La gravitation quantique

Introduction :

La conférence "Strings 2000" tenue en juillet 2000 à Ann Arbor (USA-Michigan) s'est close par une liste des 10 problèmes non-résolus les plus importants en physique fondamentale:

1. Les paramètres fondamentaux de la physique sont-ils essentiels ou contingents?
2. La gravité quantique explique-t-elle l'origine de l'univers?
3. Quelle est la durée de vie du proton?
4. La nature est-elle supersymétrique?
5. Pourquoi l'univers a-t-il une dimension temporelle et trois d'espace?
6. Quelle est la valeur de la constante cosmologique?
7. Quels sont les degrés de liberté de la M-théorie?
8. Comment résoudre le paradoxe de la perte d'information dans les trous noirs?
9. Pourquoi l'échelle gravitationelle est-elle très différente de celle des autres interactions?
10. Comment comprendre quantitativement le confinement des quarks et des gluons?

La physique fondamentale se limite-t-elle vraiment aux supercordes? Witten présente souvent les supercordes comme **la** théorie unifiée définitive. Jusqu'où a-t-il raison? En fait, deux voies principales sont suivies depuis plus d'un demi-siècle pour quantifier la gravitation, l'approche "**canonique**" et l'approche "**covariante**". La première considère la gravitation d'abord comme une théorie de l'espace-temps, et elle cherche donc à quantifier l'espace-temps. La seconde considère la gravitation d'abord comme une force s'exerçant entre des objets, et elle cherche donc à quantifier cette interaction en terme d'échange de particules, les gravitons, dans un espace-temps extérieur. Ces deux approches ont connu une histoire tourmentée et difficile, mais elles semblent bien arriver près de leurs buts respectifs. Leur aboutissement présent est représenté par le formalisme des boucles pour l'approche canonique, et par celui des cordes pour l'approche covariante.

Pourquoi vouloir quantifier la gravitation ?

On pourrait imaginer que la gravitation fournit (dans le cadre de la relativité générale d'Einstein par exemple) le cadre spatio-temporel *classique* dans lequel se dérouleraient les phénomènes *quantiques*, mais sur lequel ils n'influeraient pas. La matière et le rayonnement seraient alors décrits par une théorie quantique des champs (comme l'électrodynamique quantique) mais leur effet sur la courbure de l'espace-temps n'interviendrait qu'au travers des valeurs propres (classiques) des opérateurs (quantiques) les décrivant. L'équation d'Einstein de la relativité générale, qui a la forme classique:

$$G_{\mu\nu} = T_{\mu\nu}$$

reliant le tenseur d'Einstein $G_{\mu\nu}$ (décrivant la courbure de l'espace-temps) au tenseur énergie-impulsion $T_{\mu\nu}$ de la matière prendrait alors la forme:

$$G_{\mu\nu} = \langle T_{\mu\nu} \rangle$$

Cette approche présente malheureusement de graves inconvénients:

- la gravitation manifeste ainsi une différence irréductible avec les autres interactions (électromagnétisme ou interactions nucléaires);
- les champs de matière ne sont pas complètement déterminés par cette équation, puisque seules les valeurs moyennes sont contraintes et non leurs fonctions d'onde complètes;
- la possibilité (théorique) est ouverte d'utiliser la gravitation pour violer les axiomes de la mécanique quantique (comme les inégalités de Heisenberg).

Il aurait été surprenant de toute manière que les deux théories physiques fondamentales dont nous disposons pour décrire la nature, la relativité générale et la théorie quantique, restent cloisonnées et ne s'unissent pas malgré leurs fondations très différentes. La mécanique quantique est fondée sur des quantités non commutatives tandis que les coordonnées d'espace-temps sont commutatives en relativité générale (d'où les efforts pour construire des géométries non commutatives).

Il est permis d'espérer qu'une telle fusion soit aussi fructueuse que la fusion de l'électrodynamique et de la mécanique newtonienne qui a donné la relativité, ou celle de l'électrodynamique et de la thermodynamique qui a donné la théorie quantique.

D'autre part, les théories quantiques des champs conduisent le plus souvent à des infinis ("divergences").

En électrodynamique quantique (QED), par exemple, pour aboutir à des quantités finies pour des observables comme la masse ou la charge de l'électron, on doit utiliser l'artifice de la **renormalisation** qui revient grossièrement à supposer ces quantités infiniment négatives au départ, avant de leur ajouter une quantité infiniment positive venant de la contribution des interactions électromagnétiques à cette masse et à cette charge. Le procédé est fructueux dans le cas de la QED parce que toutes les autres quantités que l'on peut calculer dans cette théorie sont ensuite finies (et en parfait accord avec toutes la observations!): on dit que la QED est une théorie renormalisable. Mais il existe des théories des champs non-renormalisables. Dans leur cas, il n'est pas possible de se débarrasser des infinis par un nombre fini d'ajustements comme dans le cas de la QED, et la théorie perturbative perd tout caractère prédictif.

La cause profonde de la plupart de ces infinis est le caractère continu de l'espace-temps: les échanges de particules pouvant se faire sur des distances **x** aussi petites que l'on veut, ces particules peuvent transporter des énergies et des impulsions **p** aussi grandes que l'on veut (en application de l'inégalité de Heisenberg de la mécanique quantique $\Delta\mathbf{x}.\Delta\mathbf{p}>h$). Si l'espace-temps n'était pas continu à très courte distance, ou si l'interaction n'était pas ponctuelle, ces divergences disparaîtraient.

Allant plus loin, certains ont avancé l'idée de supprimer la notion d'un espace-temps sous jacent aux phénomènes et leur servant de cadre de référence, et de ne retenir que les relations entre objets, comme l'ont tour à tout préconisé Leibniz, Mach ou Penrose. L'objectif est alors de construire une physique sans espace-temps, ou plus exactement dans laquelle l'espace-temps n'apparaît que comme une limite de la théorie à grande échelle.

Enfin, la cosmologie fondée sur la relativité générale conduit à la théorie du big bang, dans laquelle l'univers a commencé son existence dans un état singulier -au sens mathématique du terme- où les équations de la relativité générale perdent leur sens. La compréhension de la singularité du big bang ne peut faire l'économie d'une théorie quantique de la gravitation.

