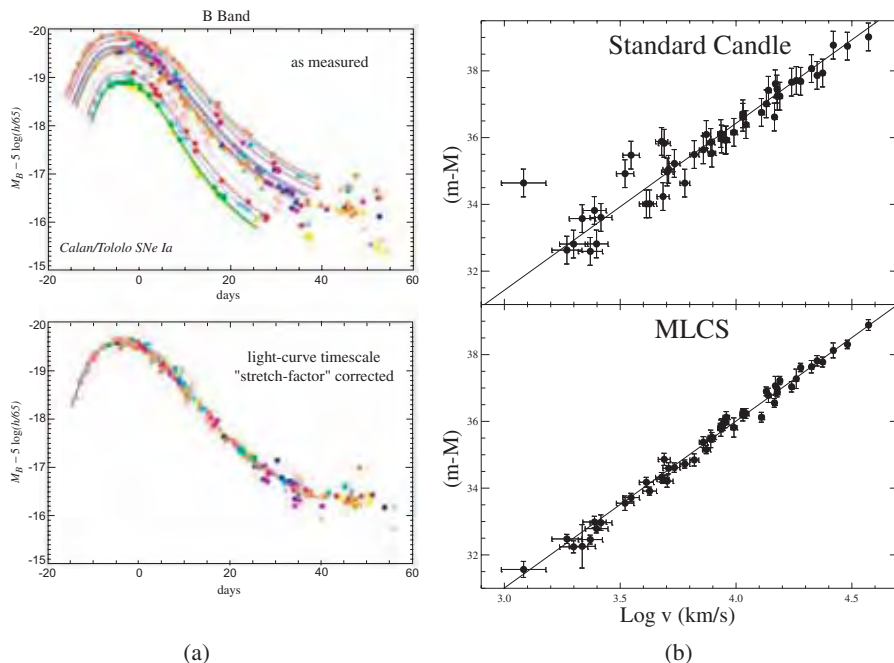


FIG. 7.11 – Mise au gabarit de Supernovae. Les figures du bas correspondent à la correction de celles du haut. (a) Effet sur les courbes de lumière qui donnent la décroissance de l'éclat en fonction du temps. (b) Diminution de la dispersion dans le diagramme de Hubble qui montre la variation du module de distance en fonction de la vitesse de récession de la galaxie hôte. (D'après Kim et al., 1997.)

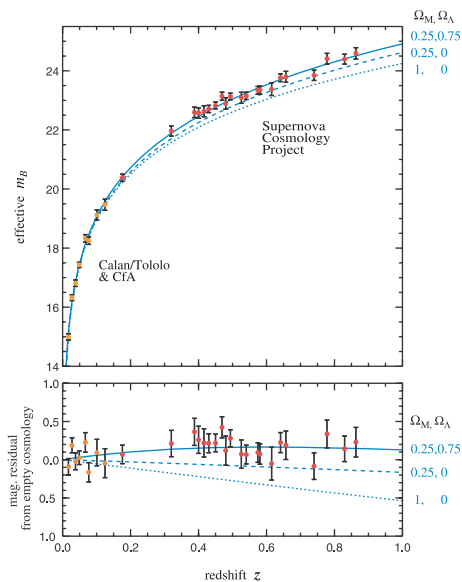
Pour assembler un échantillon de SNe Ia à grand  $z$ , une des difficultés est la nécessité de les détecter tôt, avant le maximum de la courbe de lumière, pour bien en connaître la forme et les mettre au gabarit. Il faut aussi en avoir un spectre précis, qui ne peut s'obtenir que sur les plus grands télescopes, peu après le maximum. Or le temps de montée de la courbe de lumière est relativement court, typiquement trois semaines, et les SNe Ia sont des événements aléatoires relativement rares ; il s'en produit quelques unes par siècle parmi les milliards d'étoiles d'une galaxie comme la nôtre. Comme dans le cas de la recherche d'événements de micro-lentilles pour détecter les MACHOs, la technologie n'est devenue disponible que récemment. Mais on peut aujourd'hui faire en une nuit l'image d'un million de galaxies à  $z \lesssim 0,5$ , sur un télescope de quatre mètres de diamètre équipé d'une caméra moderne à relativement grand champ (détecteurs CCD sur un quart de degré carré). Une ré-observation de suivi quelques semaines plus tard des mêmes champs permet, par différence, de repérer les objets ayant changé de magnitude. Si on est capable de traiter les données rapidement (typiquement 20 giga-octets par

nuit), on peut être sûr d'avoir identifié des candidats, avant qu'ils aient atteint le maximum de leur courbe de lumière, pour les suivre spectroscopiquement sur un très grand télescope... et pour convaincre le comité d'attribution de ce précieux temps que les quelques nuits allouées seront utilisées à bon escient, typiquement six mois plus tard, sur des candidats dont on n'a pas encore la liste. En fait, maintenant que les télescopes de 8 mètres sont moins rares et que la technique a fait ses preuves, on commence à dédier partiellement certains télescopes de 4 m à des tâches de ce type. Ceci permet de faire le relevé de mêmes champs, dans plusieurs couleurs, plusieurs fois par mois, pendant quelques mois d'affilée et d'assurer un flot soutenu de candidats.

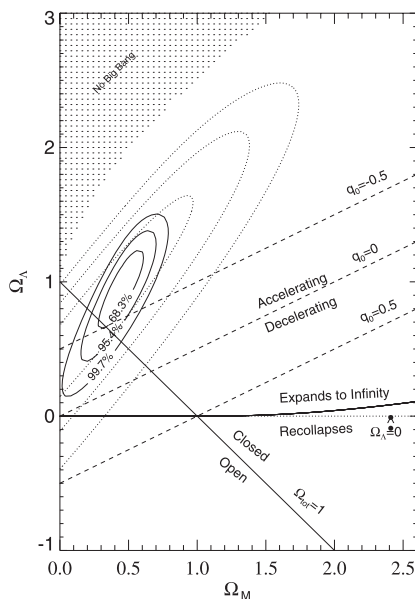
Deux équipes délocalisées, le « Supernovae Cosmology Project » (SCP) et le « High Z supernovae Team » (HZT), ont progressivement émergé dans la course aux SNe Ia à grand  $z$ . Les échantillons qu'elles ont assemblés ont permis de vérifier que les courbes de lumière des supernovae à grand  $z$  sont bien dilatées du facteur  $1+z$  attendu dans les modèles de Friedman-Lemaître. Mais les implications cosmologiques ont initialement beaucoup varié entre 1995 et 1998. Ces premières analyses ont indiqué tantôt  $\Omega_{m0} \sim 0,3$ , tantôt  $\Omega_{m0} \sim 1$ , ce qui s'explique aisément, vu la difficulté de la tâche et le fait que ces premières conclusions étaient basées sur une poignée d'événements à la spectroscopie d'assez mauvaise qualité. Puis l'équipe HZT a publié en 1998 la détection d'une valeur non nulle<sup>22</sup> de  $\Omega_{\Lambda 0}$  sur la base de 16 SNe Ia à grand  $z$  [63]. Peu de temps après l'autre équipe, SCP, a confirmé ces premiers résultats sur la base de 42 SNe Ia à grand  $z$  [64]. La figure 7.12a montre la version 2003 de leur diagramme de Hubble, nettement améliorée par rapport à la version originale de 1999, grâce à l'inclusion de 11 SNe Ia supplémentaires pour lesquelles on dispose de données de très grande qualité, qui ont été obtenues avec le télescope spatial Hubble (HST). Les courbes bleues du diagramme permettent de comparer les données aux modules de distance attendus selon différentes combinaisons de paramètres ( $\Omega_{m0}, \Omega_{\Lambda 0}$ ). On constate ainsi que les SNe Ia à  $z \simeq 0,5$  sont plus faibles d'environ *un tiers de magnitude* par rapport à un univers avec  $\Omega_{m0} = 0,25, \Omega_{\Lambda 0} = 0$ , différence qui disparaît si  $\Omega_{\Lambda 0} = 0,7$ .

Ce que mesurent ces données est une certaine combinaison des paramètres qui contrôle la variation de  $d_L$  dans ce domaine de  $z$ . L'expression (7.44) montre que la comparaison de données à  $z < 0,1$  et  $z \sim 0,5$  devrait être principalement sensible au paramètre de décélération  $q_0$  (indépendamment de  $H_0$ ), avec  $q_0 = \Omega_{m0}/2 - \Omega_{\Lambda 0}$ . L'inclusion de données à plus grand  $z$  devrait en revanche permettre de progressivement séparer les effets de  $\Omega_{m0}$  et  $\Omega_{\Lambda 0}$ . En pratique, on calcule la magnitude  $m_i^M$  qu'aurait chaque supernovae pour une valeur donnée des paramètres inconnus utilisée dans le modèle des observations. Puis on calcule la probabilité d'obtenir dans un tel modèle l'ensemble des magnitudes observées,  $m_i^O$ , étant données les incertitudes que l'on a estimées sur chaque magnitude,  $\sigma_i$ . On peut alors indiquer dans quelles régions de l'espace des paramètres cette probabilité est supérieure à

<sup>22</sup> $\Omega_{m0} = 0,24 \pm 0,1$  sous l'hypothèse  $\Omega_{k0} = 0$  ( $= \Omega_{m0} + \Omega_{\Lambda 0} - 1$ ), et une valeur non physique de  $\Omega_{m0} = -0,35 \pm 0,18$  si  $\Omega_{\Lambda 0} = 0$ .



(a)



(b)

FIG. 7.12 – (a) Diagramme de Hubble, dans lequel toutes les SNe Ia de  $\delta z < 0,01$  ont été combinées par une moyenne pondérée en un point unique. La partie basse montre l'écart à une cosmologie vide ( $\Omega_{m0} = \Omega_{\Lambda 0} = 0$ ), en expansion balistique. (b) Contraintes sur le couple de paramètre ( $\Omega_{m0}, \Omega_{\Lambda 0}$ ). (D'après [71, 72].)

un certain seuil<sup>23</sup>. La figure 7.12b montre les contours correspondant à divers seuils choisis par convention, afin que dans 68, 95 et 99 % des cas (par exemple en simulant diverses observations du modèle avec ses incertitudes et son analyse) les vraies valeurs des paramètres soient incluses dans la région délimitée par ces contours concentriques qui délimitent ainsi des intervalles de confiance. La figure montre à la fois les premières contraintes obtenues en 1998 par le HZT (contours pointillés), et les dernières analyses disponibles dans une publication de 2004 de la même équipe (contours plein). Dans ce cas, l'amélioration des contraintes tient à l'obtention par cette équipe de 16 nouvelles SNe Ia mesurées avec le HST dont 6 sont à  $z > 1,25$  ! Cela conduit à une réduction d'un facteur six de la surface comprise dans le contour à 68 %.

La figure 7.12b indique aussi (en tirets) des lignes d'accélération constante. Si les données ne mesuraient que l'accélération (i.e.  $z \lesssim 0,5$  d'après (7.44)), on aurait alors vu des ellipses extrêmement allongées, de grand axe aligné avec ces lignes. Les données de 1998 permettaient déjà de lever partiellement la dégénérescence entre divers couples de valeurs  $(\Omega_{m0}, \Omega_{\Lambda0})$  qui maintiennent  $q_0$  constant. Elles excluaient clairement le modèle Einstein-de Sitter ( $\Omega_{m0} = 1, \Omega_{\Lambda0} = 0$ ), ainsi que tout modèle en décélération ( $q_0 > 0$ ). Les données 2004 conduisent à des contours moins allongés, et plus resserrés, autour du modèle de concordance  $\Omega_{m0} \simeq 0,3, \Omega_{\Lambda0} = 1 - \Omega_{m0}$ . Nous verrons bientôt que les analyses cosmogoniques des anisotropies du rayonnement fossile indiquent très clairement que  $\Omega_{k0} \simeq 0$ , c'est-à-dire un univers plat, ce qu'indique la ligne  $\Omega_{tot} = \Omega_{m0} + \Omega_{\Lambda0} = 1$  du diagramme. Si maintenant on suppose dans l'analyse qu'effectivement  $\Omega_{k0} = 0$ , on obtient  $\Omega_{m0} = 1 - \Omega_{\Lambda0} = 0,29 \pm_{0,03}^{0,05}$  (en complet accord avec les résultats de l'autre équipe). Ce jeu de données permet de plus une première détection sans ambiguïté de la période de décélération qui a précédé la phase actuelle d'accélération<sup>24</sup>. Comme je l'ai noté précédemment, le modèle  $\Lambda$ CDM était déjà considéré très sérieusement depuis environ 1995, mais il ne fait cependant pas de doute que ces analyses de SNe Ia ont joué un rôle très important (avec les analyses du rayonnement fossile) dans le changement de paradigme qui a progressivement eu lieu entre 1995 et 2000.

Un tel modèle d'univers est néanmoins étrange, en raison notamment de la coïncidence qu'il implique, puisque nous vivons à l'époque où l'expansion change de signe. Diverses tentatives théoriques pour rendre cette coïncidence moins troublante aboutissent à modifier l'équation d'état du vide, qui pourrait

---

<sup>23</sup>La probabilité d'un modèle est  $P \propto \exp -\chi^2/2$ , avec  $\chi^2 = \sum_i (m_i^O - m_i^M)^2/\sigma_i^2$ . Si certaines quantités du modèle de la mesure – certaines dimensions de l'espace des paramètres – ne nous intéressent pas (comme la magnitude absolue des SNe Ia), il suffit de prendre l'intégrale de cette (densité de) probabilité par rapport à ces variables, ce qui assure que l'incertitude qui y est attachée est bien propagée dans le résultat final. Cette intégration sur les dimensions du modèle qui ne nous intéressent pas est appelée marginalisation. La probabilité calculée est la probabilité d'obtenir les données observées dans le modèle considéré. D'après le théorème de Bayes, c'est aussi la probabilité de ce modèle et de ses paramètres, au vu des données observées, dans la mesure où tous les modèles étaient a priori équiprobables.

<sup>24</sup>Dans un modèle phénoménologique où  $q(z) = q_0 + z(dq/dz)_0$ , les auteurs trouvent un changement de signe à  $z_{\Lambda} = 0,46 \pm 0,13$ , à comparer au  $z_{\Lambda} = 0,76$  du modèle  $\Lambda$ CDM complet qui est indiqué sur la figure 7.1.

même varier au cours de l'évolution de l'Univers. Introduisant  $w(z) = w_0 + w_1 z$  pour paramétrer la variation de  $\Omega_\Lambda = \Omega_{\Lambda 0}(1+z)^{3(1+w)}$  qui s'ensuivrait, on peut alors écrire, en négligeant rayonnement et courbure

$$d_L = \frac{c}{H_0}(1+z) \int_0^z \frac{dz'}{(1+z')^{3/2} (1 + \Omega_{\Lambda 0} [(1+z')^{3w} - 1])^{-1/2}} \quad (7.45)$$

Donc, les données de SNe Ia pourraient en principe contraindre  $w_0$  et  $w_1$ . On note cependant que le nombre de variables que l'on cherche à mesurer devient dangereusement semblable au nombre de contraintes observationnelles disponibles. Si l'on tient néanmoins pour acquis le résultat d'autres analyses cosmographiques selon lesquelles  $\Omega_{m0} = 0,27 \pm 0,04$ , l'analyse des SNe Ia donne  $w_0 = -1,31 \pm 0,22$ ,  $w_1 = -1,48 \pm 0,81$  (avec des incertitudes fortement corrélées). Comme les données sont certainement compatibles avec une équation d'état statique ( $w_1 = 0$ ), il est légitime de revenir au modèle plus simple  $w = w_0$ . L'analyse correspondante des données actuelles indique que  $w = -1,02 \pm 0,13$ , avec  $w < -0,76$  à 95 % de confiance. En d'autres termes, le terme répulsif pourrait très bien se réduire à la seule constante cosmologique  $\Lambda$ .

La beauté de ces résultats et la faiblesse des erreurs statistiques associées ne doivent cependant pas occulter le fait que les vraies sources de difficultés sont les erreurs systématiques. L'origine de ces erreurs tient à l'utilisation d'un modèle trop simplifié de la mesure qui ne prend donc pas en compte une source d'erreur cachée, comme l'existence de deux types de céphéïdes dans les mesures de distances d'Andromède par Hubble. C'est pourquoi l'essentiel du travail d'analyse des équipes a sans doute été dévolu à la recherche et à l'élimination des erreurs systématiques. L'existence potentielle d'une source d'absorption inconnue, génériquement de la poussière, est une préoccupation majeure<sup>25</sup>, puisque l'effet mesuré est un assombrissement d'une fraction de magnitude. Cet effet-là est relativement simple à prendre en compte, car cette extinction augmente avec la distance. La difficulté est plutôt que les propriétés d'extinction pourraient évoluer sur des échelles de temps cosmologiques. Et bien sûr se pose la question de l'évolution potentielle des progéniteurs de SNe Ia à de tels  $z$ .

Jusqu'à présent, rien ne semble devoir altérer les conclusions ci-dessus, et les chercheurs de ce domaine pensent que les erreurs systématiques sont inférieures aux incertitudes actuelles de nature statistique. Le problème est que si l'on pense que les SNe Ia correspondent très probablement à l'explosion thermonucléaire de naines blanches, le phénomène n'est pas complètement compris théoriquement, et le système progéniteur lui-même n'a pas été identifié avec certitude [106]. La théorie ne peut donc malheureusement pas servir

<sup>25</sup>Outre les effets classiques, comme la nécessité d'une « correction K » due au fait que l'on observe dans une bande spectrale fixe, alors que le spectre est décalé vers le rouge selon la distance ; c'est donc une partie différente du spectre au repos qui arrive dans la bande. Ce biais est néanmoins bien connu ; il peut être corrigé en prenant en compte la forme spectrale. D'autres effets doivent, et sont, aussi pris en compte, comme le biais de Malmquist et les effets de lentille gravitationnelle.