Attention: cosmologie quantique et gravitation quantique ne sont nullement synonymes! La cosmologie quantique vise à décrire l'univers dans un cadre quantique par l'intermédiaire d'une "fonction d'onde de l'univers" ce qui n'implique pas -a priori- la quantification de la gravitation. La gravitation quantique vise à décrire les phénomènes quantiques à très petite échelle, ce qui n'implique pas -a priori- l'univers dans son ensemble.

Cependant tous les motifs de fusionner relativité générale et mécanique quantique ne sont que purement théoriques car aucune observation ne conduit à remettre en cause l'une ou l'autre de ces théories. Ceci est bien sûr une conséquence du fait que les phénomènes quantiques sont souvent négligeables à grande échelle spatiale et les phénomènes gravitationnels tout autant négligeables à petite échelle spatiale.

Avant d'aborder la quantification de la gravité, précisons qu'il existe beaucoup de situations où la gravitation ne fournit -en première approximation- qu'un cadre fixe dans lequel se déroulent des phénomènes quantiques. C'est bien sûr le cas de la mécanique quantique ou de la théorie quantique des champs telles qu'elles sont utilisées aussi bien en chimie qu'en physique des particules: un espace-temps statique de Minkowski est présupposé et ne joue pas d'autre rôle que de fixer la structure causale.

On peut écrire une théorie quantique des champs dans un espace-temps courbe classique fixé. En cosmologie, on utilise ainsi un espace-temps doté en première approximation d'une métrique de Robertson-Walker: l'espace-temps est classique et les interactions entre champs de matière sont décrites quantiquement dans cet espace-temps. Cela induit dans les équations d'évolution des termes faisant intervenir l'expansion de l'univers, quand on calcule par exemple les taux de réaction lors de la nucléosynthèse ou les transitions de phase dans l'univers primordial. On peut utiliser des espaces avec des géométries et des courbures plus complexes, sous réserve qu'il reste possible d'établir une foliation globale de l'espace-temps par des hypersurfaces Σ (de genre espace à 3 dimensions), dont la normale définit un temps. La construction de la théorie quantique suit ensuite les mêmes lignes que d'habitudes, ou presque, et on retrouve par exemple une forme généralisée de l'équation de Klein-Gordon. Par contre, il est parfaitement possible que la *même* fonction d'onde, décomposée sur des bases différentes de fonctions propres, n'y ait pas la même décomposition en modes propres: cela signifie par exemple que le nombre de particules observées varie selon l'observateur, pour le même état physique. Unruh a montré en 1976 qu'un observateur accéléré dans le vide observe une émission thermique de rayonnement, le "**rayonnement de Unruh**", dont la température T est proportionnelle à l'accélération γ (l'effet est assez faible: $T \sim 1$ K pour $\gamma = 10^{19} m/s^2$). Cela indique une relation étroite entre accélération, gravitation, thermodynamique et mécanique quantique.

Bekenstein avait d'ailleurs noté en 1973 une analogie formelle entre le second principe de la thermodynamique (l'entropie d'un système isolé ne peut que croître) et le fait que la masse d'un trou noir ne puisse que croître, et il avait montré que l'analogie était quantitative si on identifiait la surface A de l'horizon à une entropie $S=Ac^3/4hG$. Un raisonnement semi-classique conduisit Hawking à montrer en 1975 qu'un trou noir de masse M était entouré d'un rayonnement thermique, le "**rayonnement de Hawking**", de température $T=hc^3/8\pi kGM$.

Retour au sommaire

L'échelle de Planck

A partir des constantes fondamentales de la physique que sont la constante de Planck *réduite* **h** (1.054×10^{-34} J.s), la vitesse **c** de la lumière (299 792 458 m/s) et la constante **G** de Newton de la gravitation (6.673×10^{-11} $m^3/kg/s^2$), il est possible de construire une quantité ayant les dimensions d'une masse, et qu'on appelle la masse de Planck:

$$M_{Planck} = [hc/G]^{1/2} = 2.18 \times 10^{-8} \text{ kg} = 1.22 \times 10^{19} \text{ GeV}/c^2$$

C'est à peu près la masse d'un acarien. On peut de la même manière définir une longueur de Planck $L_{Planck} = [hG/c^3]^{1/2} = 1.61 \times 10^{-35}$ m et un temps de Planck $t_{Planck} = [hG/c^5]^{1/2} = 5.39 \times 10^{-44}$ s.

Quelle sens donner à ces quantités? L'attraction gravitationnelle entre deux électrons est 2.4×10^{-43} fois plus faible que leur répulsion électrostatique. Si les électrons avaient la masse de Planck, leur attraction gravitationnelle serait 137 fois supérieure à leur répulsion électrostatique (137 est juste l'inverse de la "constante de structure fine" qui mesure l'intensité de l'interaction électromagnétique). On peut donc voir dans l'échelle de Planck le point où les interactions gravitationnelles se rapprochent (et peut-être s'unissent) aux autres interactions.

Mais ceci n'est qu'une conséquence du fait que ces autres interactions ont des intensités d'ordre 1 (1/137 pour l'électromagnétisme). L'échelle de Planck est bien plutôt l'échelle où les effets quantiques gravitationnels devraient devenir importants: pour des masses et des énergies (par particule) bien plus faibles, ou des temps et des longueurs bien plus grands, l'*action* d'un système gravitationnel est en effet beaucoup plus grand que le quantum d'action h, et on attend donc un comportement classique (non-quantique) du système en question.

On peut s'en rendre compte si l'on se souvient qu'une échelle de longueur apparaît naturellement en relativité générale, le rayon de Schwarzschild $R_S=GM/c^2$ d'un objet de masse M, échelle où les effets gravitationnels relativistes sont dominants. Une échelle de longueur apparaît d'autre part en mécanique quantique, la longueur d'onde de Compton $L_C=h/Mc$ d'un objet de masse M, longueur où les effets quantiques sont dominants. Un objet pour lequel le rayon de Schwarzschild égale la longueur d'onde de Compton a une masse égale à la masse de Planck (et ces deux longueurs sont alors égales à la longueur de Planck).

Caveat: la notion de masse de Planck n'a ce sens que si la constante G n'est pas renormalisée, si l'espace-temps n'a que 4 dimensions, et si la gravitation est bien telle que définie par la relativité générale (pas de termes en R^2 dans le lagrangien, de supersymétries, etc.).

Retour au sommaire

Approche canonique : du superespace au formalisme des boucles

Dans les années 1960, les théories de jauge étaient loin d'avoir leur prestige actuel car si l'électrodynamique semblait correctement décrite dans ce cadre, ce n'était pas le cas des interactions faible et forte. Le modèle de Fermi de l'interaction faible n'était pas renormalisable et l'interaction forte était trop... forte pour être traitée perturbativement. Les théoriciens exploraient plutôt les voies de l'algèbre des courants, de la matrice S, des pôles de Regge, des modèles duaux de résonances, etc. La quantification de la gravitation intéressait peu les physiciens des particules (à la notable exception de Feynman) et ce sont plutôt les "relativistes" qui s'y sont attaqués. D'où une approche naturellement géométrique (Wheeler la baptisant même "géométrodynamique").

L'approche canonique a connu des débuts difficiles, dûs à la difficulté d'identifier les "bonnes"variables dynamiques pour un système très contraint par les symétries, puis une période d'euphorie après la découverte (ou l'invention) de l'équation de Wheeler-DeWitt, analogue de l'équation de Schrödinger. Mais l'euphorie est retombée quand cette équation s'est révélée presque inexploitable. La découverte par Ashtekar de meilleures variables dynamiques, puis leur réexpression dans le formalisme des "boucles" a ouvert dans les années 1990 une nouvelle période faste aboutissant aux premières indications de quantification de l'espace-temps et aux techniques de "réseaux de spins" et de "mousses de spin". Le premier vrai *calcul* de gravitation quantique vient de retrouver la relation entre entropie et surface d'un trou noir.

Quantification canonique

L'approche canonique part d'une formulation hamiltonienne de la théorie. La première étape est donc d'identifier les variables dynamiques pertinentes (dans le lagrangien **L** par exemple), appelées en général les "coordonnées généralisées" q_i. Les dérivées du lagrangien par rapport aux dérivées par rapport au temps (dq_i/dt) des q_i sont les "moments généralisés" p_i et ils permettent de construire le hamiltonien **H**$=p_i(dq_i/dt)-L$. Deux situations peuvent se présenter:

- Le formalisme possède autant de variables que le système possède de degrés de liberté physiques. Dans ce cas la quantification revient essentiellement à remplacer les q_i par des opérateurs agissant sur les états physiques décrits par une fonction d'onde Ψ. Il en est de même des p_i (dans la "représentation de Schrödinger", p_i->id/dq_i), et les crochets de Poisson deviennent des commutateurs, etc.
- Par contre, si la description choisie possède plus de variables que de degrés de liberté, la situation se complique. Cela signifie en effet que ces variables sont reliées entre elles (par des symétries par exemple). Ces relations $C(p_i)=0$ s'appellent des "contraintes". L'alternative est donc soit de "résoudre" les contraintes en les utilisant pour réduire le nombre de variables avant la quantification, soit de quantifier la théorie d'abord et d'appliquer ensuite les contraintes. A la suite des travaux de Dirac au début des années 1950, la procédure habituelle est de suivre la seconde voie, de quantifier en conservant toutes les variables et de requérir ensuite que les états physiques Ψ soient annihilés par l'opérateur de contrainte: **C**$\Psi=0$

Sous ce rapport, la relativité générale se présente comme une théorie très contrainte en raison de l'invariance par **difféomorphisme** (les quantités physiques sont invariantes par les transformations générales des coordonnées, c'est à dire les fonctions continues de l'espace-temps sur lui-même). Remarquons que cette invariance signifie que les points de l'espace-temps n'ont pas de réalité physique par eux-mêmes. Ces difféomorphismes ressemblent aux transformations de jauge de l'électromagnétisme, où le potentiel-vecteur A_μ est équivalent au potentiel $A_\mu+d_\mu f$ pour toute fonction f, mais avec une différence importante: ces transformations de jauge se passent en un point donné de l'espace-temps, tandis qu'un difféomorphisme envoie un point de l'espace-temps vers un autre. Une quantité simple invariante par difféomorphisme est l'intégrale d'une fonction sur tout l'espace-temps, ce qui est hautement non-local. On voit ici apparaître cette idée que les quantités vraiment physiques sont inévitablement non-locales, que l'on retrouvera plus loin aussi bien dans le formalisme des boucles que dans celui des supercordes.

Résoudre les contraintes correspondant aux difféomorphismes conduit à des systèmes d'équations différentielles non-linéaires couplées, qui se sont révélés insolubles jusqu'à présent. On est donc contraint de quantifier un système contraint et d'imposer ensuite que les états physiques Ψ[géométrie] soient annihilés par les opérateurs des difféomorphismes. Notons au passage que ces états sont des fonctions des fonctions décrivant la géométrie de l'espace-temps, donc des *fonctionnelles*.

L'identification des variables "pertinentes" est loin d'être triviale dans le cas de la gravitation. Il paraît naturel d'utiliser la métrique de l'espace-temps mais la difficulté est d'aboutir à des observables indépendantes du système de coordonnées choisies (Bergmann 1956, Dirac 1959). La décomposition proposée par Arnowitt, Deser et Misner (ADM) en 1962 rend plus transparente l'interprétation géométrique des contraintes: un feuilletage de l'espace-temps par des surfaces Σ en 3 dimensions est choisi et le tenseur métrique $g_{\mu\nu}$ est décomposé en un tenseur h_{ij} codant la métrique des 3-surfaces Σ, et un 3-vecteur et un scalaire. Les contraintes de difféomorphisme se séparent classiquement en deux équations, l'une correspond à l'invariance par difféomorphisme sur les 3-surfaces Σ et l'autre n'est autre que le hamiltonien (correspondant aux translations dans le temps). L'intérêt de la décomposition ADM est que le hamiltonien ne dépend que de la **3-géométrie** décrite par h_{ij} . L'inconvénient est que la covariance *explicite* est perdue.