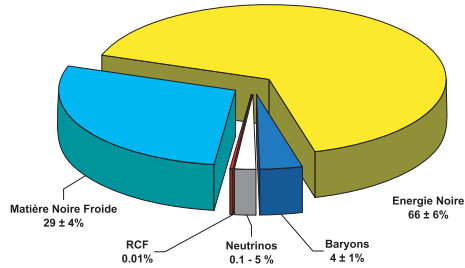


FIG. 7.13 – *Le recensement cosmologique est bien surprenant ! La matière usuelle, les atomes, ne constitue que quelques pour cent du total. . .*

de guide pour savoir si, et comment, les SNe Ia évoluent à différents  $z$ . Pour le moment la diversité constatée des SNe Ia a bien été prise en compte par les méthodes empiriques de mise au gabarit. De plus, il est rassurant de constater que les SNe Ia ont été observées dans un grand nombre d'environnements, des galaxies de tous types (elliptiques, spirales, naines et irrégulières), et que le domaine de variation des propriétés de ces environnements (métallicité, âge des étoiles) est intrinsèquement plus large que ne l'est l'évolution de la moyenne entre  $z = 0$  et  $z = 0,5$ . Remarquons toutefois qu'au cours des dernières années, des SNe Ia vraiment bizarres (hors gabarit) ont été trouvées et il se pourrait que ces déviantes soient plus nombreuses à grand  $z$  ; cela laisse à penser qu'un autre paramètre pourrait être nécessaire pour décrire (empiriquement) toute la famille des SNe Ia. Bien sûr, de nombreux efforts sont actuellement dévolus à la constitution d'échantillons plus vastes où chaque SNe Ia est observée en grand détail pour affiner les tests. Ce faisant, les levées de dégénérescence entre paramètres permettront d'augmenter le nombre de tests véritablement indépendants qui mettent à l'épreuve la cohérence du paradigme.

En conclusion, les tests cosmographiques permettent de dresser le bilan énergétique de la figure 7.13, qui représente bien la vision actuelle. Mais il est clair qu'à elles seules les indications fournies jusqu'à maintenant n'auraient sans doute pas suffi à emporter l'adhésion de la communauté. Il faut maintenant nous tourner vers la cosmogonie, dont se tirent les arguments les plus persuasifs.

## 7.4 Cosmogonie

La cosmogonie décrit l'émergence des structures de l'Univers. La figure 7.14 permet de visualiser la distribution autour de nous des galaxies du relevé SDSS dans une tranche équatoriale. Il s'agit des mêmes données que celles de la figure 7.1, mais présentées différemment. Le développement par instabilité gravitationnelle de fluctuations primordiales permet de rendre compte de l'existence et des propriétés statistiques de ces grandes structures.

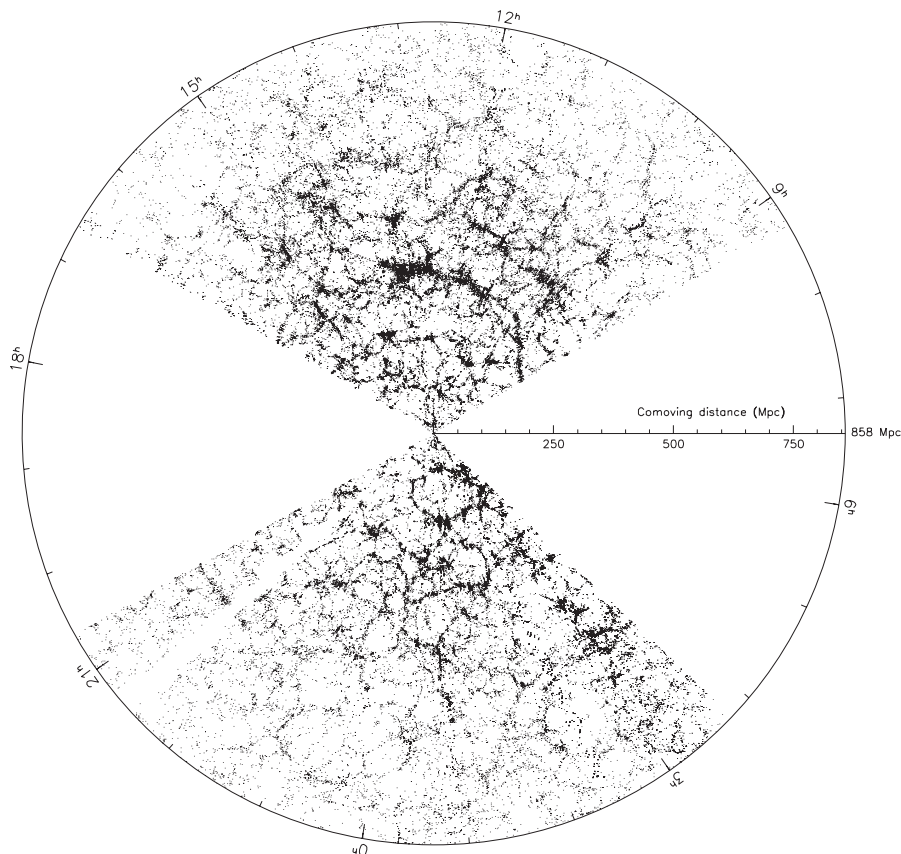


FIG. 7.14 – Tranche équatoriale du relevé SDSS. (D'après [2].)

### 7.4.1 Évolution récente des structures

La figure 7.15a montre une simulation numérique de l'évolution gravitationnelle des structures de la matière froide qui se sont développées au sein d'un cube de 150 Mpc de côté dans un scénario  $\Lambda$ CDM. L'image 7.15b permet de suivre l'évolution temporelle de la densité, dans une tranche mince extraite de cette boîte. On constate que pour ce type de scénario, la croissance des structures est clairement hiérarchique. Quand un objet se forme, le gaz baryonique suit initialement la chute de la matière sombre. Néanmoins, des chocs se produisent quand le matériau frais percute la distribution déjà relaxée. Ces chocs chauffent le gaz, qui peut cependant dissiper son énergie en rayonnant et éventuellement former un disque. La fragmentation en étoiles et trous noirs du gaz refroidi libère des photons ionisants. Ils vont réioniser l'Univers (en fait le gaz intergalactique). En outre, ils rétroagissent sur le



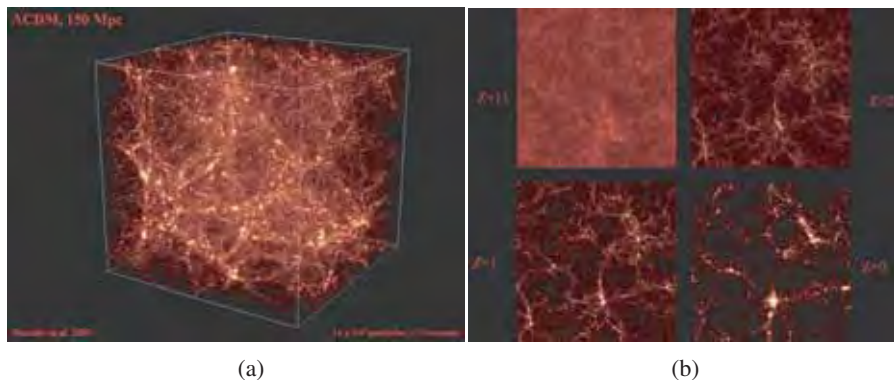


FIG. 7.15 – Simulation numérique d'une cube de 150 Mpc dans un univers  $\Lambda$ CDM ( $\Omega_{m0} = 1/3, \Omega_{\Lambda0} = 1 - \Omega_{m0}, H_0 = 2/3, n_s = 1$ ). (a) Distribution actuelle de la matière sombre (luminosité proportionnelle à la densité). (b) Évolution temporelle par instabilité gravitationnelle d'une tranche mince de la boîte de calcul (15 Mpc) qui montre le développement hiérarchique des structures au sein d'une ossature cosmique de contraste croissant, mais dont les grandes lignes sont en place dès le début.

processus d'effondrement du gaz baryonique en cours d'accrétion, sans compter l'effet des vent de supernovae qui enrichit le milieu en métaux et change ses propriétés de refroidissement. . .

Comme les baryons ne représentent à grande échelle qu'une contribution faible à la densité totale de matière, les galaxies doivent alors être considérées comme de simple traceurs, des jalons dont la répartition permet de révéler la distribution de masse sous-jacente. Une des difficultés est néanmoins qu'aux échelles plus petites, elles constituent un traceur biaisé qui ne reflète pas fidèlement la structuration de la matière sombre, à cause précisément de tous les processus de rétroaction mentionnés ci-dessus. Cette question du biais matière-lumière a engendré de multiples débats enflammés sur son amplitude en fonction de l'échelle ( $\lesssim 5$  Mpc), car elle est centrale à toute comparaison théorie-observation à ces échelles. Cependant les analyses détaillées des grands relevés comme le SDSS démontrent qu'il n'y a, à  $z \sim 0$ , effectivement pas de biais à grande échelle pour les galaxies « normales ». Quoi qu'il en soit, les comparaisons aux grandes échelles ont effectivement rencontré des succès quantitatifs considérables que laisse présager la similitude visuelle frappante de la distribution des galaxies de la figure 7.14 et celle de la matière sombre de la figure 7.15.

Les simulations nous offrent aussi la possibilité de visualiser le déplacement dans l'espace-temps que nous effectuons, quand nous considérons des observations lointaines, le long du cône de lumière passé. Ainsi la figure 7.16 montre comment l'observateur (en bas à gauche) qui effectue un relevé en pinceau très profond « voit » les structures à l'époque où elles ont émis la lumière qui nous

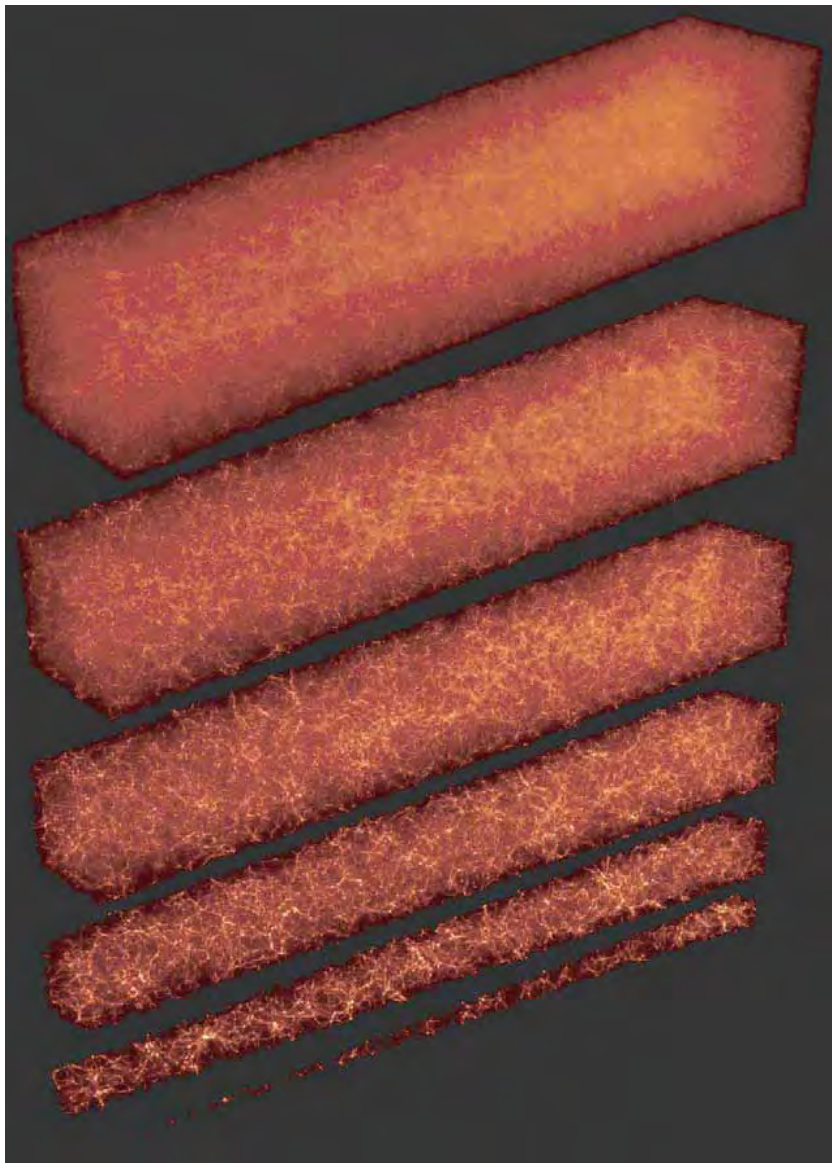


FIG. 7.16 – Cône de lumière passé, replié sur lui-même pour les besoins de la représentation. La véritable distribution de la matière de ce long pinceau s'obtient en mettant bout à bout les sections de cône successives. Cette image illustre de manière frappante que voir loin, c'est voir le passé lointain. (Source IAP-CNRS, S. Colombi.)

parvient aujourd'hui. On remonte donc le temps en s'éloignant ; ainsi voit-on l'instabilité gravitationnelle à l'œuvre en transformant un champ gaussien en un réseau cosmique, qui n'est pas sans rappeler l'ossature d'une éponge. En haut à droite, le décalage vers le rouge du spectre d'une galaxie serait de  $\Delta\lambda/\lambda \sim 6$  (comme les quasars les plus lointains du SDSS). Évidemment cette visualisation nous affranchit, et c'est son mérite, des autres effets intrinsèques à de véritables observations, comme l'échantillonnage de plus en plus parcimonieux à grandes distances. Outre leur évolution propre, les monstres lumineux qui nous servent de jalons éloignés offrent, par leur rareté même, une représentation de plus en plus biaisée des structures sous-jacentes. Cette imbrication d'effets cosmologiques et observationnels explique pourquoi les simulations numériques ont joué au cours des vingt dernières années un si grand rôle dans le développement de la cosmologie. La figure 7.19 nous en offrira un autre exemple.

À elles seules, ces images ne permettent cependant pas de savoir en quoi ce scénario  $\Lambda$ CDM est une recette cosmogonique satisfaisante. Pour répondre à cette question, il faut maintenant nous intéresser à la physique des premières phases de l'évolution.

## 7.4.2 Premières phases de l'évolution des structures

L'évolution des fluctuations de la matière est très simple, tant que les contrastes sont faibles. Soit  $\delta(x)$  le contraste de densité en un point  $x$ ,

$$\delta(x) = \frac{\rho(x) - \bar{\rho}}{\bar{\rho}} \quad (7.46)$$

où  $\bar{\rho}$  dénote la densité moyenne et où la position  $x$  est repérée dans des coordonnées comobiles (qui sont « entraînés par l'expansion », Éq. (7.12)). Il s'avère en fait plus commode de considérer ce contraste de densité comme une superposition d'ondes planes<sup>26</sup>, de modes de longueurs d'onde  $\lambda = 2\pi/k$ , et d'amplitudes  $\delta_{\vec{k}}$ . Tant que les contrastes sont faibles,  $\delta \ll 1$ , l'évolution est linéaire, et chacun de ces modes évolue indépendamment des autres. On peut donc sans perte de généralité ne s'intéresser qu'à ces fluctuations sinusoïdales, puisque toute autre fluctuation se ramène à une superposition de ces fluctuations élémentaires. Le champ de densité est une variable aléatoire, et seules ses propriétés statistiques nous intéressent. Dans le cas des fluctuations primordiales, l'hypothèse minimale est qu'il s'agissait d'un champ régi par la

---

<sup>26</sup>Il s'agit d'une décomposition en modes de Fourier,  $\delta_{\vec{k}}$ , et on a  $\delta(x) = \int d^3k/(2\pi)^3 e^{-i\vec{k}\cdot\vec{x}} \delta_{\vec{k}}$  ; chacun des modes s'obtient par la transformation inverse  $\delta_{\vec{k}} = \int d^3x e^{i\vec{k}\cdot\vec{x}} \delta(x)$ .

statistique gaussienne<sup>27</sup>, sans corrélation entre les modes. La valeur moyenne de l'amplitude des fluctuations de nombre d'onde  $k$  est nulle,  $\langle \delta_k \rangle = 0$ , car les surdensités sont aussi probables que les sous-densités. En revanche, l'amplitude typique des fluctuations à l'échelle  $k$  est donnée par  $\langle |\delta_k|^2 \rangle^{1/2}$ . La quantité

$$P(k) = \langle |\delta_k|^2 \rangle \tag{7.47}$$

est appelée le spectre de puissance. C'est une quantité fondamentale qui suffit à caractériser complètement les propriétés statistiques initiales du champ de densité; elle indique l'amplitude relative des fluctuations selon l'échelle. La contribution de l'échelle  $\lambda$  aux fluctuations de  $\delta$  sont alors données<sup>28</sup> par  $[k^3 P(k)]^{1/2}$  pris à  $k = 2\pi/\lambda$ . Nous allons considérer le cas où le spectre initial,  $P_i$ , est en loi de puissance,

$$P_i(k) \propto k^n \tag{7.48}$$

Soit  $d_H = cH^{-1}$  le rayon de Hubble. Dans ce qui suit, cette échelle joue un rôle important, car elle indique la limite *instantanée* des processus « causaux ». En effet si  $\Gamma$  est la probabilité par unité de temps qu'une interaction se produise, la réaction correspondante sera à l'équilibre si son temps caractéristique  $\Gamma^{-1}$  est inférieur au temps caractéristique d'expansion  $H^{-1}$ ; dans le cas contraire, elle sera « gelée », c'est-à-dire incapable de réagir assez rapidement aux conditions changeantes créées par l'expansion. Par ailleurs, la portée d'une interaction étant  $\lesssim c\Gamma^{-1}$ , une réaction ne pourra affecter que des échelles inférieures au rayon de Hubble (qui vaut  $\simeq 3h^{-1}$  Gpc aujourd'hui). Nous avons vu (Éq. (7.27)) que  $a \propto t^{1/2}$  pendant l'ère du rayonnement (quand  $\rho_{tot} \simeq \rho_r \propto a^{-4}$ , à  $a(t) < a_{eq}$ ), et  $a \propto t^{2/3}$  par la suite, dans l'ère de la matière (tant que  $\rho_{tot} \simeq \rho_m \propto a^{-3}$ ). Comme le rayon de Hubble est proportionnel au temps pendant ces deux ères, on constate donc que  $d_H \propto a^2$  durant l'ère de rayonnement, et  $d_H \propto a^{3/2}$  par la suite. En fait il est plus commode de considérer des longueurs comobiles, et on notera  $\lambda_H = d_H a_0/a(t)$ . On a

<sup>27</sup>On a  $\langle \delta_k \delta_{k'}^* \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P(k)$ , et la probabilité qu'un mode ait une amplitude  $\delta_k$  est  $P(\delta_k) \propto \exp -|\delta_k|^2/[2P(k)]$ , indépendamment de sa direction (pour maintenir l'isotropie en moyenne). Les phases sont, elles aussi, sans aucune corrélation, avec une distribution équiprobable entre 0 et  $2\pi$ . La transformée de Fourier inverse du spectre de puissance  $P(k)$  est la fonction de corrélation à deux points du champ,  $\xi(r) = \langle \delta(x)\delta(x+r) \rangle$ , et en  $r = 0$ ,  $\xi(0) = \langle |\delta|^2 \rangle = \int P(k) d^3k / (2\pi)^3$ .

<sup>28</sup>L'intégrale du spectre de puissance donne la variance du champ correspondant  $\langle \delta^2 \rangle = \int P(k) d^3k / (2\pi)^3 \propto \int k^3 P(k) d \ln k$  (voir la note 27). Chaque intervalle logarithmique de  $k = 2\pi/\lambda$  contribue donc  $\sim k^3 P(k)$  à la variance.