Equation de Wheeler-DeWitt

En principe, le hamiltonien exprime le contenu dynamique d'une théorie, l'évolution temporelle des systèmes décrits. Mais, ici, il joue aussi le rôle d'une contrainte. Après quantification, l'opérateur correspondant annihile donc les états physiques Ψ. L'équivalent pour la gravitation de l'équation de Schrödinger $H\Psi=id\Psi/dt$ est juste $H\Psi=0$, que l'on appelle l'équation de Wheeler-DeWitt (1967). De façon plus explicite, on a en gravitation pure (sans champs de matière ϕ):

$$H\,\Psi[h] = G^2\,[h_{ik}\,h_{jl} + h_{jk}\,h_{il} - h_{ij}\,h_{kl}]\,(d^2\Psi/dh_{ij}dh_{kl}) + \det(h)\,R_3[h]\,\Psi[h] = 0$$

où R_3 est le scalaire de courbure de la 3-géométrie. L'équation de Wheeler-DeWitt ne dépend que de la 3-géométrie h_{ij} (et -éventuellement- de champs de matière ϕ) **mais non du temps**. Cela a donné lieu à une abondante littérature consacrée à l'interprétation de cette absence du temps. Le temps joue de toute manière un rôle très différent en relativité générale, où il est une variable dynamique, et en mécanique quantique, où il n'est qu'un paramètre extérieur au système étudié et non un opérateur (il n'est donc pas *mesuré*).

Une possibilité "raisonnable" de résoudre la "question du temps" est de considérer qu'une fonction d'onde Ψ solution de l'équation de Wheeler-DeWitt décrit une corrélation entre 3-géométrie et les champs de matière présents. Cela ouvre la possibilité qu'une fonction de ces champs (par exemple leur densité d'énergie) ou une fonction de la métrique (par exemple un paramètre d'échelle) joue le rôle d'un temps effectif. Mais l'accord est très loin d'être général!

De plus, l'équation de Wheeler-DeWitt, comme toute équation différentielle de la physique, peut avoir un très grand nombre de solutions parmi lesquelles il faut sélectionner celle qui correspond à des conditions aux limites données (dont le choix est un problème de physique).Il faut aussi être capable de calculer ces solutions, ce qui est un problème mathématique rarement soluble pour cette équation! Il n'est même pas certain, bien au contraire, que l'équation elle-même soit mathématiquement bien définie. Elle possède en effet un nombre infini de degrés de liberté (la 3-géométrie h_{ij} et les champs de matière ϕ en tout point de l'espace). L'ensemble de ces degrés de liberté a été baptisé superespace par Wheeler, et une grande activité pendant plusieurs dizaines d'années a consisté à tronquer ce superespace en ne gardant qu'un petit nombre (fini) de ces degrés de liberté pour constituer un mini-superespace et y résoudre l'équation de Wheeler-DeWitt.

Notons au passage que l'équation de Wheeler-DeWitt suppose implicitement l'espace(-temps) continu.

Un mini-superespace de Friedmann

Il est possible de se faire une idée des résultats auxquels on peut parvenir par cette approche en considérant un mini-superespace particulièrement simple, celui qui décrit un univers homogène et isotrope. L'évolution d'un tel univers est classiquement décrite par l'équation de Friedmann:

$$(a'/a)^2 = 8\pi G\rho/3 - k/a^2 + \Lambda/3$$

où a est le "paramètre d'échelle" ne dépendant que du temps t, a' sa dérivée par rapport au temps, ρ la densité d'énergie de l'univers, k la constante de courbure spatiale et Λ l'éventuelle constante cosmologique. Dans une vision semi-newtonienne, cette équation peut être interprétée comme T=-U, où T est l'énergie cinétique du "fluide cosmique" et U son énergie potentielle. Elle peut aussi être interprétée comme l'équation du mouvement dérivant du lagrangien:

$$\text{Lagrangien} = - 3\pi a^3/4G\ [(a'/a)^2 + 8\pi G\rho/3 - k/a^2 + \Lambda/3]$$

La variable dynamique est le paramètre d'échelle a et le moment conjugué P (=dLagrangien/da') est $-3\pi aa'/2G$. On retrouve bien l'équation de Friedmann comme équation d'Euler-Lagrange dP/dt= dLagrangien/da. Quel est alors le hamiltonien?

$$H = Pa' - \text{Lagrangien}$$

$$= 3\pi a^3/4G\ [(a'/a)^2 - 8\pi G\rho/3 + k/a^2 - \Lambda/3]$$

$$= 0$$

On constate bien que le hamiltonien est identiquement nul, ce qui est une autre façon de voir l'absence du temps. La quantification va entraîner le remplacement de P par l'opérateur -id/da et l'équation de Wheeler-DeWitt $H\Psi=0$ s'écrit tout simplement:

$$\{-d^2/da^2 + 9\pi^2 /4G^2\ [-8\pi G\rho a^4/3 + ka^2 - \Lambda a^4/3]\ \}\ \Psi = 0$$

qui a exactement la forme de l'équation de Schrödinger à une dimension décrivant le mouvement d'une particule de coordonnée a dans un potentiel V(a).

Pour un univers dénué de matière (ρ=0), l'univers de de Sitter de la cosmologie classique, le potentiel V(a) a la forme simple

$$V(a) = ka^2 - \Lambda a^4/3$$

Si k=-1 ou si k=0, la "particule s'éloigne de a", ce que l'on peut interpréter comme un univers en expansion (rappelons que le paramètre d'échelle a joue en cosmologie le rôle du temps). Si k=+1, il n'y a classiquement de solution que si $a>a_0 = [3/\Lambda]^{1/2}$. Cela correspond classiquement à un univers de de Sitter fermé dont le paramètre d'échelle passe par un minimum pour cette valeur.

Quantiquement, au contraire, une "particule" initialement placée en $a<a_0$, par exemple en a=0, peut émerger de l'autre côté de la barrière de potentiel, en $a=a_0$, avant de s'éloigner vers $a>>a_0$. Cela se manifeste par la création apparemment spontanée d'un univers de paramètre d'échelle a_0 (un univers-bulle, un bébé-univers comme on le dit parfois). Bien entendu, toute solution Ψ de l'équation d'onde dépend des conditions aux limites imposées, ce qui a fait l'objet de débats interminables au fil des décennies.

Avoir négligé la densité de matière ρ ne change pas grand-chose à ce scénario: si cette densité provient essentiellement du rayonnement, comme celui-ci décroît en $1/a^4$, cela revient à ajouter au potentiel V(a) un terme constant, ce qui ne change pas les conclusions. Si cette densité est essentiellement due à la matière non-relativiste qui décroît en $1/a^3$, cela revient à ajouter un terme linéaire à V(a), ce qui change pas non plus qualitativement les conclusions. Quantitativement, seront exponentiellement favorisées par l'effet tunnel les situations minimisant la barrière de potentiel et donc celles où la contribution de la matière est maximale, ce qui favorise les modèles d'inflation (où l'univers passe quelque temps dans un faux minimum supérieur au vrai minimum).

Caveat: on est subrepticement passé de la première à la seconde quantification puisqu'on parle de création d'univers. Il faudrait reprendre toute la discussion en termes d'intégrales de chemin, par exemple, ce qui a occupé nombre de très éminents physiciens, tels Hawking ou Vilenkin, pendant plusieurs *dizaines* d'années, sans résultat bien probant.

Connexions, holonomie et variables d'Ashtekar

Le programme canonique a pris un nouveau départ en 1987 quand Abhay Ashtekar a proposé d'utiliser comme variable dynamique non pas la **métrique** h_{ij} à 3 dimensions mais la **connexion**, comme on le faisait depuis quelques années dans les théories de jauge, ouvrant une voie éventuelle à une unification ultérieure.

C'était d'ailleurs presque un retour aux sources, dans la mesure où les connexions avaient été introduites par Weyl et Poincaré dans les années 1920 pour décrire la relativité générale. A la suite des travaux de Levi-Civita en 1917 définissant le transport parallèle sur une variété, Weyl et Poincaré ont introduit la notion de connexion, qui permet de passer d'un point à un autre d'une variété en suivant un chemin donné. Des chemins différents conduisent à des résultats différents, et par exemple un vecteur pivote en général différemment selon le chemin suivi: le vecteur $\mathbf{V}(\mathbf{x})$ devient $\mathbf{V'}(\mathbf{x'})$. Au premier ordre, $\mathbf{x'}=\mathbf{x}+d\mathbf{x}$ et $\mathbf{V'}=\mathbf{V}+d\mathbf{V}$ et la rotation $d\mathbf{V}$ est proportionnelle à $\mathbf{V}$ et au déplacement $d\mathbf{x}$. En composantes, $dV^i=\Gamma_i{}^j{}_k dx_j V^k$, faisant apparaître la connexion affine $\Gamma_i{}^j{}_k$. La notion de connexion ne présuppose pas l'existence d'une métrique sur la variété, mais elle peut l'intégrer: si une métrique est définie sur la variété, la connexion est équivalente au symbole de Christoffel. Une géodésique est aussi bien une courbe dont les tangentes restent parallèles les unes aux autres (aspect connexion) qu'une courbe de longueur minimale (aspect métrique).

Les connexions jouent un rôle central dans la formulation moderne des **théories de jauge**. Dans ces théories, les champs de matière possèdent un groupe de symétrie continue, et cette invariance locale (imposée en tout point de l'espace-temps) entraîne l'existence d'un champ compensateur, le champ de jauge. Du point de vue de la géométrie différentielle, le champ de jauge est une connexion définie sur la variété (espace ou espace-temps). Concrètement, imaginons une particule décrivant une trajectoire dans l'espace. Son état final *interne* peut différer de son état initial par une phase (ou, plus généralement, par une transformation du groupe de symétrie), et la connexion a précisément pour objet d'établir le lien entre la position spatiale et la phase. Elle mesure le changement de phase d'un point à l'autre de l'espace. On dit qu'on a une structure d'**espace fibré** qui est simplement ici le produit tensoriel de l'espace de base par l'espace interne décrivant la phase (fibre).

Une courbe fermée dans l'espace de base n'est généralement pas une courbe fermée dans le fibré, mais il existe une transformation passant d'une extrémité à l'autre de cette courbe. Les transformations qui relient les extrémités de la courbe sur le fibré forment un groupe, appelé le groupe d'**holonomie** de la connexion. Dans le cas de la particule dont la phase change après un tour complet, ce groupe est simplement le groupe U(1). Dans le cas d'un vecteur sur une variété en *n* dimensions, une boucle fermée le ramène au point de départ mais tourné: le groupe d'holonomie est SO(*n*), le groupe des rotations en *n* dimensions.

Il est important de noter que le changement de phase (au sens large) subi par la particule ne dépend pas seulement des valeurs de la métrique le long du chemin suivi: il dépend des valeurs de la connexion et celles-ci intègrent des informations sur la métrique à l'intérieur de la boucle. En électrodynamique, ceci est illustré par l'effet Aharonov-Bohm, dans lequel un électron peut traverser une zone où le champ électromagnétique (l'équivalent de la métrique) est nul et voir cependant sa phase modifiée si le potentiel électromagnétique (la connexion), lui, n'y est pas nul.

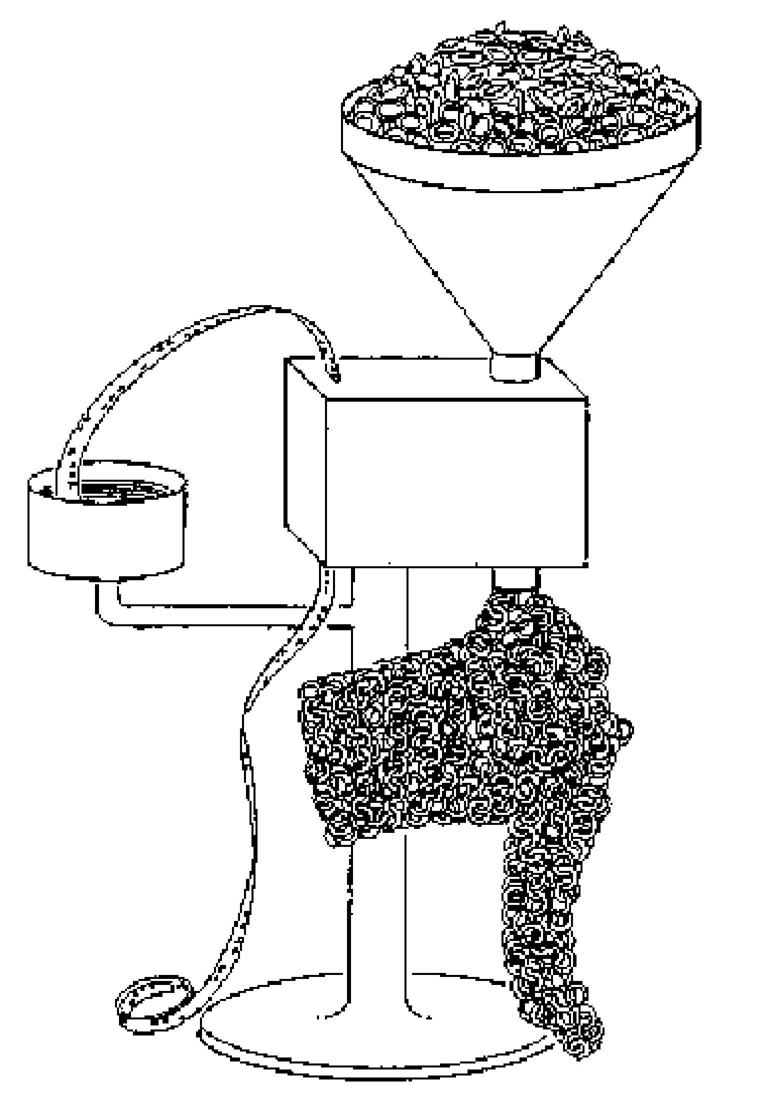
L'idée d'Ashtekar fut donc d'utiliser non la *métrique* de l'espace courbe (la 3-géométrie), mais la *connexion*. En pratique, il n'utilisa pas la connexion affine de Weyl mais une connexion spinorielle (ce qui lui permet de traiter les fermions) dont le groupe d'holonomie est SU(2), le recouvrement universel de SO(3). Cette connexion A_i^a porte un indice d'espace i et un indice "interne" a de SU(2). A ce stade, il ne s'agissait encore que d'une reformulation de la relativité générale *classique*. Mais, pour des raisons qui ne sont pas encore parfaitement élucidées, les variables d'Ashtekar conduisent à une formulation de la quantification plus simple que les approches antérieures. L'idée était que l'espace des fonctions continues Ψ(A) serait l'espace des états quantiques gravitationnels (l'espace de Hilbert de la théorie).