Pour plus de précision, on peut procéder en considérant  $\bar{\delta}(x, \lambda)$ , la valeur de  $\delta$  moyennée sur, par exemple, une sphère (ou une boule gaussienne) d'échelle caractéristique  $\lambda$  centrée en  $x$ , notée  $W_\lambda(x) : \bar{\delta}(x, \lambda) = \int \delta(x') W_\lambda(x - x') d^3x'$ . Les fluctuations typiques de  $\delta$  lissées à l'échelle  $\lambda$  sont donc indiquées par la variance  $\sigma_\delta^2(\lambda) \equiv \langle \bar{\delta}^2(x, \lambda) \rangle \propto \int k^3 P(k) \bar{W}_\lambda^2(k) d^3k$ , où  $\bar{W}_\lambda(k) = \int W_\lambda(x') \exp(ikx') d^3x$  est la transformée de Fourier de la fonction fenêtre. La contribution d'une gamme d'échelles  $\Delta$  centrée autour de  $\lambda_c$  peut alors s'obtenir en considérant une fonction fenêtre centrée autour de  $\lambda_c$  et de largeur  $\Delta$ . On peut choisir par exemple la gaussienne  $\propto \exp -(\lambda - \lambda_c)^2 / (2\Delta^2)$ , dont la transformée est aussi une gaussienne,  $\propto \exp -\Delta^2(k - k_c)^2 / 2$  centrée autour de  $k_c = 2\pi/\lambda_c$ , ce qui permettra de raffiner l'estimation  $k_c^3 P(k_c) \Delta / \lambda_c$ .

alors  $\lambda_H \propto a$  et  $a^{1/2}$  dans, respectivement, l'ère du rayonnement et celle de la matière.

Selon la relativité générale, aux échelles  $\lambda > \lambda_H$  ( $k < k_H$ ), le contraste augmente proportionnellement au carré du facteur d'échelle pendant l'ère du rayonnement, et comme le facteur d'échelle pendant l'ère de la matière<sup>29</sup>. Il est intéressant de remarquer que ce contraste croît comme le carré de la croissance du rayon de Hubble,  $\delta_k \propto \lambda_H^2(a)$ . Il s'ensuit que  $k^3 P(k) \propto k^3 P_i(k) \lambda_H^4(a) \propto k^3 P_i(k) k_H^{-4} = k^{(3+n)}/k_H^4(a)$ . On voit donc que dans le cas particulier où  $n = 1$ , on a  $k^3 P(k) \propto (k/k_H)^4$ , i.e. chaque gamme d'échelle apporte la même contribution à la variance des fluctuations au moment où cette échelle est égale au rayon de Hubble. Pour parler de manière imagée, on décrit en général cet instant comme le moment où une échelle « entre » dans le rayon de Hubble (qui signale l'échelle maximale d'influence des processus causaux). Donc dans le cas  $n = 1$ , la décroissance de l'amplitude initiale avec l'échelle est juste compensée par le surplus de croissance permis par le temps supplémentaire pour que ces échelles plus grandes entrent à leur tour dans le rayon de Hubble. Le démarrage des processus causaux part dans ce cas d'une situation invariante, d'où le nom de spectre invariant d'échelle; ce cas  $n = 1$  est en fait le spectre dit de Harrison-Peebles-Zeldovich<sup>30</sup>.

Aux échelles  $\lambda < \lambda_H$ , les fluctuations croissent, comme aux plus grandes échelles, proportionnellement à  $a(t)$  pendant l'ère de la matière. Il suffit donc de spécifier le spectre au moment de l'équivalence, puisque toutes les échelles croissent d'un même facteur  $a(t)$  après. Pour obtenir le spectre de fluctuation

---

<sup>29</sup>Ce qui peut se comprendre grâce à la comparaison suivante. Soit l'évolution d'un morceau sphérique d'univers, U2, légèrement surdense, par rapport à un univers U1 servant de fond de référence et ayant  $\Omega = 1$ . On a alors

$$H_1^2 = \frac{8\pi G\bar{\rho}}{3} \quad \text{et} \quad H_2^2 = \frac{8\pi G\rho}{3} - \frac{k}{a^2} \tag{7.49}$$

avec  $\rho = \bar{\rho}(1 + \delta)$ . Notons que si la taille de la fluctuation,  $\lambda$  est supérieure à  $\lambda_H$ , il existe une ambiguïté due à l'impossibilité à des observateurs, respectivement à l'intérieur de U2 et à l'extérieur, de « comparer leurs notes », du moins tant que  $\lambda > \lambda_H$ . Cette ambiguïté peut être levée par le choix d'une jauge particulière, dite de flot de Hubble uniforme; c'est-à-dire qu'on compare  $\rho$  et  $\bar{\rho}$  quand  $H_1 = H_2$ . C'est d'ailleurs un choix naturel, si  $\lambda < \lambda_H$  et  $\sigma_\rho \ll 1$ . On a alors

$$\frac{\delta\rho}{\bar{\rho}} = \frac{k/a^2}{8\pi G\bar{\rho}/3} \tag{7.50}$$

On voit que si  $\delta \ll 1$ , i.e. dans le cas linéaire, on peut prendre  $a_2 \sim a_1 \equiv a$ , et on trouve bien  $\delta \propto a^2$  ( $\propto t$ ) pendant l'ère de rayonnement ( $\bar{\rho} \propto a^{-4}$ ) et  $\delta \propto a$  ( $\propto t^{2/3}$ ) dans l'ère de matière ( $\bar{\rho} \propto a^{-3}$ ).

<sup>30</sup>Nous venons de discuter le comportement de perturbations aux échelles plus grandes que la longueur de Hubble. Ceci semble impliquer qu'il faut renoncer à créer ces fluctuations par un processus physique de l'Univers primordial et les mettre « ab initio » puisque aucun processus causal ne semble pouvoir jouer sur ces échelles. Le modèle d'inflation apporte une solution élégante à ce problème, quoiqu'on puisse en imaginer d'autres solutions (cordes, textures, etc.). En effet, au cours de cette période, on a  $H = \text{constante}$ , avec une inflation exponentielle tant que dure cette phase de Sitter. Certaines échelles peuvent donc être successivement plus petites que  $\lambda_H$ , puis devenir plus grandes, avant de finalement redevenir plus petites. En d'autres termes, l'horizon causal global peut être beaucoup plus grand que le rayon de Hubble qui ne fixe que la limite instantanée des processus causaux.

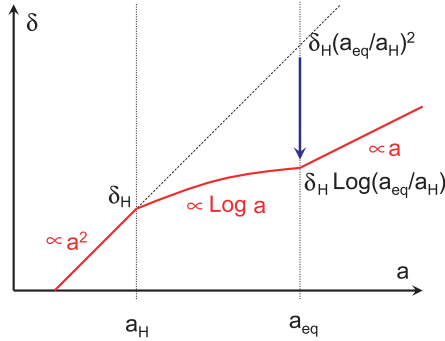


FIG. 7.17 – La faible croissance (logarithmique) de fluctuations « entrées dans le rayon de Hubble » ( $\lambda < \lambda_H(t)$ ) pendant l'ère de rayonnement se traduit par un amortissement relatif de ces petites échelles que la figure 7.18 illustre.

aujourd'hui si l'évolution avait toujours été linéaire, nous n'aurons qu'à multiplier le spectre à l'équivalence par le carré du taux de croissance depuis cette époque<sup>31</sup>  $\sim 1 + z_{eq}$ . En revanche, pendant l'ère de rayonnement où l'expansion est plus rapide, les fluctuations ne croissent presque pas, logarithmiquement tout au plus. Comme on peut le voir sur la figure 7.17, il s'ensuit un déficit relatif de croissance des échelles qui sont « entrées » sous le rayon de Hubble pendant l'ère de rayonnement. Appelons  $a_H(\lambda)$  cet instant. Le facteur de suppression  $A = \log(a_{eq}/a_H)/(a_{eq}/a_H)^2$  se traduit par une modification de l'amplitude relative des fluctuations aux différentes échelles. La valeur de  $a_H(\lambda)$  est déterminée par  $\lambda = \lambda_H(a_H(\lambda))$ . Pendant l'ère de rayonnement, il faut résoudre  $\lambda = a_H(\lambda)/a_{eq}\lambda_H(t_{eq})$ . Si on note  $\lambda_0$  le rayon de Hubble à l'équivalence, l'amortissement relatif de l'échelle  $\lambda$  est donné par  $A = \log(\lambda_0/\lambda) (\lambda_0/\lambda)^{-2} = \log(k/k_0) (k/k_0)^{-4}$ . Le spectre transformé est alors le spectre au moment de l'entrée dans le rayon de Hubble,  $P_H(k)$  fois  $A^2(k)$ . Dans le cas invariant d'échelle  $n = 1$ , on a

$$P(k) \propto \begin{cases} k & \text{quand } k < k_0 \\ k^{-3} \log^2(k/k_0) & k > k_0 \end{cases} \quad (7.52)$$

Les figures 7.18a et 7.18b montrent en bleu le spectre  $P(k)$  et  $[k^3 P(k)]^{1/2}$  qui correspondent au scénario SCDM, dans lequel  $n = 1$  avec  $\Omega_{m0} = 1 - \Omega_{r0}$  et  $h^2 = 0,5$ . En vert, on a plutôt choisi  $\Omega_{m0}h^2 = 0,15$  (cas où  $\Omega_{m0} \sim 0,3$ ).

<sup>31</sup>Le taux de croissance  $D(t)$  pendant l'ère de matière est donné par le facteur d'expansion. En revanche, si  $\Omega_{m0}$  n'est pas égal à  $1 - \Omega_{r0}$ , la croissance après l'ère de la matière devient plus faible voire s'annule quand  $\Omega_{k0}$  ou  $\Omega_{\Lambda 0}$  devient le terme dominant de l'équation de Friedman. La formule

$$\frac{D(t)}{a(t)} = \frac{5}{2} \Omega_m(a) \left[ \Omega_m^{4/7}(a) - \Omega_\Lambda(a) + \left(1 + \frac{\Omega_m}{2}\right) \left(1 + \frac{\Omega_\Lambda}{70}\right) \right] \quad (7.51)$$

décrit à quelques pour cent près le facteur correctif.

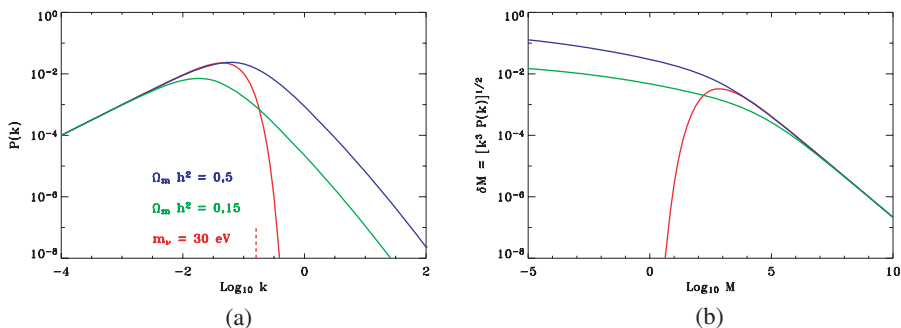


FIG. 7.18 – Comparaisons des fluctuations disponibles dans l'ère de matière pour former les structures, en fonction du type et de la quantité de matière sombre. (a) Spectre de puissance. (b) Amplitude des fluctuations à une échelle de masse donnée qui s'en déduit. Dans le cas de matière sombre chaude (en rouge) ce sont de grandes échelles qui deviennent non-linéaires en premier, et la formation des structures procède ensuite par fragmentation. On sait aujourd'hui que la formation est plutôt hiérarchique.

La transition rayonnement-matière ayant lieu plus tard, l'amortissement affecte un plus grand nombre d'échelles (des  $k$  plus petits), où si l'on préfère, le basculement entre un comportement logarithmique à petite échelle et le comportement  $\propto M^{-2/3}$  de  $[k^3 P(k)]^{1/2}$  se produit à plus grande échelle. D'un point de vue observationnel, c'est d'abord aux petites échelles qu'on a bien mesuré les fluctuations; la déficience du scénario CDM dans sa forme initiale s'est alors manifestée par le fait qu'il n'y avait (relativement) pas assez de puissance à grande échelle dans ce scénario. C'est ce qu'a montré nettement l'analyse du relevé de galaxies APM par Maddox et collaborateurs en 1990 [107]. Ceci est un exemple typique des analyses cosmogoniques qui ont progressivement amené à considérer sérieusement un scénario de type  $\Lambda$ CDM, bien avant toute indication plus directe comme celle fournie par les SNe Ia. On peut juger de l'accord entre le spectre de puissance mesuré aujourd'hui dans le relevé SDSS et une description  $\Lambda$ CDM sur la figure 7.26.

La figure 7.18 montre aussi en rouge le spectre qu'on aurait dans le cas de matière sombre chaude. Comme je l'ai indiqué précédemment, les fluctuations de petite échelle sont dans ce cas amorties. En effet ces particules diffusent librement, à une vitesse proche de  $c$ , tant qu'elles sont relativistes. Ceci introduit une coupure exponentielle du spectre à  $k_{FS} = 0,16 \text{ Mpc}^{-1} (m_\nu/30 \text{ eV})$  (indiquée par un trait tireté sur la figure 7.18b), qui est en fait très proche de  $k_0$  pour  $m_\nu \simeq 30 \text{ eV}$ . Dans ce cas (cf. 7.18c), on n'a pas une croissance hiérarchique, mais plutôt la formation de structures de masse  $M \simeq M_{FS} \simeq 10^{14} M_\odot$  (amas et superamas).

La figure 7.19 compare l'évolution non-linéaire récente de la texture des grandes structures selon le spectre de fluctuation adopté. La colonne de gauche

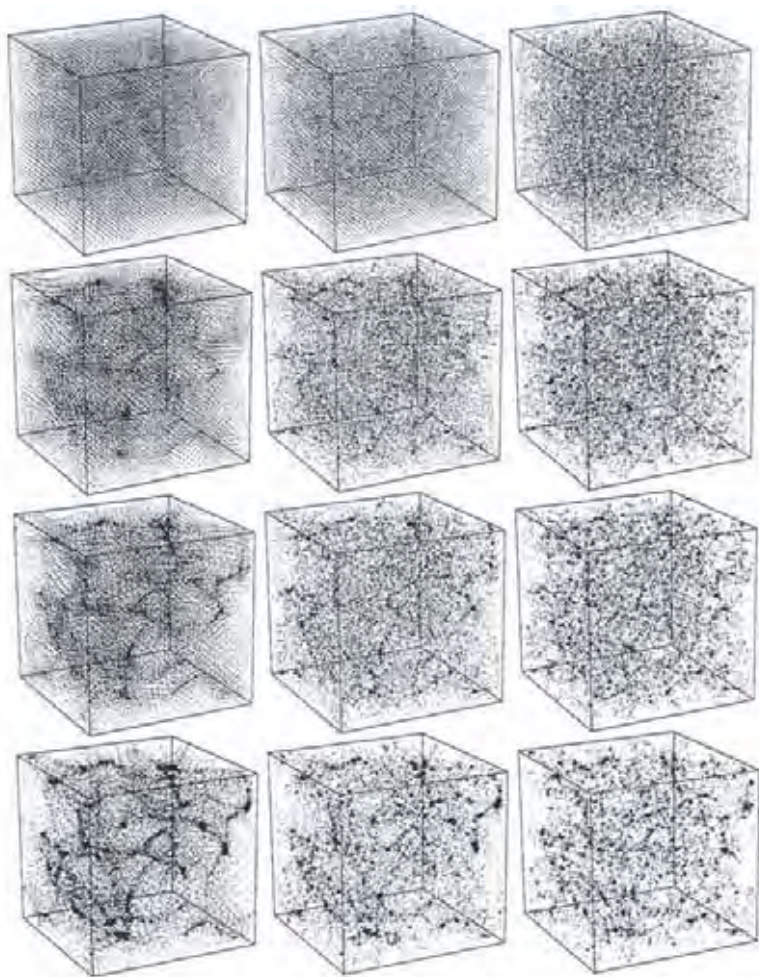


FIG. 7.19 – Développement temporel des grandes structures pour divers types de conditions initiales. De gauche à droite, il y a de plus en plus de puissance à petite échelle relativement aux grandes, ce que le développement temporel (de haut en bas) révèle. La colonne de gauche illustre ainsi le cas d'un scénario à base de matière chaude, et celle du centre celui d'un scénario à base de matière froide (la colonne de droite correspond à du bruit blanc, où le spectre est invariant en fonction de l'échelle). Cette figure illustre le type de fluctuations effectuées vers 1990. (D'après [108].)

montre l'évolution d'un univers dont la densité d'énergie est dominée par de la matière noire chaude, si bien qu'il y a très peu de puissance à petite échelle dès les conditions initiales de la simulation. Dans un scénario de ce type,



les très grandes structures se forment d'abord ; les structures plus petites se forment ensuite par fragmentation de la composante baryonique (scénario dit « top-down »). Dans la colonne centrale, qui correspond à un univers dominé par de la matière sombre froide, il y a suffisamment de puissance aux petites échelles pour que les petites structures (sub-galactiques) se forment en premier pour s'agglutiner et fusionner en des structures de plus en plus vastes (scénario « bottom-up »). La colonne de droite montre l'évolution de conditions initiales de type bruit blanc (et donc sans aucune échelle caractéristique), ayant encore plus de puissance aux petites échelles. Ce cas sert de contrôle et d'extrême dans une série, même si ce type de conditions initiales n'a jamais été envisagé comme réaliste.

Ce type d'études et leurs comparaisons aux données ont ainsi permis de progressivement sélectionner les ingrédients nécessaires à un scénario cosmogonique satisfaisant. Mais c'est l'étude des anisotropies fossiles qui par leur immédiateté font penser que nous avons sans doute finalement abouti à une description appropriée dans ses grandes lignes.