Boucles, noeuds, tissages et broderies

Dès 1988, Smolin et Rovelli montraient qu'il serait encore plus astucieux d'utiliser non la connexion A_i^a directement mais une variable de **boucle** (*loop* variables), c'est-à-dire une fonction *f* de l'élément U(γ,A) du groupe d'holonomie SU(2) correspondant à une boucle fermée γ dans l'espace (en général on prend la trace qui est un invariant, ce qu'on appelle une *boucle de Wilson*). Ils ont montré qu'il existe une sorte de "dualité" entre les deux variables: la donnée de l'ensemble des *f*(γ,A) pour toutes les boucles γ dans l'espace courbe est équivalente à la donnée de la connexion A en tous les points de cet espace. D'ailleurs on peut aussi bien voir en *f*(γ,A) une fonction sur l'espace des boucles qu'une fonction sur l'espace des connexions. Cette dualité signifie intuitivement que la connaissance de la rotation subie par un vecteur (ou un spineur) au terme d'une boucle fermée quelconque permet de reconstituer partout la courbure de l'espace. L'espace des états quantiques gravitationnels est alors l'espace de ces fonctions *f*(γ,A) , les opérateurs agissant sur elles brisant, redirigeant et recollant les boucles. Il n'est pas encore absolument sûr que cet espace soit mathématiquement bien défini (en particulier que les états aient bien une norme finie).

Quel est l'intérêt de passer de la connexion à la boucle? D'abord il s'agit de solutions exactes de l'équation de Wheeler-DeWitt. Et surtout il est beaucoup plus simple dans ce cas de tenir compte de l'invariance de la théorie par difféomorphisme: on passe simplement d'une boucle à la même boucle déformée. Les véritables quantités physiques sont donc en fait les classes d'équivalence des boucles, ce qui correspond à la notion mathématique de **noeud**. Pour le mathématicien, un noeud est une courbe fermée qui ne se recoupe pas, et qui ne dépend ni de la taille ni de la forme ni de l'orientation des brins noués. En 2 dimensions, le seul noeud est le noeud "trivial" tandis qu'en 4 dimensions tout noeud peut se dénouer: la notion n'a donc d'intérêt qu'en 3 dimensions. La grande difficulté de la théorie des noeuds est de classer tous les noeuds possibles en identifiant les noeuds qui ne sont que des déformations les uns des autres (noeuds *isotopes*). Jusqu'à maintenant, toutes les méthodes de classement proposées (nombre minimal de croisements, surface sur laquelle le noeud peut être tracé sans croisement, groupe des classes d'homotopie des chemins autour du noeud, genre, polynômes d'Alexander, de Jones, de HOMFLY, de Kauffman, invariants de Vassiliev) confondent certains noeuds. C'est donc un domaine en pleine effervescence. Le problème se complique d'ailleurs pour la gravitation quantique car les noeuds s'entrelacent en formant des tresses ou entrelacs (*braids* en anglais).

Comment s'effectue le lien entre une collection de ces quanta d'espace que sont les noeuds et l'espace continu de la relativité générale? Un état quantique est ici un état superposant un ensemble de boucles entrelacées. A cette échelle, l'espace-temps est discret, mais l'ensemble de ces boucles tisse une sorte de "cottes de mailles" qui, à plus grande échelle, apparaît comme un tissu continu (*weave* en anglais) comme Wheeler en avait l'intuition dès 1973). Dans cette approche, le graviton n'a pas d'existence propre. Il n'apparaît que dans une limite linéarisée, sous la forme d'une sorte de vague sur un océan à peu près plat (une broderie sur le tissu, en somme). Si l'océan se creuse, il n'est plus identifiable en tant que tel (Ashtekar, Rovelli et Smolin 1992).



Plus techniquement, on cherche à construire des opérateurs correspondant aux notions géométriques de base, telles que l'aire d'une surface à 2 dimensions, et en recherchant à quels états correspondent des valeurs propres approchant de la géométrie classique. Il est par exemple possible de construire des opérateurs *A* et *V* et de montrer que ses valeurs propres sont des multiples respectivement du carré et du cube de la longueur de Planck: on les identifie donc naturellement aux opérateurs *Aire* et *Volume*. Bien sûr une aire mesurable en relativité générale ne peut correspondre qu'à des degrés de liberté physiques, de la matière ou de la gravitation, sinon elle ne serait pas invariante par difféomorphisme.

La construction d'un opérateur *Aire* a permis de démontrer la relation postulée par Bekenstein puis Hawking entre l'aire de l'horizon d'un trou noir et son entropie, premier résultat quantitatif de la gravitation quantique (Rovelli 1996, Ashtekar *et al.* 1998).

Réseaux et mousses de spins

Le formalisme des variables de boucle ne fait déjà plus explicitement appel à la continuité de l'espace (il l'utilise pour les définir à partir des notions classiques, mais une fois la construction terminée, il est possible d'oublier l'échafaudage de construction). Mais on peut aller plus loin: au lieu de considérer seulement des boucles (ou des noeuds), il est possible d'utiliser des réseaux de spin. L'espace en 3 dimensions est remplacé par un réseau, ou un graphe, c'est à dire ensemble de points (les vertexs) reliés par des arêtes. Ces arêtes portent un opérateur du groupe d'holonomie SU(2), donc un indice de spin, et les vertexs assurent la cohérence de ces indices à la jonction des arêtes. Les boucles correspondent ainsi à un cas particulier de réseau. Ces réseaux de spins ont été introduits par Penrose en 1964, bien avant la révolution d'Ashtekar, pour dégager la physique d'un espace-temps continu: seules importent les relations (arêtes) entre objets (vertexs), dans le droit fil des idées de Leibniz et Mach. C'est en fait en utilisant des réseaux de spins que Rovelli et Smolin ont montré en 1995 la quantification des aires et des volumes.

La notion de réseau de spins en 3 dimensions s'étend à la notion de mousse de spins en 4 dimensions. Le but est de décrire ainsi l'espace-temps et non seulement l'espace. Dans ce cas, on a des faces (portant des indices de spin), des arêtes et des vertexs (portant des opérateurs assurant la cohérence des indices de spin aux jonctions). En 4 dimensions, le groupe d'holonomie est SU(2)xSU(2), le recouvrement universel de SO(4), et chaque face devrait donc porter une paire d'indices de spins. Il semble toutefois que ces deux indices soient systématiquement égaux, sinon les amplitudes sont nulles. Mais nous arrivons là aux limites actuelles des recherches en cours.

Retour au sommaire

Approche covariante : d'une théorie des champs comme les autres aux supercordes

L'approche covariante utilise comme objet fondamental le graviton, quantum du champ gravitationnel, et son objectif était de traiter la gravitation exactement comme l'électrodynamique ou les interactions nucléaires. Elle a connu, elle-aussi, une histoire contrastée alternant périodes d'euphorie et périodes de découragement. L'établissement des règles de Feynman pour les diagrammes faisant intervenir des gravitons s'est révélée beaucoup plus difficile que pour la théorie de l'électrodynamique quantique, et elle n'est parvenue à maturité qu'à la fin des années 1960 en forgeant des outils qui allaient d'ailleurs se révéler essentiels pour la théorie des champs de jauge. Cependant la démonstration de la non-renormalisabilité de la gravitation pure allait lancer les théoriciens dans une sorte de fuite en avant avec l'invention de la supersymétrie, de la supergravité, des supergravités étendues, des supercordes et de la M-théorie. Le premier vrai *calcul* de gravitation quantique vient cependant de retrouver la relation entre entropie et surface d'un trou noir, comme dans l'approche canonique mais par des voies radicalement différentes.

La gravitation comme théorie de jauge

Le point de départ (Rosenfeld 1930, Pauli et Fierz 1939, Gupta 1952) est d'écrire la métrique $g_{\mu\nu}$ de l'espace-temps comme la somme d'une métrique fixe $\sigma_{\mu\nu}$ d'arrière plan (en général la métrique plate et statique de Minkowki) et d'une perturbation $h_{\mu\nu}$. **Seule cette dernière est quantifiée.** Elle se comporte comme un champ de spin 2 et de masse nulle (pour que la gravitation soit une force à longue portée), le graviton. La gravitation est ainsi interprétée comme résultant de l'échange de gravitons entre des masses exactement comme l'électrodynamique résulte de l'échange de photons entre des charges, échanges ayant lieu dans les deux cas dans un espace-temps purement classique.

Effectivement, quand on injecte la décomposition $g_{\mu\nu} = \sigma_{\mu\nu} + h_{\mu\nu}$ dans le calcul du scalaire de courbure R qui intervient dans la formulation lagrangienne de la relativité générale, on obtient une série de perturbation dont le premier terme correspond à une théorie de particules de spin 2, de masse nulle et sans interaction, et dont les termes suivants décrivent des couplages entre ces particules, avec un couplage κ égal à la racine carrée de la constante de Newton G.

Dans les années 1960, un petit nombre de physiciens, au premier rang desquels Feynman et DeWitt, se sont attaqué à la très difficile construction des "règles de Feynman" pour cette théorie. Feynman montre en 1963 que les diagrammes en arbres donnent bien la théorie classique linéaire, mais observe ensuite que les diagrammes à une boucle violent l'unitarité. DeWitt élabore entre 1964 et 1967 les règles corrigeant ce problème lors d'un choix de jauge particulier par l'introduction de fermions fictifs, les "fantômes"de Faddev-Popov. L'invariance par difféomorphisme a, sous forme infinitésimale ($h_{\mu\nu} \rightarrow h_{\mu\nu} + d_\mu \xi_\nu + d_\nu \xi_\mu$), exactement la forme d'une transformation de jauge standard et conduit à (l'analogue) des identités de Ward. Les progrès faits dans ces domaines ont d'ailleurs servi aux théories de jauge des interactions faible et forte (unification électrofaible et chromodynamique quantique). A la fin des années 1960, les "règles de Feynman" étaient établies, et il devenait possible de calculer les amplitudes de diffusion entre gravitons ou entre gravitons et champs de matière.

Non-renormalisabilité

Depuis longtemps, on soupçonnait la gravitation de ne pas être une théorie des champs renormalisable. Le soupçon venait d'un simple argument dimensionnel (Heisenberg 1938): à la différence de la constante de couplage de l'électromagnétisme (la constante de structure fine a), de la constante de couplage de l'interaction faible (dans la théorie de Weinberg-Salam) et de l'interaction forte (dans la chromodynamique quantique), la constante de couplage κ du graviton a la dimension de l'inverse d'une masse (la masse de Planck bien entendu).