### 7.4.3 Anisotropies fossiles

Qu'observe-t-on au juste quand on pointe le télescope dans une direction quelconque pour observer le fond diffus du rayonnement fossile (qu'on nommera dorénavant Rayonnement Cosmologique de Fond – RCF en abrégé) ? Pour cela, reprenons le cours de l'histoire thermique de l'Univers. Quelques minutes après le Big Bang, le gaz ionisé quasi homogène qui remplit l'espace est constitué de protons, d'électrons, de photons (ainsi que des quelques noyaux de deutérium et autres éléments légers qui se sont formés au cours de la nucléosynthèse et qu'on peut négliger ici). La trajectoire d'un photon est déviée chaque fois qu'il subit une interaction (de type Compton) avec un des électrons libres. L'Univers est alors un plasma, opaque au rayonnement, comme un brouillard (les électrons jouant le rôle des gouttelettes d'eau diffusantes). Au cours du temps, les distances augmentent, et cette expansion se traduit par une baisse de température. Mais pendant longtemps il ne se passe pas grand-chose ; si on regarde ce qu'il advient dans une région donnée de l'Univers, l'image à différents temps est identique, pourvu qu'on ramène le cube à sa taille précédente par un changement d'échelle. Autrement dit, on utilise des coordonnées *comobiles* qui annulent l'effet de l'expansion sur les distances, si bien qu'un cube est représenté de la même taille à un moment ultérieur, même si, physiquement, sa taille a augmenté. Dans ces coordonnées, l'image du cube est statistiquement invariante.

Cependant, quand la température a suffisamment baissé, des atomes vont pouvoir se former, sans que les collisions avec les photons soient suffisantes pour tous les dissocier. Un certain nombre d'atomes neutres commence donc à apparaître, ce qui diminue le nombre d'électrons avec lesquels les photons peuvent interagir. C'est l'époque de la combinaison en atomes. En conséquence, le parcours moyen entre deux interactions a tendance à augmenter,

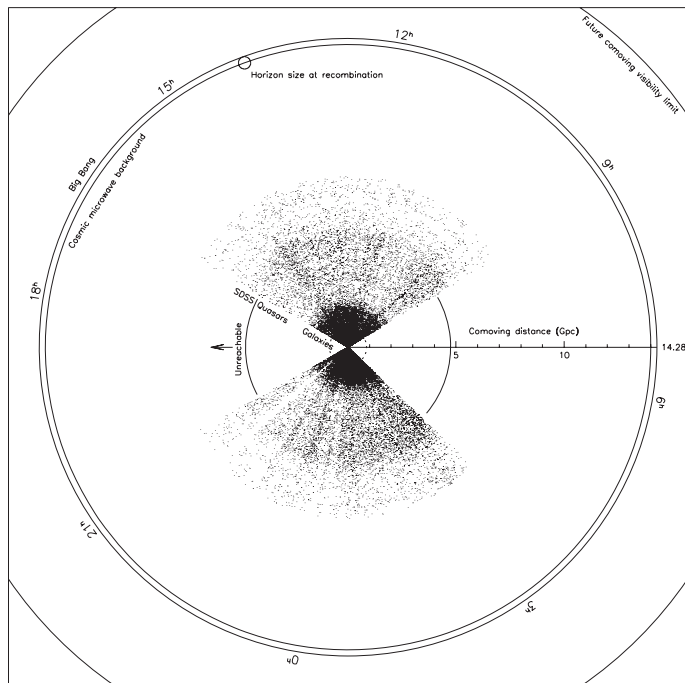


FIG. 7.20 – La surface de dernière diffusion en perspective. L’observateur est au centre. Le cercle en pointillé à 850 Mpc montre l’extension de la distribution des galaxies du relevé SDSS que montrait la figure 7.14. Les quasars les plus lointains détectés sont à plus de 10 Gpc. Cette distance est comparable au rayon de 14 Gpc de la surface de dernière diffusion (SDD) des photons du rayonnement fossile (voir la fin du paragraphe 7.3.2 pour les autres distances indiquées). (D’après [2].)

et au bout d’un certain temps – qui est en fait très court à l’échelle cosmologique – le milieu est presque totalement neutre et le libre parcours des photons a tellement augmenté qu’il est quasiment infini. En d’autres termes, le milieu jusqu’ici opaque, est rapidement devenu transparent. Ce découplage des photons se produit environ 380 000 ans après le Big Bang, à  $z_{dec} = 1100$ . L’observateur du RCF observe donc des photons provenant d’une coquille sphérique dont l’épaisseur, faible, est donnée par la durée relativement brève qu’il a fallu pour que l’Univers devienne transparent ( $\Delta z \sim 80$ ). On peut en général négliger cette épaisseur et considérer que nous observons la « Surface » de Dernière Diffusion – SDD en abrégé – des photons sur les électrons du gaz ionisé primordial (à tout le moins dans les directions du ciel où il n’existe pas de zones faisant écran ou superposant leur émission à l’émission primordiale). La figure 7.20 permet de voir où se positionne en coordonnées comobiles la SDD par rapport aux objets du SDSS déjà représentés en figure 7.1.

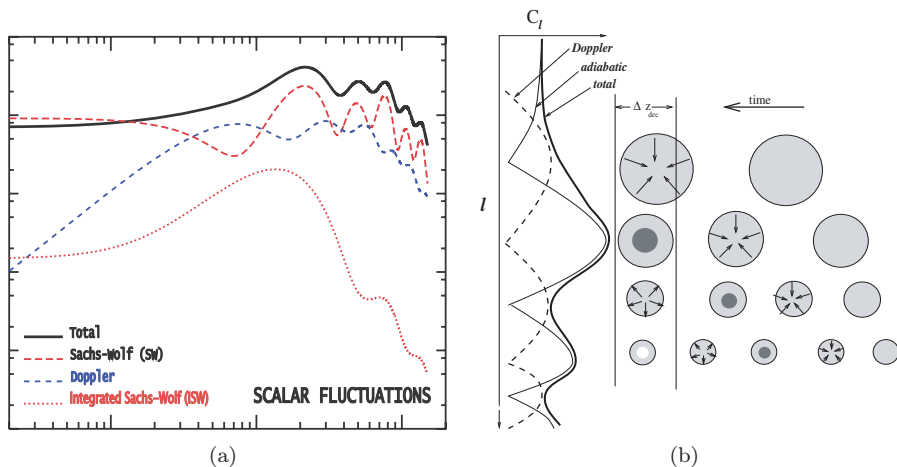


FIG. 7.21 – (a) Forme attendue du spectre des anisotropies de température,  $l^2 C(\ell)$  et contributions relatives des divers effets (on a supposé qu'il n'y avait que des fluctuations adiabatiques scalaires et le diagramme est en coordonnées log-log. (b) Au cours du temps, des fluctuations de plus en plus grande taille (extension spatiale) commencent à osciller et laissent leur empreinte caractéristique sur le spectre de puissance angulaire. (D'après [109].)

Le RCF porte bien sûr le témoignage des conditions physiques qui régnaient au moment du découplage. En particulier, les petites inhomogénéités à l'origine des grandes structures laissent leurs empreintes. Pour l'observateur, elles se traduisent par des variations de la température observées dans diverses directions ; en d'autres termes il existe des anisotropies de température qui constituent un champ aléatoire sur la sphère. Pour analyser les propriétés statistiques de ce champ, on procède de la même manière que pour les inhomogénéités de densité, sauf que l'espace à trois dimensions est remplacé par la surface de la sphère ; la transformation de Fourier en ondes planes élémentaires est alors remplacée par la transformation en harmoniques sphériques, de coefficients  $a(\ell, m) \equiv a_{\ell m}$ . On caractérise les fluctuations par leur spectre de puissance angulaire,  $C(\ell) = \langle a_{\ell m}^2 \rangle$ . En fait, il ne peut dépendre que du multipole  $\ell$ , et non des orientations des modes indexées par  $m$ , pour assurer l'isotropie des propriétés du champ (tout comme le spectre de puissance spatial ne pouvait dépendre que du module  $k = |\vec{k}|$ ). Un multipole donné correspond à une échelle angulaire  $\theta \sim 180^\circ/\ell$ . Cette statistique à deux points caractérise complètement un champ gaussien. La figure 7.21a montre la forme attendue du  $C(\ell)$  dans une large gamme de modèles cosmogoniques. Enfin, comme pour le contraste de densité, on s'intéresse la plupart du temps à la contribution par intervalle logarithmique d'échelle,  $\ell(2\ell + 1) C(\ell)$  (il y a  $2(2\ell + 1)$  valeurs de  $m$  pour chaque multipole  $\ell$ ).

La forme spécifique du  $C(\ell)$  résulte de l'action combinée de plusieurs phénomènes. L'un est « l'effet Sachs-Wolf » [42]. C'est la perte d'énergie subie par un photon qui doit s'extraire d'un puits de potentiel sur la surface de dernière diffusion (pour à terme nous atteindre) ; cet effet se superpose aux fluctuations intrinsèques de température. Nous avons vu que les fluctuations de la matière sombre ne croissent presque pas pendant l'ère de la radiation ; l'amortissement du spectre spatial aux échelles inférieures au rayon de Hubble à l'équivalence,  $k_0$  en est la résultante. Puis toutes les échelles croissent identiquement d'un facteur  $\propto a$ . Mais ceci ne décrit que le comportement de la matière sombre non baryonique. Le rayonnement avant la combinaison est, lui, si fortement couplé à la matière baryonique (via ses interactions Thomson incessantes avec les électrons) que l'ensemble photons + baryons constitue en fait un fluide unique (les baryons suivant à leur tour, via les collisions électrons-nucléons). Si l'évolution de ce fluide est indiscernable de celle de la matière sombre aux échelles plus grandes que le rayon de Hubble, elle est en revanche assez différente aux échelles plus petites où la pression du rayonnement peut se faire sentir. En effet, quand la gravité fait augmenter le contraste d'une surdensité, la pression croît, puis arrête l'effondrement ; elle peut même inverser le mouvement du fluide (mais pas celui de la matière non-baryonique). La phase d'expansion qui s'ensuit diminue la pression ; la gravité peut alors reprendre le dessus, et engendrer une nouvelle phase de contraction... La figure 7.21b offre une vue d'artiste de ce processus.

Il y a donc des oscillations acoustiques du fluide baryons + photons, aux échelles suffisamment petites pour que la pression puisse jouer, c'est-à-dire quand  $k > k_A$ , où l'échelle acoustique  $k_A$  est fixée par l'inverse de la distance parcourue à la vitesse du son à l'instant  $\eta_*$  considéré. Aux échelles plus grandes que l'horizon sonore ( $k < k_A \propto 1/(c_S \eta_*)$ ), le contraste initial est simplement amplifié. L'amplification est maximale à  $k = k_A$ , tandis qu'à  $k = 2k_A$  la fluctuation a eu le temps de rebondir complètement. De manière générale, les multiples impairs de  $k_A$  sont compressés au maximum, tandis que c'est l'inverse pour les multiples pairs de  $k_A$ . La figure 7.22 montre l'évolution temporelle approchée de l'amplitude de différents modes de Fourier.

L'inertie que les baryons apportent au fluide a pour effet de déplacer le point zéro des oscillations. L'écart type de l'amplitude des modes (diagramme de droite) présente donc un renforcement relatif des pics impairs de compression par rapport aux pics pairs de raréfaction. Cette différence est directement proportionnelle à la quantité de baryons,  $\Omega_{b0} h^2$  (en effet puisque  $\Omega_X$  est le rapport de la densité de  $X$  à la densité critique, qui décroît avec le temps comme  $h^{-2}$ ,  $\Omega_X h^2$  est effectivement proportionnel à la densité physique de  $X$ ).

Supposons un moment que la transition d'opaque à transparent soit instantanée, à  $\eta = \eta_*$ . Ce que nous verrions alors en observant le RCF serait juste l'image directe de ces ondes sur la surface de dernière diffusion ; on attend donc une série de pics à des multipoles  $\ell_A = k_A \times D_*$ , où  $D_*$  est la distance angulaire à la SDD, qui dépend de la géométrie de l'espace-temps. Bien sûr, un  $k$  donné contribue à un intervalle de  $\ell$  (quand  $k$  est perpendiculaire à la ligne

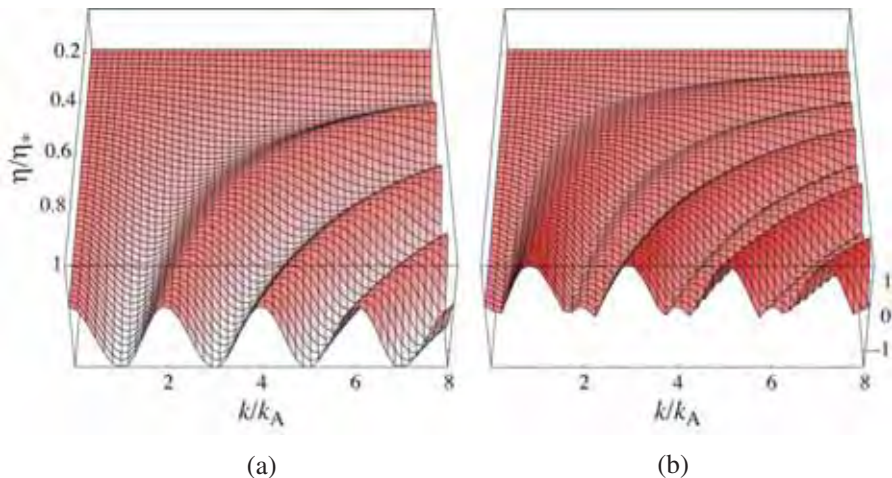


FIG. 7.22 – Évolution temporelle approchée de la température effective qui additionne les effets (de signe opposé) de la température intrinsèque et des fluctuations du potentiel newtonien. (a) Amplitude. Remarquez le déplacement du point zéro qui conduit à un accroissement relatif des compressions. (b) Écart quadratique moyen qui montre l'augmentation des pics impairs (d'après [103]).

de visée, il contribue à des  $\ell$  plus petits que quand l'angle décroît), mais cet étalement est assez limité. Comme l'échelle angulaire acoustique  $\ell_A$  dépend de la géométrie et de la vitesse du son, cette dépendance entraîne sa variation en fonction de la valeur de trois paramètres cosmologiques. On a ainsi

$$\frac{\Delta \ell_A}{\ell_A} \simeq -1,1 \frac{\Delta \Omega_{\text{tot},0}}{\Omega_{\text{tot},0}} - 0,24 \frac{\Delta \Omega_{m0} h^2}{\Omega_{m0} h^2} + 0,07 \frac{\Delta \Omega_{b0} h^2}{\Omega_{b0} h^2} \quad (7.53)$$

au voisinage d'un modèle plat  $\Omega_{\text{tot}} = 1$  ayant 15 % de matière ( $\Omega_{m0} h^2 = 0,15$ ) et 2 % de baryons [110]. La position des pics (et en particulier la position du premier) dépend donc principalement du contenu énergétique total  $\Omega_{\text{tot}}$  (et donc de la géométrie spatiale), la quantité de matière totale  $\Omega_{m0} h^2$  n'intervenant somme toute qu'assez peu.

Comme je l'ai mentionné précédemment, les baryons augmentent l'inertie du fluide baryons + photons et déplacent le point zéro des oscillations. Une densité baryonique plus importante a tendance à augmenter le contraste entre pics pairs et impairs; on peut donc utiliser cette information sur le contraste en hauteur pour connaître la densité baryonique. L'influence de la matière non-baryonique est plus subtile. En fait, elle n'agit pas en tant que telle, mais elle contribue à la densité de matière totale. Une augmentation de la quantité de matière entraîne une transition plus précoce entre ère du rayonnement et ère dominée par la matière, à  $1 + z_{eq} = \Omega_{m0}/\Omega_{r0} \simeq 2,3 \times 10^4 \Omega_{m0} h^2$  (Éq. (7.31)). Cette transition a donc lieu un peu avant la combinaison à

$z \simeq 1100$ . L'amortissement du potentiel gravitationnel des modes qui oscillent dans l'ère de rayonnement (quand le contraste croît à peine) accroît l'amplitude des modes qui ont commencé à osciller durant cette ère, c'est-à-dire les modes  $k > k_{boost} \sim k_0 \gtrsim k_A$ . Avancer la transition revient à augmenter  $k_{boost}$ , et donc seuls des modes à plus grand  $k$  sont amplifiés. L'effet net est de globalement *décroître* l'amplitude des pics quand la quantité de matière augmente (en plus du petit changement d'échelle dû à la variation de  $\ell_A$  mentionnée précédemment). L'effet sur les pics du spectre de la matière dans son ensemble est donc assez différent de celui de la seule composante baryonique. La forme du spectre permet donc de séparer la contribution des deux types de matière. Cela suggère que la dégénérescence de l'effet de ces trois paramètres ( $\Omega_{tot,0}, \Omega_{m0}, \Omega_{b0}$ ) peut être levée avec des mesures suffisamment précises des anisotropies.

Comme le fluide oscille, il y a aussi un *effet Doppler* dans la direction  $\vec{k}$  (ligne pointillée bleue de la figure 7.21a), qui vaut zéro aux pics acoustiques et qui est maximal entre eux. Cet effet s'ajoute en quadrature à l'effet Sachs-Wolf. En effet, si l'on imagine une onde acoustique avec  $\vec{k}$  perpendiculaire à la ligne de visée, il n'y a pas d'effet Doppler, tandis que pour  $\vec{k}$  parallèle à la ligne de visée, l'effet Doppler est alors maximal et l'effet Sachs-Wolf est nul. Ce déphasage lisse la structure en pics et creux, mais pas complètement, car l'effet Doppler est un peu plus faible que l'effet Sachs-Wolf (d'un facteur  $\propto \Omega_B h^2$ ). Une augmentation de l'abondance baryonique augmente donc aussi le contraste pic/creux (en plus du contraste entre pics pairs et impairs).

Nous avons jusqu'à maintenant considéré le fluide comme parfait et la transition vers la transparence comme instantanée ; ni l'un ni l'autre ne sont complètement vrais. Les photons diffusés par les électrons exécutent une marche aléatoire de type brownien et peuvent parcourir une distance proportionnelle à la racine carrée du nombre d'interactions, et donc du temps (en coordonnées comobiles, ce qui prend en compte l'effet de l'expansion). Comme les photons sont beaucoup plus nombreux que les électrons (d'un facteur de plusieurs milliards), ils entraînent les électrons dans ce mouvement de diffusion (ce qui entraîne en cascade les protons). Il s'ensuit que toutes les fluctuations plus petites que l'échelle de diffusion des photons sont sévèrement amorties. Cet effet, appelé l'amortissement Silk, est accru par l'augmentation rapide de l'échelle de diffusion durant la combinaison rapide (mais non instantanée) des électrons et des protons qui conduit à la transparence. La finitude de l'épaisseur de la SDD et l'imperfection du fluide font qu'il existe une *coupure exponentielle de la partie à grand- $\ell$*  du spectre de puissance angulaire. Résultat : il ne reste pas grand chose de primordial à observer aux échelles plus petites que  $\sim 5'$  ( $\ell \gtrsim 2000$ ).

Après la combinaison, les photons doivent se mouvoir au sein des grandes structures en développement afin d'atteindre l'observateur. Ils peuvent perdre de l'énergie en ayant à s'extraire de puits de potentiel plus profonds qu'ils ne l'étaient quand ils y sont tombés. Cet effet dépend donc de la rapidité de développement des structures, qui est elle-même fonction du recensement

cosmologique. La réciproque est vraie aussi, bien sûr, et les photons peuvent tirer de l'énergie de vides en formation. Ces deux effets ont tendance à se moyennner à zéro, à petite échelle, puisque l'observateur ne voit que l'effet globalement intégré le long de la ligne de visée. La ligne rouge pointillée de la figure 7.21 indique la forme typique de cette contribution *Sachs-Wolf Intégrée* (SWI). L'effet SWI est anti-corrélé avec l'effet Sachs-Wolf, si bien que le spectre de puissance total,  $C(\ell)$ , est en fait un peu plus petit que le spectre de chacune des composantes prise séparément.

D'autres fluctuations secondaires peuvent aussi laisser leur empreinte, comme par exemple l'effet de lentille des grandes structures qui lisse un peu le spectre. Mais cela aussi peut se prédire précisément. Et le fait que le noyau de lissage soit modifié selon la valeur des paramètres cosmologiques introduit des petits effets qui peuvent aider à lever les dégénérescences résiduelles entre différents paramètres cosmologiques. Les autres effets secondaires qui s'impriment après la combinaison sont en général beaucoup plus faibles (aux échelles  $> 5'$ ). Excepté les fluctuations des diverses émissions astrophysiques d'avant-plan, comme celles de notre Galaxie. On peut néanmoins utiliser des observations multi-fréquences pour les séparer plutôt bien.

Jusqu'à présent, nous n'avons considéré que des perturbations scalaires. Des perturbations primordiales vectorielles ont tendance à s'amortir avec l'expansion. Elles ne créent pas de fluctuations de température dans le cas standard de fluctuations engendrées une fois pour toute dans l'Univers très primitif<sup>32</sup>. Il est néanmoins attendu que le processus même qui a engendré les fluctuations primordiales scalaires ait aussi créé un fond stochastique d'ondes gravitationnelles. Cette contribution (faible) de nature tensorielle ne contribue de façon significative que dans la partie à bas- $\ell$  du spectre, avant le premier pic (car l'amplitude de ces sources diminue dès que leur longueur d'onde devient plus petite que le rayon de Hubble dont le multipôle correspondant est proche). Cette contribution est souvent ignorée au niveau de précision des expériences actuelles. Mais elle est beaucoup plus conséquente pour la partie polarisée de l'émission. La théorie prévoit en effet que les anisotropies du rayonnement fossile doivent être légèrement polarisées.

Une diffusion Thomson peut créer de la polarisation, si la radiation incidente n'est pas isotrope ; ceci peut se produire s'il existe des gradients de vitesse dans le fluide. Avant la combinaison, les diffusions successives sont incohérentes. On s'attend donc à ce qu'une fraction de polarisation notable ne soit créée qu'au moment de la combinaison. Cette polarisation est partiellement corrélée avec la température dans la même direction. En effet les variations de température sont partiellement issues de l'effet Doppler dû aux vitesses dans le fluide. Comme l'effet des vitesses est maximal entre les pics du spectre de température, on s'attend à ce que la structure en pics du spectre de la polarisation soit décalée. Le degré de polarisation prédit est de l'ordre de

---

<sup>32</sup>Ce n'est en revanche plus le cas dans les scénarios où l'on considère des sources actives de fluctuations agissant continuellement au cours de l'Histoire, comme dans le cas où il existerait des cordes cosmiques.

10 %. Des mesures de polarisation du RCF peuvent donc vérifier la cohérence interne des mesures, et aider à résoudre certaines dégénérescences. Mais le plus important, à terme, est que des mesures de polarisation nous fournissent probablement la meilleure façon de détecter le fond d'ondes gravitationnelles primordiales.

Nous nous sommes cantonnés jusqu'ici au cas de perturbations adiabatiques, qui se développent sous l'influence de leur gravité propre, quand elles sont plus grandes que le rayon de Hubble. Une fois entrées dans le rayon de Hubble, la composante photonique (et ses baryons) oscille jusqu'à la combinaison. Toutes les perturbations ont été supposées en phase, comme si elles avaient toutes été engendrées quasi simultanément ; ce qui conduit aux pics « acoustiques » des spectres de puissance résultants. Cette condition est effectivement satisfaite dans la théorie de l'inflation. Dans le cas où il existe des sources actives de fluctuations, comme dans les théories qui incluent des « défauts » topologiques, des perturbations sont engendrées tout le temps avec des propriétés bien différentes<sup>33</sup>. Un seul pic large est alors attendu, ce qui n'est pas observé. C'est pourquoi les données actuelles démontrent que ces défauts ne peuvent pas engendrer à eux seuls la croissance des grandes structures de l'Univers.

*En résumé, les fluctuations progénitrices des grandes structures doivent avoir laissé une empreinte sur le rayonnement cosmologique de fond ; les propriétés statistiques de cette empreinte peuvent être prédites en détail, en fonction des propriétés des perturbations primordiales et de celles de l'Univers homogène. Réciproquement, la mesure de ces anisotropies est une sonde précise et directe de ces propriétés cosmologiques.*

#### 7.4.4 Observations des anisotropies du RCF

Les sites les plus divers ont été utilisés pour effectuer des observations du RCF ; que ce soit sur le toit du laboratoire à Princeton, dans les Alpes, au pôle Sud, en ballon stratosphérique, ou à bord de satellites. Les critères décisifs pour juger de la qualité d'un site sont l'absorption atmosphérique autour de 3 millimètres (100 GHz) et les émissions de lumière parasite. L'atmosphère terrestre est une source intense aux longueurs d'onde proches du maximum du corps noir à 2,7 K, et cette source n'est ni uniforme ni même vraiment stationnaire. Même si des mesures à bord de ballons stratosphériques au plafond, vers 40 km, permettent de minimiser cette source de confusion avec de véritables anisotropies astrophysiques, seuls les satellites permettent de s'en affranchir. Un satellite permet aussi de beaucoup s'éloigner de la Terre ce qui offre l'avantage de diminuer la lumière parasite pouvant potentiellement atteindre les détecteurs par une suite de réflexions et réfractions en cascade. Le produit des 250 K de la température de brillance de la Terre par l'angle solide

---

<sup>33</sup>Elles sont en général isocourbes au lieu d'être adiabatiques, non-gaussiennes, et les phases des perturbations à une échelle donnée sont incohérentes, ce qui ne conduit pas à de multiples oscillations comme dans le cas cohérent de l'inflation.



sous lequel elle est vue doit être largement inférieur au produit de la sensibilité instrumentale qu'on vise à atteindre par la taille angulaire du lobe d'observation, idéalement des micro-kelvins dans un faisceau de  $\sim 5$  arc minutes. Au sol, en ballon, ou en orbite basse, la Terre occupe environ la moitié du ciel, alors que le faisceau de mesure en représente plutôt un demi-millionième. Il faudrait alors atteindre un taux de réjection de la lumière hors axe de  $2 \times 10^{14}$ , ce qui serait extrêmement difficile à faire, et encore plus à contrôler par la mesure. Le point de Lagrange L2 est à environ 1,5 million de kilomètres de la Terre, dans la direction opposée au soleil. La Terre ne couvre alors plus que 17 millièmes du ciel, et le facteur de réjection nécessaire s'abaisse en conséquence d'un facteur  $\sim 28\,000$ , tout en continuant à présenter un défi instrumental formidable.

La première détection sans ambiguïté des anisotropies du RCF a été faite en avril 1992 par l'expérience DMR<sup>34</sup> à bord du satellite COBE en orbite terrestre, avec une résolution effective de la carte des anisotropies de 10 degrés et un rapport signal sur bruit par pixel de l'ordre de 1 (cf. Fig. 7.24). Ceci a permis une mesure de l'effet Sachs-Wolf aux grandes échelles (bas- $\ell$ ). L'invariance constatée de  $\ell^2 C(\ell)$  dont témoigne la figure 7.23a indique que la pente logarithmique du spectre de puissance primordiale,  $n$ , ne peut pas être très différente de l'unité. Et la hauteur de ce plateau  $\sim (30 \mu\text{K})^2$  fournit une estimation directe de la normalisation du spectre,  $A$  (en supposant une théorie minimale, sans grande vérification possible des autres prédictions).

Au cours des quatre années suivantes (Fig. 7.23a), un certain nombre d'expériences explorant de petites régions du ciel ont commencé à suggérer une augmentation de puissance autour de l'échelle du degré, c'est-à-dire vers  $\ell \sim 200$ . Comme le montre la figure 7.23b, il y avait dès 1999 une indication claire, corroborée par un grand nombre d'expériences prises collectivement, qu'un premier pic avait été détecté. Mais ni la hauteur ni l'échelle angulaire ne pouvaient être déterminées avec précision, et l'esprit sceptique pouvait encore évoquer des incertitudes de calibration relative et des erreurs systématiques résiduelles. La situation a changé en mai 2000, quand les collaborations BOOMERanG et Maxima ont chacune annoncé une détermination relativement précise du spectre de  $\ell \sim 50$  à  $\ell \gtrsim 600$  (Fig. 7.23c). L'existence d'un pic autour de  $\ell \simeq 220$  est alors devenu incontestable, avec l'implication immédiate que  $\Omega_{\text{tot}}$  doit être proche de un. Ce résultat a eu instantanément une résonance très large, puisqu'il indiquait clairement, après des décennies de recherches intenses, que la géométrie spatiale de l'Univers doit être proche de la géométrie plate (puisque l'incertitude sur les autres paramètres cosmologiques ne permet pas d'éviter cette conclusion, cf. Éq. (7.53)). En conjonction avec le résultat des SNe Ia, il devenait clair que le scénario de concordance  $\Lambda\text{CDM}$  avec  $\Omega_{m0} \simeq 0,3$ ,  $\Omega_{\Lambda0} \simeq 1 - \Omega_{m0}$  doit être pris très au sérieux.

Comme je l'ai indiqué auparavant, le scénario de formation des structures à partir de fluctuations adiabatiques initialement en phase prédit génériquement

<sup>34</sup>Acronyme de *Differential Microwave Radiometer*. L'expérience FIRS a aussi annoncé une détection très peu de temps après.

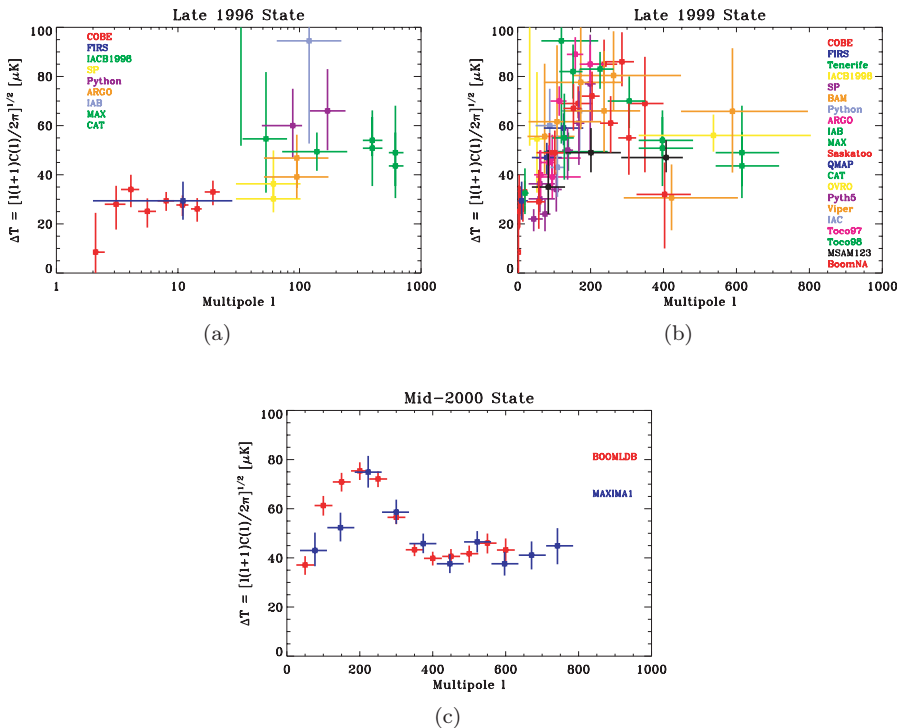


FIG. 7.23 – Mesures successives du spectre de puissance angulaire des anisotropies de température. (a) Disponibles fin 1996; (b) à la fin 1999; (c) à la mi- 2000.

l'existence d'une série de pics acoustiques, dont le contraste relatif donne une indication assez directe de l'abondance baryonique. L'existence d'une coupure exponentielle du spectre à grand- $\ell$  est elle aussi une prédiction générique puisqu'elle traduit l'imperfection du fluide photons + baryons et l'épaisseur finie de la SDD. Cette forme générale du spectre a bien été observée et vérifiée, dès la fin 2002 par l'ensemble des expériences au sol et en ballon, DASI, VSA, CBI, Archeops... Les premiers résultats du satellite américain WMAP (pour Wilkinson Microwave Anisotropy Probe) ont été annoncés début 2003. La plus grande surprise est sans doute qu'il n'y ait pas eu de surprises<sup>35</sup>, en dépit d'une très grande amélioration des données.

Ce satellite a d'abord fourni des cartes de tout le ciel dans 5 bandes de fréquence, allant de 20 à 90 gigahertz. Ces résultats ont exigé un an de prises de données (la mission durera encore au moins trois ans). Pour éviter au maximum le rayonnement parasite du Soleil, le satellite gravite très loin de la

<sup>35</sup>Ce n'est pas tout à fait exact. Les données indiquent une valeur vraiment élevée de la profondeur optique,  $\tau$ . Ce paramètre reflète la façon dont l'Univers est devenu réionisé par le flux des premiers objets non-linéaires.

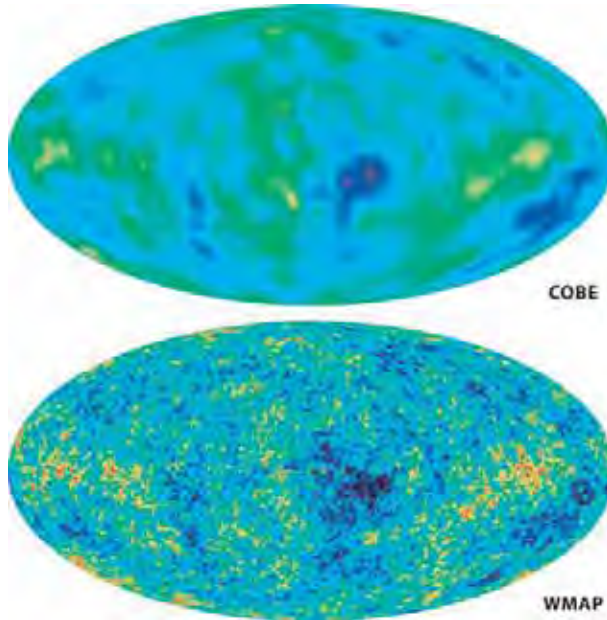


FIG. 7.24 – Carte des anisotropies du RCF sur tout le ciel telles que vues, par COBE puis par WMAP.

Terre autour du point de Lagrange numéro 2. L'équipe de WMAP a ensuite combiné entre elles les 5 cartes, mesurées à différentes fréquences, avec des coefficients choisis pour que les émissions issues de notre Galaxie s'annulent au mieux et ne laissent que le RCF (et du bruit de mesure) dans la carte résultante. Une manière de voir cette opération est d'imaginer que l'on utilise les 4 cartes à plus basse fréquence, là où l'émission Galactique domine le plus par rapport à celle du RCF, pour construire un modèle de l'émission Galactique. On utilise ensuite ce modèle pour « nettoyer » au mieux la carte à plus haute fréquence. La figure 7.24 permet de mesurer le chemin accompli en comparant la carte du RCF produite par COBE/DMR et celle produite par WMAP en utilisant la technique que nous venons de décrire. La résolution de la carte WMAP est environ 40 fois meilleure. La procédure de nettoyage n'est cependant pas parfaite, ne serait-ce qu'à cause de la présence de bruit dans les mesures. Évidemment les résidus éventuels seront d'autant plus importants que le signal à nettoyer est fort. C'est pourquoi les mesures de spectre de puissance (voir Fig. 7.26) n'utilisent pas les zones de la carte les plus sujettes à caution, pour l'essentiel celles qui sont proches du plan de notre Galaxie.

L'amélioration des contraintes posées par l'analyse des anisotropies a été spectaculaire au cours des dernières années. Pour en juger, considérons un modèle cosmologique tout à fait minimal, que 7 paramètres,  $p = (\tau, \Omega_k, \Omega_\Lambda, \omega_b, \omega_c, A_s, n_s)$ , suffisent à spécifier. Le paramètre  $\tau$  indique quand

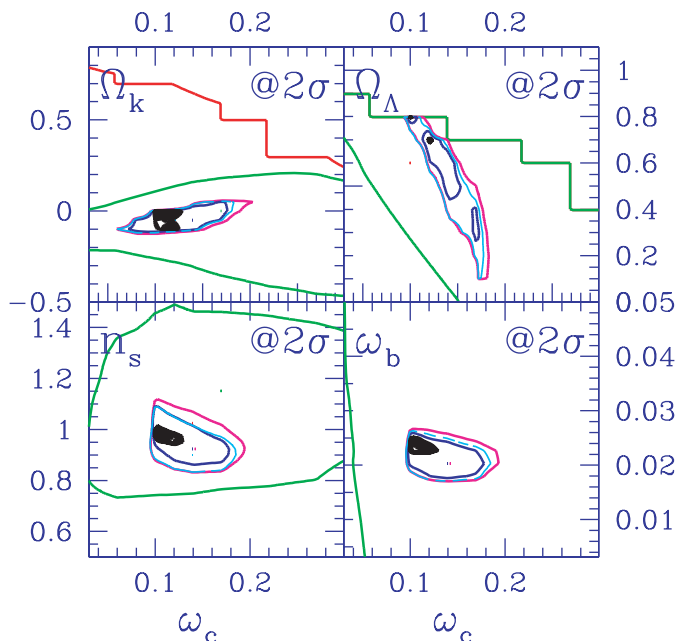


FIG. 7.25 – Contraintes successives, entre 1992 et 2003, apportées par l'analyse des anisotropies pour quelques couples de paramètres.

l'Univers a été réionisé; sa valeur affecte en particulier la coupure exponentielle du spectre. Les 4 paramètres  $(\Omega_k, \Omega_\Lambda, \omega_c, \omega_b)$  déterminent le recensement énergétique et donc le modèle d'univers (on a noté  $\omega_c$  et  $\omega_b$  la densité physique de matière sombre froide et de baryons,  $\omega_c \equiv \Omega_c h^2$  et  $\omega_b \equiv \Omega_b h^2$ , et j'ai omis l'indice 0 pour alléger les notations). On suppose de plus que les fluctuations primordiales sont gaussiennes, scalaires, adiabatiques, sans échelle caractéristique; elles sont donc entièrement décrites par un simple spectre de puissance à deux paramètres,  $P_i(k) = A_s k_s^n$  (j'avais noté la pente logarithmique  $n$  jusqu'à présent). L'ajustement des prédictions de ce modèle aux données permet de vérifier s'il existe une plage de paramètres capable de fournir une bonne description des données. C'est le cas!

La figure 7.25 montre les contraintes successives sur quatre couples de paramètres du modèle qu'on obtient en fonction du jeu de données employé. Les restrictions a priori,  $0,9 > h > 0,45$ ,  $t_0 > 10 \text{ Ga}$ ,  $\Omega_{m0} > 0,1$ , que l'on a imposées constituent la seule information cosmologique issue d'autres analyses que l'on utilise; cela fait longtemps que la communauté cosmologique considère que ces bornes sont assez solidement établies, par de multiples observations. Le contour en rouge indique que la seule utilisation de COBE ne permet que de contraindre faiblement<sup>36</sup>  $\Omega_k$ . L'ensemble des données collectées

<sup>36</sup>Mais COBE a en revanche mieux contraint l'amplitude  $A_s$  du spectre primordial.

jusqu'en janvier 2000 permettait de poser les limites en vert qui dénotent principalement une courbure relativement faible. Les contours en violet, turquoise et bleu indiquent les contraintes disponibles, respectivement, en janvier 2002, juin 2002 et janvier 2003. La mesure d'autres pics que le premier permet notamment de mesurer l'abondance baryonique (qui est proportionnelle au contraste entre pics pairs et impairs). La zone noire donne la situation de mars 2003, dès que furent connus les premiers résultats du satellite WMAP. Cette figure montre que les anisotropies du RCF permettent donc aujourd'hui de sélectionner à elles seules un modèle de type  $\Lambda$ CDM, à peu près plat ( $\Omega_{k0} \simeq 0$ ) avec  $\Omega_{\Lambda 0} \simeq 0,7$ . Les baryons n'y constituent que 15 % de la matière totale ( $\omega_b/\omega_c \simeq 0,02/0,13 \sim 15$  %), la constante de Hubble vaut de l'ordre de  $70 \text{ km.s}^{-1}.\text{Mpc}^{-1}$  (car  $\Omega_m = (\omega_c + \omega_b)h^{-2}$ , et donc  $h^2 \simeq (0,13 + 0,02)/0,3 = 0,5$ ). Les structures s'y développent à partir de fluctuations adiabatiques ayant un spectre de conditions initiales invariant d'échelle « naturel » (Harrison-Peebles-Zeldovich,  $n_s \simeq 1$ ).

### 7.4.5 Concordance, précision et justesse

Les anisotropies du RCF ont l'avantage d'offrir une comparaison directe des données à la théorie, dans un régime où les fluctuations sont toujours de très faible amplitude. Cela permet de calculer leur évolution avec précision en utilisant des équations linéarisées que l'on sait traiter mathématiquement. En outre, les anisotropies sont uniques au sens où leurs propriétés sont sensibles à pratiquement tous les aspects du modèle cosmologique, une fois atteint un niveau de précision des mesures suffisant. Ces anisotropies ne constituent cependant qu'une partie de l'information cosmogonique disponible. La confrontation des résultats aux autres sources d'information va nous permettre d'une part de tester la cohérence du modèle, et d'autre part de lever les ambiguïtés inhérentes à chacun des ensembles de données pris indépendamment..

Un grand relevé de positions de galaxies comme le SDSS permet notamment de mesurer le spectre de puissance de la distribution de la matière aujourd'hui, à condition que l'on sache relier la répartition des galaxies à celle de la matière sombre sous-jacente. L'utilisation cosmologique du spectre de puissance de la distribution des galaxies est d'autant plus difficile que l'on considère des échelles plus petites qui ont subi une évolution fortement non-linéaire. En revanche, si l'on se restreint aux échelles  $\gtrsim 8 h^{-1} \text{ Mpc}$  (soit  $k \lesssim 0,8 h^{-1} \text{ Mpc}$ ) telles que  $\delta < 1$ , cette difficulté peut être contournée, en supposant simplement qu'il existe un biais constant en fonction de l'échelle, tel que  $P_{gal} = b^2 P_{mat}$ . On sait aujourd'hui que cette approximation est raisonnable, aux grandes échelles actuelles.

La figure 7.26 résume la connaissance des spectres de fluctuations à la mi-2004. Les mesures sont représentées par des croix noires. Il s'agit de valeurs moyennes par intervalles de  $\ell$  ou  $k$ , intervalles dont les tailles sont indiquées par l'extension horizontale des croix ; l'extension verticale de chaque croix montre quant à elle l'incertitude sur la mesure (à  $1\sigma$ ) correspondante. Sur le

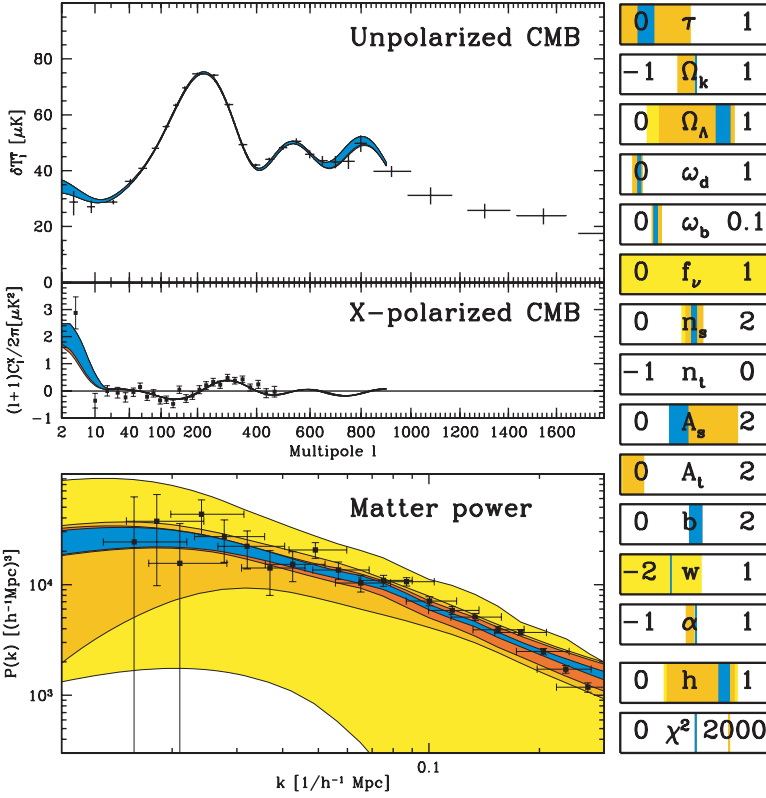


FIG. 7.26 – *Connaissance actuelle des spectres de puissance des anisotropies du RCF (en haut) et des grandes structures (SDSS, en bas). Un modèle plat à sept paramètres ajusté sur les données du RCF suffit à prédire (zone orangé) le spectre des fluctuations de matière observées aujourd’hui autour de nous. L’ajustement simultané de tous les spectres (zone bleu) permet de réduire encore très substantiellement l’incertitude sur certains paramètres cosmologiques. (D’après [111].)*

panneau du haut apparaît le spectre des anisotropies de température mesurées par WMAP, CBI et ACBAR (CBI et ACBAR fixant la partie à  $\ell \geq 600$ ). WMAP a par ailleurs apporté une première mesure du spectre  $C_\ell^{\chi}$  de la corrélation entre température et polarisation. Ce spectre est tracé dans le panneau du milieu. Le spectre de puissance des grandes structures (mesuré dans le relevé SDSS) est présenté dans le panneau du bas en dessous. Cette figure permet donc de parcourir, d’un seul coup d’œil, une partie importante de l’information statistique dont nous disposons aujourd’hui pour sélectionner un modèle d’univers, et en vérifier la cohérence interne et externe.

Considérons un modèle cosmologique à 12 paramètres,

$$p = (\tau, \Omega_k, \Omega_\Lambda, \Omega_d, \Omega_b, f_\nu, w, A_S, n_S, A_T, \alpha, b) \quad (7.54)$$

Par rapport au modèle minimal précédent, on a introduit pour décrire le recensement (1) une équation d'état plus générale,  $p = w\rho$ , pour l'énergie du vide  $\Omega_\Lambda$  (au lieu d'imposer  $w = -1$ ), et (2) la possibilité qu'une fraction  $f_\nu$  de la matière sombre soit « chaude », sous la forme de neutrinos massifs. En ce qui concerne le spectre des fluctuations primordial, en plus de l'amplitude et de la pente du spectre scalaire,  $A_s, n_s$ , on a introduit  $r$  qui dénote l'amplitude<sup>37</sup> de fluctuations tensorielles éventuelles (elles correspondraient à un fond d'ondes gravitationnelles) ainsi qu'un paramètre  $\alpha = d \ln n_s / d \ln k$  qui permettrait de détecter un écart à une loi de puissance du spectre primordial.

La région en jaune de la figure 7.26 indique la zone acceptable pour les valeurs des paramètres (colonne de droite) et pour les spectres qui s'en déduisent, dans le cas du modèle complet ajusté aux seules données du RCF. La zone jaune du panneau du bas constitue donc la prédiction d'où l'on doit trouver le spectre de puissance de la matière aujourd'hui, sur la base de la cosmogonie sélectionnée. La zone orange-clair montre la réduction des incertitudes quand on se restreint aux modèles où  $f_\nu = 0$ ,  $w = -1$ . La zone orange-foncé s'ensuit, si l'on considère de plus que  $\Omega_k = r = \alpha = 0$ . La plage bleue indique enfin la zone correspondant à l'ajustement simultané des spectres d'anisotropies du RCF et de celui du SDSS. Le premier résultat de cet exercice est de retrouver qu'un modèle plat minimal à 7 paramètres, dans lequel l'énergie du vide est indiscernable d'une constante cosmologique ( $w = -1$ ), la contribution des neutrinos et des ondes gravitationnelles est négligeable ( $f_\nu = r = 0$ ), avec un spectre primordial en simple loi de puissance ( $\alpha = 0$ ), décrit très bien les anisotropies du RCF (toutes les plages concernant les anisotropies sont pratiquement superposées). En outre, ce modèle *prédit* très précisément le spectre des grandes structures observé aujourd'hui dans le relevé SDSS ! Ce modèle est précisément le modèle de concordance minimal qui avait en fait été élaboré « à l'envers », en utilisant plutôt des considérations de simplicité et d'élégance théoriques en conjonction avec les observations à  $z = 0$ . Aujourd'hui nous pouvons mesurer directement l'empreinte des fluctuations sur le rayonnement fossile, sélectionner les modèles acceptables et en déduire les résultats attendus à  $z = 0$ .

Comme la prédiction tirée des anisotropies est en excellent accord avec les mesures du SDSS, y compris dans le cas du modèle minimum, il est légitime d'aller plus loin et d'examiner les enseignements qu'on peut tirer d'une analyse conjointe, en considérant une par une quelques extensions par rapport au modèle minimal plat à 7 paramètres<sup>38</sup>,  $p_M = (\tau, \Omega_\Lambda, \omega_b, \omega_c, A_s, n_s, b)$ . Commentons par relaxer la contrainte sur le paramètre de courbure que les anisotropies

<sup>37</sup>La pente  $n_T$  du spectre tensoriel a été supposée obéir à la relation  $n_T = -\tau/8$  qui est une prédiction générique de la plupart des modèles d'inflation (ceux qui satisfont la condition de roulement lent).

<sup>38</sup>Le paramètre de biais et l'épaisseur optique ont tous deux un statut un peu spécial, au sens où ils paramétrisent notre impuissance actuelle à les prédire précisément à partir de la valeur des autres paramètres. Ceci est en principe possible, mais exigerait d'avoir construit la théorie complète de la formation de toutes les structures de l'Univers et de leurs émissions, y compris celles qui ont réionisé le milieu intergalactique.

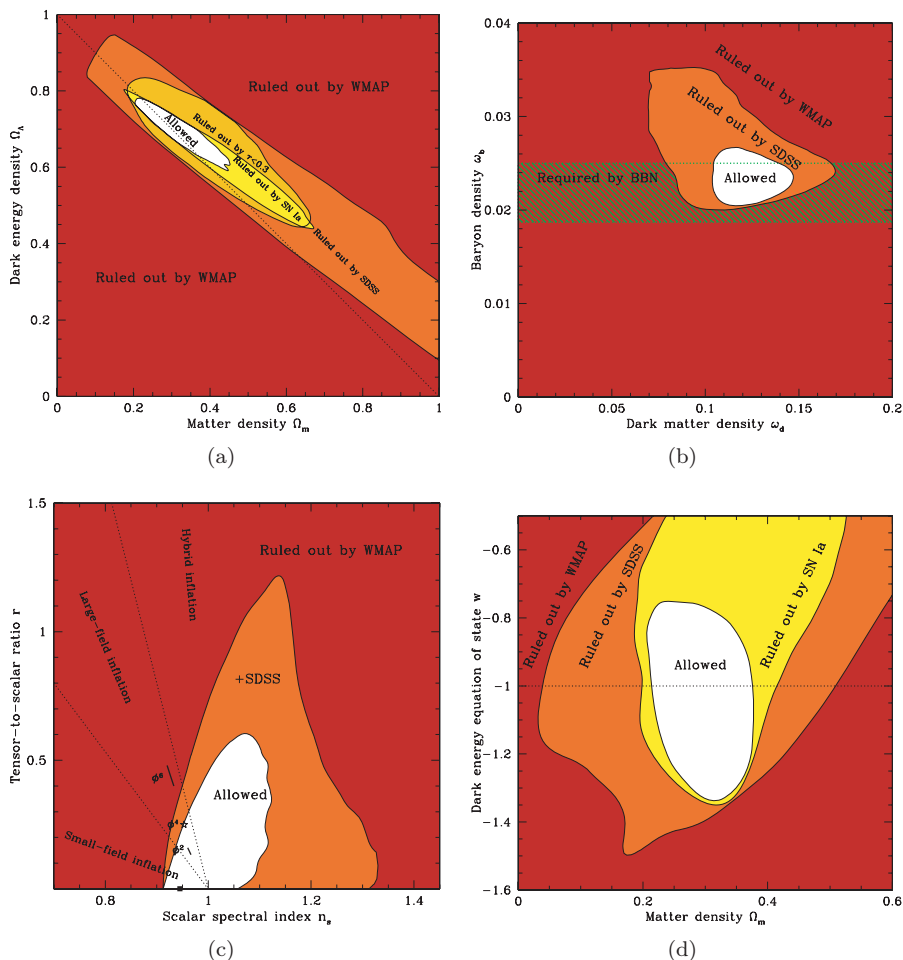


FIG. 7.27 – Exemples de contraintes tirées d'analyses conjointes. (D'après [111].)

sont seules à contraindre directement. La figure 7.27a montre que les données WMAP (en fait WMAP + CBI + ACBAR) excluent clairement (toutes les limites sont à  $2\sigma$ ) un modèle ouvert  $(\Omega_m; \Omega_\Lambda) = (0,3; 0)$ ; elles ne sont pas compatibles non plus avec un modèle Einstein-de Sitter  $(\Omega_m, \Omega_\Lambda) = (1, 0)$ . En fait, ces données sélectionnent une bande légèrement oblique par rapport à la ligne diagonale  $\Omega_m + \Omega_\Lambda = 1$ , sans délimiter à elles seules la part relative entre ces deux quantités. Il existe plusieurs manières de lever cette ambiguïté. On peut par exemple imposer que la constante de Hubble soit entre 58 et 86  $\text{km.s}^{-1}.\text{Mpc}^{-1}$ , l'intervalle à 90 % de confiance qu'indiquent diverses méthodes astrophysiques (reposant sur l'utilisation des céphéïdes). Cela suffirait à sélectionner la zone blanche de la figure. Alternativement, on peut utiliser les données du SDSS, ce qui exclut la zone rouge clair. Les cosmologues



considèrent en général que  $\tau$  ne peut pas être plus grand que 0,3. L'imposer revient à exclure la zone orangée. Enfin, la prise en compte des résultats de l'analyse des SNe Ia restreint à nouveau la zone permise à la zone blanche, centrée autour de  $\Omega_m = 0,3, \Omega_\Lambda = 0,7$ , indépendamment des analyses de  $H_0$ .

Le panneau adjacent (Fig. 7.27b) permet une autre vérification de cohérence ; il compare la contrainte sur l'abondance baryonique qu'indique WMAP seul, ou en conjonction avec le SDSS (quand  $\Omega_{\text{tot}} = 1$ ), et la contrainte issue de la confrontation de l'abondance observée de deutérium<sup>39</sup> avec celle prédite par les calculs de nucléosynthèse primordiale. La physique nucléaire classique appliquée aux premières minutes de modèles de Friedman-Lemaître donne le même résultat que l'analyse des perturbations de la métrique et de leur trace imprimée 380 000 ans plus tard ! La cohérence est maintenue, même si on inclut les contraintes issues de la comparaison des prédictions du développement de ces perturbations aux observations des grandes structures, 13 milliards d'années plus tard.

Les données actuelles ne permettent pas encore d'imposer une limite véritablement contraignante sur l'existence d'un fond d'ondes gravitationnelles primordiales d'amplitude  $r = A_s/A_t$  non nulle. Le panneau 7.27c montre qu'en revanche la pente  $n_s$  du spectre primordial ne peut pas être très différente de l'unité. Si un grand nombre de modèles d'inflation suggère des valeurs de  $n_s$  légèrement inférieures à un (voire des écarts à une pure loi de puissance), les données présentes sont là encore insuffisantes pour différencier les modèles.

Le dernier panneau (Fig. 7.27d) permet enfin de juger des contraintes actuelles sur l'équation d'état de l'énergie noire. Il est clair que les données d'aujourd'hui ne signalent pas de déviation par rapport au cas « classique »  $w = -1$  pour lequel  $\Omega_\Lambda$  peut s'interpréter comme une constante cosmologique. Ces contraintes ne sont pas non plus très serrées.

Mentionnons enfin que l'amplitude des fluctuations primordiales,  $A_s$ , qui est directement mesurée par l'analyse des anisotropies et, indirectement, par le contraste actuel de la distribution des galaxies, peut aussi se contraindre par d'autres tests de l'amplitude des fluctuations de l'univers proche. C'est le cas en particulier de mesures récentes des caractéristiques du champ de cisaillement gravitationnel créé par les grandes structures. Comme les premières détections ne datent que de 2002, cette méthode extrêmement prometteuse ne fournit pas encore de tests serrés du paradigme, mais la situation devrait changer rapidement.

Même si ce survol de résultats cosmogoniques récents est très incomplet, il suffit à démontrer que les modèles de Friedman-Lemaître fournissent un cadre satisfaisant au développement d'une théorie cohérente et prédictive de la formation et de l'évolution des grandes structures de l'Univers. Le scénario a le mérite de la simplicité et de l'économie de moyens : de petites fluctuations

---

<sup>39</sup>Les abondances d'autres éléments légers, comme  ${}^4\text{He}$  ou le  ${}^7\text{Li}$ , ont tendance à indiquer des valeurs plus faibles de l'abondance baryonique (jusqu'à un facteur deux), mais la déduction des abondances primitives à partir des observations est plus difficile que dans le cas du deutérium.

primordiales se sont développées par instabilité gravitationnelle pour former la distribution à grande échelle que nous a révélée la répartition très inhomogène des galaxies. Ce processus est initié par des perturbations scalaires, adiabatiques, de statistique gaussienne, et ayant un spectre invariant d'échelle. Le développement de ces perturbations dans un Univers plat, dont le bilan énergétique est dominé par de l'énergie sombre et de la matière sombre non baryonique, autorise un ajustement précis de données à des échelles spatiales et temporelles multiples. Ces conclusions sont robustes, au sens où l'élargissement de la classe des modèles qu'on ajuste n'a pas permis de trouver une façon satisfaisante de contourner ces conclusions.

## 7.5 Au-delà de la concordance...

Nous avons vu comment, grâce aux tests cosmogoniques, on connaît l'abondance des composantes de chaque type tel que défini par son équation d'état à travers leurs traces sur des échelles critiques. Ces abondances déterminent par exemple le rayon de Hubble au moment de l'égalité des contributions énergétiques de la matière et du rayonnement, et au moment de leur découplage, ce qui détermine les taux de croissance relatifs ou la phase d'oscillation des diverses échelles. Il s'ensuit que les oscillations acoustiques sont un outil pour connaître la géométrie, savoir quelle fraction de la matière est non baryonique, et limiter la masse des particules qui y contribuent. La comparaison avec les fluctuations de la matière de l'Univers local permet d'affiner le diagnostic, en levant les ambiguïtés qui subsistent lorsqu'on analyse uniquement les anisotropies du RCF.

Les mesures d'anisotropies sont relativement uniques parmi l'ensemble des observations astrophysiques utilisées pour contraindre les modèles cosmologiques. Elles sont du même type que des expériences de physique fondamentale ; elles relient des paramètres physiques fondamentaux, qui décrivent notre monde, à des manifestations observationnelles bien spécifiées, qui peuvent être prédites longtemps à l'avance, avec une grande précision. La plupart des prédictions théoriques génériques du modèle ont maintenant été détectées ; qu'il s'agisse du plateau dû à l'effet Sachs-Wolf, de la série de pics à écarts constants qui commence à  $\ell \simeq 220$  (pour un univers plat), de la coupure exponentielle à grand- $\ell$ , tout aussi bien que les premières mesures de polarisation qui se trouvent au niveau attendu, avec la structure en pics décalés prévue. Il reste néanmoins à mesurer en détail les spectres de polarisation, afin d'améliorer encore les contraintes issues de l'analyse des anisotropies.

Quoi qu'il en soit, les paramètres qu'on déduit actuellement de l'analyse des anisotropies sont compatibles avec les contraintes diverses issues d'autres sondes cosmologiques, et il n'y a à présent pas de signe d'incohérences évidentes. Le modèle  $\Lambda$ CDM était déjà en place avant l'obtention des résultats du satellite WMAP ; il aurait aisément pu être infirmé. Sa confirmation n'en est que plus remarquable, d'autant plus que les anisotropies contraignent

simultanément l'ensemble des paramètres. Nous sommes donc finalement entrés dans l'ère où les données empiriques sont telles qu'il est possible de pratiquer des ajustements globaux plutôt que d'en varier seulement un ou deux à la fois en gardant les autres fixés à des valeurs « raisonnables », comme cela a longtemps été le cas. Et comme il y a maintenant plus de contraintes précises qu'il n'y a de paramètres libres pour les ajuster, nous pouvons soumettre nos idées à des tests de cohérence qui sondent de multiples aspects du modèle, y compris la théorie elle-même qui sert de cadre. Les tests de cohérence actuels demandent bien sûr à être largement étendus : ils sont sans doute encore trop limités pour exclure avec certitude que les cohérences ne soient que des coïncidences (ou le résultat de la tentation inconsciente de s'arrêter dans la recherche d'effets trompeurs, une fois acquise la « bonne réponse » selon le paradigme en cours). Il n'en reste pas moins que nous avons appliqué avec un incroyable succès la physique classique dans le cadre cosmologique des modèles de Friedman-Lemaître, au point de laisser penser que nos analyses à  $T < 1$  TeV sont probablement justes, en plus d'être maintenant précises.

À ce jour, le modèle minimal décrit jusqu'alors est suffisant pour interpréter les données dont nous disposons, et la plupart des autres modèles théoriques considérés ont été disqualifiés par les observations. Cependant, certaines possibilités restent très mal contraintes, voire pas du tout. Nous avons vu que les contraintes sur l'équation d'état de l'énergie sombre sont encore faibles. S'il s'avérait que  $w \neq -1$ , nous disposerions alors d'un indice essentiel pour mieux comprendre la physique correspondante, ce qui pourrait peut-être nous aider à enfin résoudre le problème de longue date du niveau minimum d'énergie pour les particules et les champs. Pourquoi l'énergie du vide est-elle à un niveau tel qu'elle domine aujourd'hui le bilan énergétique, tout en étant incroyablement faible par rapport à son échelle naturelle<sup>40</sup> ? Un champ quantique dit de « quintessence » relaxant lentement vers zéro permettrait de fournir une explication dynamique à cette coïncidence. Il s'agit de remplacer le vide par un champ d'ordinaire négligeable, mais dont la densité d'énergie,  $\rho_\Lambda$ , dominerait aux échelles cosmologiques et fournirait une constante cosmologique effective, qui ne pourrait varier que lentement, aussi bien spatialement que temporellement. Les tests cosmologiques des prochaines années offrent une fenêtre sans doute unique pour mettre à l'épreuve l'existence et les propriétés de cette extension au modèle standard de la physique des particules [112].

En ce qui concerne la matière sombre, froide et sans collision, nous disposons d'informations concordantes à de nombreuses échelles différentes – de l'intervalle compris entre 10 000 et 10 Mpc sondé par le RCF à l'intervalle entre 100 et 1 Mpc des grandes structures. Mais pour les structures aux échelles  $< 1$  Mpc, la gravitation n'est qu'un phénomène parmi nombre d'autres qu'il faut inclure (hydrodynamique, processus collisionnels et radiatifs, etc.). Les difficultés rencontrées (nombre de sous-structures et profils de densité près du

---

<sup>40</sup>L'échelle d'énergie de l'énergie noire,  $E_\Lambda = \rho_\Lambda^{1/4}$  est environ 30 ordres de grandeur plus petite que l'énergie de Planck  $E_{Planck} = G^{-1/2} \dots$

centre de structures comme les galaxies et les amas) viennent sans doute de la physique complexe à mettre en œuvre pour en rendre compte. Les problèmes sont donc sans doute de nature astrophysique. À moins qu'ils soient aussi un signe annonciateur de révisions nécessaires par rapport au cas le plus simple. Diverses modifications des propriétés physiques de la matière sombre froide ont d'ores et déjà été proposées. Il s'agit par exemple de modifier le taux d'auto-interaction (pris nul dans le cas standard), ou le temps de vie de la particule constituante (pris infini jusqu'ici), ou la force de son interaction en fonction de l'échelle. Chacune de ces modifications affecte le scénario cosmogonique de manière différente; elles changent l'époque de formation et le nombre de certains objets en fonction de leur masse, ou bien le profil de densité en fonction de l'échelle des structures formées. La cosmologie devrait donc pouvoir contraindre ces possibilités et continuer à nous aider à développer la physique fondamentale.

Les tests cosmogoniques permettent de sonder les mécanismes de création des fluctuations ayant initié la croissance des structures. Les données actuelles indiquent que le modèle naturel est une approximation satisfaisante. Comme ce modèle a été introduit dès les années soixante-dix sur la base d'arguments généraux, ce succès ne peut fournir que peu d'indications sur la physique à l'œuvre. Toute déviation par rapport aux hypothèses de base de fluctuations gaussiennes, scalaires, adiabatiques, invariantes d'échelle donnerait en revanche de précieuses indications pour restreindre les possibilités théoriques. La théorie de l'inflation introduite au début des années quatre-vingt [50–52] offre l'exemple d'une classe de mécanismes satisfaisante. Elle n'a néanmoins pas de justification fondamentale et n'a pas encore fait de prédiction inattendue qui ait été vérifiée par la suite. Heureusement, la théorie montre que les conditions initiales « naturelles » de Harrison-Peebles-Zeldovich ne sont, après tout, pas si naturelles, en ce sens que les fluctuations créées dépendent du détail des champs qui causent l'inflation. En fait des déviations sont attendues, car ce qui est générique dans le cadre de l'inflation, c'est une dépendance logarithmique. Les propriétés des anisotropies du RCF permettent en principe d'évaluer ces déviations à l'invariance d'échelle (pente  $n \neq 1$ ), pour peu qu'on augmente beaucoup la précision des mesures. La théorie a une autre prédiction générique, l'existence d'un fond d'ondes gravitationnelles primordiales que la polarisation des anisotropies du RCF pourrait révéler. Diverses manières de réaliser une phase d'inflation conduisent à des relations différentes entre amplitude du fond gravitationnel et pente du spectre de fluctuations scalaires. L'augmentation de précision des mesures devrait donc favoriser une classe particulière de modèle. La mesure de la pente du spectre de fluctuations tensorielles, encore plus difficile, permettrait quant à elle de vérifier une relation générique des modèles d'inflation. La vérification de cette relation permettrait véritablement d'inclure cette théorie dans le paradigme cosmologique standard.

Cependant d'autres idées commencent aujourd'hui à émerger de la confrontation entre physique fondamentale et cosmologie. Les travaux sur les cordes

fondamentales ont notamment conduit à un renouveau de l'idée de Kaluza de dimensions supplémentaires, dont la présence permet une description unifiée des forces. L'univers usuel serait confiné à un sous-espace ou « brane », l'espace à quatre dimensions ordinaire, tandis que la gravité baignerait l'intégralité des dimensions supplémentaires, certaines pouvant ne pas être compactes. Notre Big Bang pourrait alors être la manifestation de la collision entre deux branes. De nombreux travaux cherchent actuellement à prédire les conséquences cosmologiques de cette idée. L'idée d'univers cyclique est aussi de retour. Dans ce cas, les processus qui ont donné naissance aux grandes structures de l'Univers pourraient s'être déroulés durant une phase de contraction lente avant le Bang, plutôt qu'au cours d'une période d'expansion rapide après le Bang ; la contribution du fond d'ondes gravitationnelles aux anisotropies du RCF devrait dans ce cas être négligeable. Quoiqu'il en soit, il est certain que la cosmologie continuera de jouer un rôle essentiel en physique fondamentale en mettant en œuvre les moyens de l'astronomie.

Au terme de ce survol d'aspects de la cosmologie contemporaine, il est clair que l'héritage d'Einstein dans ce domaine de la physique est toujours autant d'actualité car la relativité générale a permis de fonder et développer la cosmologie moderne. Cette théorie suffit pour élaborer une histoire de l'Univers, hormis celle de ses tout premiers instants, et pour construire une cosmogonie cohérente et prédictive. En ce qui concerne la constante cosmologique, erreur d'Einstein ou pas, on ne sait toujours pas ce qu'il en est. En revanche, il est maintenant bien établi qu'il existe une composante faiblement variable, spatialement et temporellement, qui accélère l'expansion, et qui peut être le reflet de cette constante cosmologique, l'énergie actuelle du vide, ou un nouveau fluide dont les propriétés restent à mieux cerner. Quoiqu'il en soit, la cosmologie construite sur la relativité est devenue un outil pour poursuivre la recherche d'une vision unifiée de la physique. La théorie des cordes fondamentales est l'approche la plus prometteuse dont nous disposons actuellement. C'est sans doute vers elle qu'il faut chercher la justification ultime des propriétés des fluctuations primordiales, et de la physique du secteur sombre, matière et énergie noires, dont nous commençons à tracer les contours.

## Pour aller plus loin

Niveau deuxième cycle : [15, 113, 114].

Livres de référence (en anglais) pour étudiants, par ordre chronologique : [16, 115–117].

## 7.6 Références

- [1] K. Popper, *La logique de la découverte scientifique*, Paris, Payot (pour la traduction française) (1973).

- [2] J.R. Gott, M. Juric, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark, N. Bahcall, J. Brinkmann, *A Map of the Universe*, astro-ph/0310571 (2003).
- [3] E. Hubble, *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*, Proceedings of the National Academy of Science **15**, 168–173 (1929).
- [4] H.S. Leavitt, *Annals of Harvard College Observatory* **173** (1912).
- [5] E. Opik, *An estimate of the distance of the Andromeda Nebula*, Ap. J. **55**, 406–410 (1922).
- [6] E.P. Hubble, *A spiral nebula as a stellar system: Messier 33*, Ap. J. **63**, 236–274 (1926).
- [7] E.P. Hubble, *Cepheids in spiral nebulae*, The Observatory **48**, 139–142 (1925).
- [8] W. de Sitter, *Einstein's theory of gravitation and its astronomical consequences. Third paper*, MNRAS **78**, 3–28 (1917).
- [9] A. Friedman, S.-B. Preuss. Akad. WissZ. Phys. **10**, 377 (1922).
- [10] G. Lemaître, *Ann. Soc. Sci. Bruxelles* **47A**, 49 (1927).
- [11] G. Lemaître, *Expansion of the universe, The expanding universe*, MNRAS **91**, 483–490 (1931).
- [12] H. Bondi, T. Gold, *The Steady-State Theory of the Expanding Universe*, MNRAS **108**, 252 (1948).
- [13] F. Hoyle, *A New Model for the Expanding Universe*, MNRAS **108**, 372 (1948).
- [14] A. Einstein, de Sitter W., *Proc. Natl. Acad. Sci. USA* **213** (1932).
- [15] J.-P. Luminet, *L'invention du Big Bang. Le seuil*, Point Sciences (2004).
- [16] P.J.E. Peebles, *Principles of physical cosmology*, Princeton Series in Physics, Princeton, NJ, Princeton University Press (1993).
- [17] G. Gamow, *The Origin of Elements and the Separation of Galaxies*, Phys. Rev. **74**, 505–506 (1948).
- [18] R.A. Alpher, R.C. Herman, *On the Relative Abundance of the Elements*, Phys. Rev. **74**, 1737–1742 (1948).
- [19] J.C. Mather, E.S. Cheng, R.E. Eplee, R.B. Isaacman, S.S. Meyer, R.A. Shafer, R. Weiss, E.L. Wright, C.L. Bennett, N.W. Boggess, E. Dwek, S. Gulkis, M.G. Hauser, M. Janssen, T. Kelsall, P.M. Lubin, S.H. Moseley, T.L. Murdock, R.F. Silverberg, G.F. Smoot, D.T. Wilkinson, *A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite*, Ap. J. Lett. **354**, L37–L40 (1990).
- [20] R.H. Dicke, P.J.E. Peebles, P.G. Roll, D.T. Wilkinson, *Cosmic Black-Body Radiation*, Ap. J. **142**, 414 (1965).
- [21] D.J. Fixsen, G. Hinshaw, C.L. Bennett, J.C. Mather, *The Spectrum of the Cosmic Microwave Background Anisotropy from the Combined COBE FIRAS and DMR Observations*, Ap. J. **486**, 623 (1997).
- [22] G. Lemaître, *Ann. Soc. Sci. Bruxelles* **53A**, 51 (1933).
- [23] E.M. Lifshitz, *J. Phys. USSR* **10**, 116 (1946).

- [24] I.D. Novikov, Soviet Phys. JETP **19**, 467 (1964).
- [25] E.R. Harrison, *Fluctuations at the Threshold of Classical Cosmology*, Phys. Rev. D **1**, 2726–2730 (1970).
- [26] P.J.E. Peebles, J.T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, Ap. J. **162**, 815 (1970).
- [27] Y.B. Zeldovich, *A hypothesis, unifying the structure and the entropy of the Universe*, MNRAS **160**, 1P (1972).
- [28] F. Zwicky, Helv. Phys. Acta **6**, 124 (1933).
- [29] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, Ap. J. **86**, 217 (1937).
- [30] R. Cowsik, J. McClelland, *Gravity of Neutrinos of Nonzero Mass in Astrophysics*, Ap. J. **180**, 7–10 (1973).
- [31] A.S. Szalay, G. Marx, *Neutrino rest mass from cosmology*, A&A **49**, 437–441 (1976).
- [32] Y.B. Zel'Dovich, *Gravitational instability: an approximate theory for large density perturbations*, A&A **5**, 84–89 (1970).
- [33] V.A. Lubimov, E.G. Novikov, V.Z. Nozik, E.F. Tretyakov, V.S. Kosik, *An estimate of the ( $\nu_e$ ) mass from the  $\beta$ -spectrum of tritium in the valine molecule*, Phys. Lett. B **94**, 266–268 (1980).
- [34] A.A. Klypin, S.F. Shandarin, *Three-dimensional numerical model of the formation of large-scale structure in the Universe*, MNRAS **204**, 891–907 (1983).
- [35] J. Centrella, A.L. Melott, *Three-dimensional simulation of large-scale structure in the universe*, Nature **305**, 196–198 (1983).
- [36] T. Kuhn, *Structure des révolutions scientifiques*, Paris, Flammarion (pour la traduction française) (1972).
- [37] G. Marx, A.S. Szalay, *Three-dimensional simulation of large-scale structure in the universe*, Neutrino '72 Proc, Technoinform, Budapest **1**, 213 (1972).
- [38] H. Pagels, J.R. Primack, *Supersymmetry, cosmology, and new physics at teraelectronvolt energies*, Phys. Rev. Lett. **48**, 223–226 (1982).
- [39] G.R. Blumenthal, H. Pagels, J.R. Primack, *Galaxy formation by dissipationless particles heavier than neutrinos*, Nature **299**, 37 (1982).
- [40] J.R. Bond, A.S. Szalay, M.S. Turner, *Formation of galaxies in a gravitino-dominated universe*, Phys. Rev. Lett. **48**, 1636–1639 (1982).
- [41] P.J.E. Peebles, *The Black-Body Radiation Content of the Universe and the Formation of Galaxies*, Ap. J. **142**, 1317 (1965).
- [42] R.K. Sachs, A.M. Wolfe, *Perturbations of a Cosmological Model and Angular Variations of the Microwave Background*, Ap. J. **147**, 73 (1967).
- [43] J. Silk, *Cosmic Black-Body Radiation and Galaxy Formation*, Ap. J. **151**, 459 (1968).
- [44] M.S. Longair, R.A. Sunyaev, *Fluctuations in the Microwave Background Radiation*, Nature **223**, 719–721 (1969).
- [45] R.A. Sunyaev, Y.B. Zeldovich, *Small-Scale Fluctuations of Relic Radiation*, Astrophys. Space Sci. **7**, 3 (1970).

- [46] J. Silk, M.L. Wilson, *Large-scale anisotropy of the cosmic microwave background radiation*, Ap. J. Lett. **244**, L37–L41 (1981).
- [47] P.J.E. Peebles, *Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations*, Ap. J. Lett. **263**, L1–L5 (1982).
- [48] G.R. Blumenthal, S.M. Faber, J.R. Primack, M.J. Rees, *Formation of galaxies and large-scale structure with cold dark matter*, Nature **311**, 517–525 (1984).
- [49] M. Davis, G. Efstathiou, C.S. Frenk, S.D.M. White, *The evolution of large-scale structure in a universe dominated by cold dark matter*, Ap. J. **292**, 371–394 (1985).
- [50] A.H. Guth, *Inflationary universe: A possible solution to the horizon and flatness problems*, Phys. Rev. D **23**, 347–356 (1981).
- [51] A.D. Linde, *A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems*, Phys. Lett. B **108**, 389–393 (1982).
- [52] A. Albrecht, P.J. Steinhardt, *Cosmology for grand unified theories with radiatively induced symmetry breaking*, Phys. Rev. Lett. **48**, 1220–1223 (1982).
- [53] F.R. Bouchet, *Cosmic strings - an introduction to their formation, evolution, and their microwave background signature*, dans *NATO ASIC Proc. 359: The Infrared and Submillimetre Sky after COBE*, 101–127 (1992).
- [54] P.J.E. Peebles, J. Silk, *A cosmic book*, Nature **335**, 601–606 (1988).
- [55] P.J.E. Peebles, J. Silk, *A cosmic book of phenomena*, Nature **346**, 233–239 (1990).
- [56] V. De Lapparent, M.J. Geller, J.P. Huchra, *A slice of the Universe*, Ap. J. Lett. **302**, L1–L5 (1986).
- [57] E.J. Groth, R. Juszkiewicz, J.P. Ostriker, *An estimate of the velocity correlation tensor - Cosmological implications*, Ap. J. **346**, 558–565 (1989).
- [58] K.M. Gorski, M. Davis, M.A. Strauss, S.D.M. White, A. Yahil, *Cosmological velocity correlations - Observations and model predictions*, Ap. J. **344**, 1–19 (1989).
- [59] G.F. Smoot, C.L. Bennett, A. Kogut, E.L. Wright, J. Aymon, N.W. Boggess, E.S. Cheng, G. de Amici, S. Gulkis, M.G. Hauser, G. Hinshaw, P.D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S.S. Meyer, S.H. Moseley, T. Murdock, L. Rokke, R.F. Silverberg, L. Tenorio, R. Weiss, D.T. Wilkinson, *Structure in the COBE differential microwave radiometer first-year maps*, Ap. J. Lett. **396**, L1–L5 (1992).
- [60] E.L. Wright, S.S. Meyer, C.L. Bennett, N.W. Boggess, E.S. Cheng, M.G. Hauser, A. Kogut, C. Lineweaver, J.C. Mather, G.F. Smoot, R. Weiss, S. Gulkis, G. Hinshaw, M. Janssen, T. Kelsall, P.M. Lubin, S.H. Moseley, T.L. Murdock, R.A. Shafer, R.F. Silverberg, D.T. Wilkinson,



- Interpretation of the cosmic microwave background radiation anisotropy detected by the COBE Differential Microwave Radiometer*, Ap. J. Lett. **396**, L13–L18 (1992).
- [61] J.P. Ostriker, P.J. Steinhardt, *The Observational Case for a Low Density Universe with a Non-Zero Cosmological Constant*, Nature **377**, 600 (1995).
- [62] P.J.E. Peebles, *Tests of cosmological models constrained by inflation*, Ap. J. **284**, 439–444 (1984).
- [63] A.G. Riess, A.V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P.M. Garnavich, R.L. Gilliland, C.J. Hogan, S. Jha, R.P. Kirshner, B. Leibundgut, M.M. Phillips, D. Reiss, B.P. Schmidt, R.A. Schommer, R.C. Smith, J. Spyromilio, C. Stubbs, N.B. Suntzeff, J. Tonry, *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, Astron. J. **116**, 1009–1038 (1998).
- [64] S. Perlmutter, G. Aldering, G. Goldhaber, R.A. Knop, P. Nugent, P.G. Castro, S. Deustua, S. Fabbro, A. Goobar, D.E. Groom, I.M. Hook, A.G. Kim, M.Y. Kim, J.C. Lee, N.J. Nunes, R. Pain, C.R. Pennypacker, R. Quimby, C. Lidman, R.S. Ellis, M. Irwin, R.G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B.J. Boyle, A.V. Filippenko, T. Matheson, A.S. Fruchter, N. Panagia, H.J.M. Newberg, W.J. Couch, *The Supernova Cosmology Project, Measurements of Omega and Lambda from 42 High-Redshift Supernovae*, Ap. J. **517**, 565–586 (1999).
- [65] A.E. Lange, P.A. Ade, J.J. Bock, J.R. Bond, J. Borrill, A. Boscaleri, K. Coble, B.P. Crill, P. De Bernardis, P. Farese, P. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V.V. Hristov, A. Iacoangeli, A.H. Jaffe, L. Martinis, S. Masi, P.D. Mauskopf, A. Melchiorri, T. Montroy, C.B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J.E. Ruhl, F. Scaramuzzi, D. Sforna, *Cosmological parameters from the first results of Boomerang*, Phys. Rev. D **63**, 42001 (2001).
- [66] A. Balbi, P. Ade, J. Bock, J. Borrill, A. Boscaleri, P. De Bernardis, P.G. Ferreira, S. Hanany, V. Hristov, A.H. Jaffe, A.T. Lee, S. Oh, E. Pascale, B. Rabii, P.L. Richards, G.F. Smoot, R. Stompor, C.D. Winant, J.H.P. Wu, *Constraints on Cosmological Parameters from MAXIMA-1*, Ap. J. Lett. **545**, L1–L4 (2000).
- [67] S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P.G. Ferreira, V.V. Hristov, A.H. Jaffe, A.E. Lange, A.T. Lee, P.D. Mauskopf, C.B. Netterfield, S. Oh, E. Pascale, B. Rabii, P.L. Richards, G.F. Smoot, R. Stompor, C.D. Winant, J.H.P. Wu, *MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of  $10^{-5}$  K*, Ap. J. Lett. **545**, L5–LL9 (2000).
- [68] A. Balbi, P. Ade, J. Bock, J. Borrill, A. Boscaleri, P. De Bernardis, P.G. Ferreira, S. Hanany, V. Hristov, A.H. Jaffe, A.T. Lee, S. Oh, E. Pascale,

- B. Rabii, P.L. Richards, G.F. Smoot, R. Stompor, C.D. Winant, J.H.P. Wu, *Erratum: Constraints on Cosmological Parameters from MAXIMA-1*, *Ap. J. Lett.* **558**, L145–LL145 (2001).
- [69] A.T. Lee, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. De Bernardis, P.G. Ferreira, S. Hanany, V.V. Hristov, A.H. Jaffe, P.D. Mauskopf, C.B. Netterfield, E. Pascale, B. Rabii, P.L. Richards, G.F. Smoot, R. Stompor, C.D. Winant, J.H.P. Wu, *A High Spatial Resolution Analysis of the MAXIMA-1 Cosmic Microwave Background Anisotropy Data*, *Ap. J. Lett.* **561**, L1–LL5 (2001).
- [70] R. Stompor, M. Abroe, P. Ade, A. Balbi, D. Barbosa, J. Bock, J. Borrill, A. Boscaleri, P. De Bernardis, P.G. Ferreira, S. Hanany, V. Hristov, A.H. Jaffe, A.T. Lee, E. Pascale, B. Rabii, P.L. Richards, G.F. Smoot, C.D. Winant, J.H.P. Wu, *Cosmological Implications of the MAXIMA-1 High-Resolution Cosmic Microwave Background Anisotropy Measurement*, *Ap. J. Lett.* **561**, L7–L10 (2001).
- [71] R.A. Knop, G. Aldering, R. Amanullah, P. Astier, G. Blanc, M.S. Burns, A. Conley, S.E. Deustua, M. Doi, R. Ellis, S. Fabbro, G. Folatelli, A.S. Fruchter, G. Garavini, S. Garmond, K. Garton, R. Gibbons, G. Goldhaber, A. Goobar, D.E. Groom, D. Hardin, I. Hook, D.A. Howell, A.G. Kim, B.C. Lee, C. Lidman, J. Mendez, S. Nobili, P.E. Nugent, R. Pain, N. Panagia, C.R. Pennypacker, S. Perlmutter, R. Quimby, J. Raux, N. Regnault, P. Ruiz-Lapuente, G. Sainton, B. Schaefer, K. Schahmaneche, E. Smith, A.L. Spadafora, V. Stanishev, M. Sullivan, N.A. Walton, L. Wang, W.M. Wood-Vasey, N. Yasuda, *New Constraints on  $\Omega_M$ ,  $\Omega_\Lambda$ , and  $w$  from an Independent Set of 11 High-Redshift Supernovae Observed with the Hubble Space Telescope*, *Ap. J.* **598**, 102–137 (2003).
- [72] A.G. Riess, L. Strolger, J. Tonry, S. Casertano, H.C. Ferguson, B. Mobasher, P. Challis, A.V. Filippenko, S. Jha, W. Li, R. Chornock, R.P. Kirshner, B. Leibundgut, M. Dickinson, M. Livio, M. Giavalisco, C.C. Steidel, T. Benítez, Z. Tsvetanov, *Type Ia Supernova Discoveries at  $z > 1$  from the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution*, *Ap. J.* **607**, 665–687 (2004).
- [73] C.L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S.S. Meyer, L. Page, D.N. Spergel, G.S. Tucker, E. Wollack, E.L. Wright, C. Barnes, M.R. Greason, R.S. Hill, E. Komatsu, M.R. Nolte, N. Odegard, H.V. Peiris, L. Verde, J.L. Weiland, *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, *Ap. J. Suppl. Ser.* **148**, 1–27 (2003).
- [74] H.P. Robertson, *Kinematics and World-Structure*, *Ap. J.* **82**, 284 (1935).
- [75] A.D. Walker, *Proc. London Math. Soc.* **42**, 90 (1936).
- [76] G. Lemaitre, *Proc. Nat. Acad. Sci. USA* **20**, 12 (1934).
- [77] M. Fukugita, C.J. Hogan, P.J.E. Peebles, *The Cosmic Baryon Budget*, *Ap. J.* **503**, 518 (1998).

- [78] K.G. Begeman, A.H. Broeils, R.H. Sanders, *Extended rotation curves of spiral galaxies – Dark haloes and modified dynamics*, MNRAS **249**, 523–537 (1991).
- [79] M. Milgrom, *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis*, Ap. J. **270**, 365–370 (1983).
- [80] M. Milgrom, *MOND-theoretical aspects*, New Astron. Rev. **46**, 741–753 (2002).
- [81] J.F. Navarro, C.S. Frenk, S.D.M. White, *The Structure of Cold Dark Matter Halos*, Ap. J. **462**, 563 (1996).
- [82] W.J.G. De Blok, S.S. McGaugh, A. Bosma, V.C. Rubin, *Mass Density Profiles of Low Surface Brightness Galaxies*, Ap. J. Lett. **552**, L23–L26 (2001).
- [83] A.D. Lewis, D.A. Buote, J.T. Stocke, *Chandra Observations of A2029: The Dark Matter Profile Down to below  $0.01r_{vir}$  in an Unusually Relaxed Cluster*, Ap. J. **586**, 135–142 (2003).
- [84] S.W. Allen, S. Ettori, A.C. Fabian, *Chandra measurements of the distribution of mass in the luminous lensing cluster Abell 2390*, MNRAS **324**, 877–890 (2001).
- [85] S.W. Allen, R.W. Schmidt, A.C. Fabian, *Chandra observations of RX J1347.5-1145: the distribution of mass in the most X-ray-luminous galaxy cluster known*, MNRAS **335**, 256–266 (2002).
- [86] Y.B. Zeldovich, R.A. Sunyaev, *The Interaction of Matter and Radiation in a Hot-Model Universe*, Astrophys. Space Sci. **4**, 301 (1969).
- [87] S.D.M. White, J.F. Navarro, A.E. Evrard, C.S. Frenk, *The Baryon Content of Galaxy Clusters – a Challenge to Cosmological Orthodoxy*, Nature **366**, 429 (1993).
- [88] F. Zwicky, *Nebulae as Gravitational Lenses*, Phys. Rev. **51**, 290–290 (1937).
- [89] A. Einstein, *Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field*, Science **84**, 506–507 (1936).
- [90] J.-P. Kneib, R.S. Ellis, I. Smail, W.J. Couch, R.M. Sharples, *Hubble Space Telescope Observations of the Lensing Cluster Abell 2218*, Ap. J. **471**, 643 (1996).
- [91] D. Walsh, R.F. Carswell, R.J. Weymann, *0957 + 561 A, B – Twin quasar objects or gravitational lens*, Nature **279**, 381–384 (1979).
- [92] R. Lynds, V. Petrosian, *Giant Luminous Arcs in Galaxy Clusters*, BAAS **18**, 1014 (1986).
- [93] G. Soucail, Y. Mellier, B. Fort, G. Mathez, F. Hammer, *Further data on the blue ring-like structure in A 370*, A&A **184**, L7–L9 (1987).
- [94] J.N. Hewitt, E.L. Turner, D.P. Schneider, B.F. Burke, G.I. Langston, *Unusual radio source MG1131+0456 – A possible Einstein ring*, Nature **333**, 537–540 (1988).
- [95] J.A. Tyson, G.P. Kochanski, I.P. dell’Antonio, *Detailed Mass Map of CL 0024+1654 from Strong Lensing*, Ap. J. Lett. **498**, L107 (1998).

- [96] F. Valdes, J.F. Jarvis, J.A. Tyson, *Alignment of faint galaxy images – Cosmological distortion and rotation*, Ap. J. **271**, 431–441 (1983).
- [97] J.A. Tyson, F. Valdes, J.F. Jarvis, A.P. Mills, *Galaxy mass distribution from gravitational light deflection*, Ap. J. Lett. **281**, L59–L62 (1984).
- [98] J.A. Tyson, R.A. Wenk, F. Valdes, *Detection of systematic gravitational lens galaxy image alignments – Mapping dark matter in galaxy clusters*, Ap. J. Lett. **349**, L1–L4 (1990).
- [99] H. Bonnet, Y. Mellier, B. Fort, *First detection of a gravitational weak shear at the periphery of CL 0024+1654*, Ap. J. Lett. **427**, L83–L86 (1994).
- [100] N. Kaiser, G. Squires, *Mapping the dark matter with weak gravitational lensing*, Ap. J. **404**, 441–450 (1993).
- [101] G. Squires, N. Kaiser, G. Fahlman, A. Babul, D. Woods, *A Weak Gravitational Lensing Analysis of Abell 2390*, Ap. J. **469**, 73 (1996).
- [102] M. Pierre, J.F. Le Borgne, G. Soucail, J.P. Kneib, *X-ray analysis and matter distribution in the lens-cluster Abell 2390*, A&A **311**, 413–424 (1996).
- [103] R.H. Sanders, *Observational Cosmology*, Second Aegean Summer School on the Early Universe (2004).
- [104] C.T. Kowal, *Absolute magnitudes of supernovae*, Astron. J. **73**, 1021–1024 (1968).
- [105] D. Branch, G.A. Tammann, *Type IA supernovae as standard candles*, Ann. Rev. Astr. Astrophys. **30**, 359–389 (1992).
- [106] B. Leibundgut, *Cosmological Implications from Observations of Type Ia Supernovae*, Ann. Rev. Astr. Astrophys. **39**, 67–98 (2001).
- [107] S.J. Maddox, G. Efstathiou, W.J. Sutherland, J. Loveday, *Galaxy correlations on large scales*, MNRAS **242**, 43P–47P (1990).
- [108] F.R. Bouchet, L. Hernquist, *Gravity and count probabilities in an expanding universe*, Ap. J. **400**, 25–40 (1992).
- [109] C. Lineweaver, *Inflation and the Cosmic Microwave Background*, Proceedings of the New Cosmology Summer School, Colless M. ed., World Scientific & astro-ph/0302213 (2003).
- [110] W. Hu, *CMB temperature and polarization anisotropy fundamentals*, Annals of Physics **303**, 203–225 (2003).
- [111] M. Tegmark, M.A. Strauss, M.R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D.H. Weinberg, I. Zehavi, N.A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M.S. Vogeley, A. Berlind, T. Budavari, A. Connolly, D.J. Eisenstein, D. Finkbeiner, J.A. Frieman, J.E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R.C. Nichol, J.P. Ostriker, A. Pope, R. Scranton, U. Seljak, R.K. Sheth, A. Stebbins, A.S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F.J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessy, D.W. Hogg, Ž. Ivezić, G.R. Knapp, D.Q. Lamb, B.C. Lee, R.H. Lupton, T.A. McKay, P. Kunszt, J.A. Munn, L. O’Connell, J. Peoples, J.R. Pier, M. Richmond, C. Rockosi, D.P.

- Schneider, C. Stoughton, D.L. Tucker, D.E. vanden Berk, B. Yanny, D.G. York, *Cosmological parameters from SDSS and WMAP*, Phys. Rev. D **69**(10), 103501 (2004).
- [112] P.J. Peebles, B. Ratra, *The cosmological constant and dark energy*, Rev. Modern Phys. **75**, 559–606 (2003).
- [113] M. Lachièze-Rey, *Initiation à la cosmologie*. Dunod (2003).
- [114] P. Peter, A. Gangui, *Cordes cosmiques et autres trous de l'espace-temps*, CNRS Editions (2003).
- [115] E.W. Kolb, M.S. Turner, *The early universe*, Frontiers in Physics, Reading, MA: Addison-Wesley (1988, 1990).
- [116] T. Padmanabhan, *Structure formation in the universe*, Cambridge, UK, Cambridge University Press (1993).
- [117] J.A. Peacock, *Cosmological physics*, Cambridge, UK, Cambridge University Press (1999).