Rappelons d'abord que la théorie de la renormalisation montre qu'une constante de couplage "effective" G varie avec l'énergie E de l'interaction, cette dépendance étant donnée par l'équation de Callan-Symanzik $dG/d\ln E = \beta(G)$, la fonction $\beta(G)$ pouvant se développer en série: $\beta(G) = (n-d)G + aG^2 + bG^3 + ...$ Dans cette expression, n est la dimension de l'espace-temps (normalement n=4) et d est une "dimension critique" dépendant de la théorie. Si n<d, la constante de couplage tend vers zéro à haute énergie, et la théorie est dite "super-renormalisable", alors que si n>d, la constante de couplage tend rapidement vers l'infini à haute énergie et la théorie est dite "non-renormalisable": un développement perturbatif n'a aucun sens. Si n=d, la théorie est renormalisable et le comportement de la constante de couplage dépend du signe du coefficient a: si a<0, la constante de couplage diminue logarithmiquement à haute énergie (théorie "asymptotiquement libre, comme la chromodynamique quantique), tandis qu'elle diverge pour une valeur finie de E si a>0 (pôle de Landau comme en électrodynamique quantique pure).

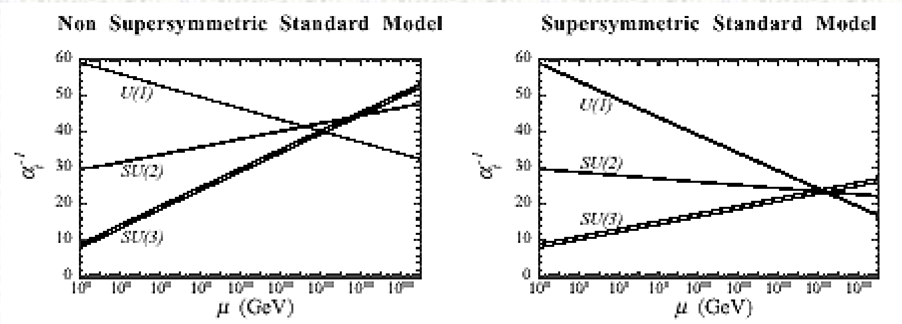
Dans le cas de la gravitation, l'action S[g], intégrale de $d^n x \, \det(g)^{1/2} \{R[g]/8\pi G\}$, est sans dimension (dans les unités des physiciens où tout est rapporté en unités d'énergie en fixant la constante de Planck ***h*** à 1 de même que la vitesse **c** de la lumière), le scalaire de courbure R[g] fait intervenir des carrés de dérivées premières de la métrique g et a donc pour dimension 2, et $d^n x$ a dimension -n. La constante de Newton G a donc pour dimension n-2 (en unités d'énergie). D'après l'équation de Callan-Symanzik G doit varier comme E^{n-d}, ce qui permet d'identifier la dimension critique d=2. Par conséquent, la gravitation est une théorie renormalisable en 2 dimensions d'espace-temps, mais pas en 4!

Il restait l'espoir -ténu- que les divergences de la gravitation se compenseraient miraculeusement en raison des symétries de la théorie, comme elles le faisaient dans les théories de Yang-Mills dont t'Hooft et Veltman montrèrent en 1971 la renormalisabilité. Cependant t'Hooft montra en 1973 que les divergences ne se compensaient pas et qu'il apparaissait à une boucle des termes absents du lagrangien initial: la gravitation était clairement non-renormalisable.

Il était alors logique d'ajouter au lagrangien des termes proportionnels aux divergences, c'est à dire des termes quadratiques dans la courbure. Stelle montra effectivement en 1977 qu'en ajoutant au scalaire de courbure R des termes en R^2 et en $R^{\mu\nu}R_{\mu\nu}$, la théorie redevenait renormalisable pour des choix particuliers des coefficients de ces termes. Malheureuseusement, l'unitarité était de nouveau perdue. Plusieurs années d'efforts intensifs ne parvinrent pas à la récupérer malgré de nombreux faux espoirs.

Supersymétrie et supergravité

Une alternative qui suscita un immense intérêt fut de compenser la divergence des boucles de gravitons (bosons) par la divergence de signe opposé de boucles de fermions (Ferrara, van Nieuwenhuizen et Freedman 1976, Deser et Zumino 1976). Cela exigeait bien sûr une exacte symétrie entre ces bosons et ces fermions (mêmes masses et mêmes couplages), une **supersymétrie**. L'idée était apparue quelques années plus tôt en physique des particules, pour des raisons voisines (compenser les divergences des bosons par celles des fermions), sous la forme d'une symétrie globale. La difficulté, bien sûr, est que la supersymétrie conduit à doubler chaque particule par un "partenaire supersymétrique" et à associer au photon un photino de spin1/2 et de masse nulle, à l'électron un s-électron de spin 0 et de même masse que l'électron, etc. Aucune de ces particules n'ayant été observée, il faut donc supposer que la supersymétrie est une symétrie brisée, et que cette brisure induit une différence de masse entre partenaires supersymétriques. Non seulement la supersymétrie améliore la renaormalisabilité de la théorie standard des particules, mais elle permet d'unifier l'interaction forte avec l'interaction électrofaible à très haute énergie, sans que les effets d'une très haute énergie d'unification ne perturbent la théorie à beaucoup plus basse énergie, là où elle est en aprfait accord avec les observations (cela s'appelle le problème de la stabilisation de la hiérarchie: les petits aimeraient bien devenir gros eux aussi).



L'intérêt de la supersymétrie ne se limite pas à améliorer les propriétés de renormalisation d'une théorie. Plus intéressant encore est le fait que l'anti-commutateur de deux générateurs de supersymétrie est une translation d'espace-temps: une supersymétrie locale inclut de ce fait automatiquement la gravitation.

Ainsi, en ajoutant au graviton de spin 2 un **gravitino** de spin 3/2 dont les couplages sont dictés par la supersymétrie, on obtient une théorie renormalisable à une boucle, et même à deux boucles! Pour aller plus loin, l'idée se fit jour d'utiliser plusieurs supersymétries. Chaque générateur fait varier le spin de 1/2 unité, et si l'on interdit que des particules fondamentales aient un spin supérieur à 2, le nombre maximal de générateurs est de 8, faisant passer d'un état d'hélicité +2 à un état d'hélicité -2. Dans cette supergravité N=8 (Cremmer, Julia et Scherk 1978), on a donc 1 graviton de spin 2, 8 gravitinos de spin 3/2, 28 bosons de spin 1, 56 fermions de spin 1/2, et 70 scalaires de spin 0. La supersymétrie peut bien sûr être définie dans un espace-temps ayant plus de quatre dimensions. La supergravité N=8 en 4 dimensions peut aussi s'écrire comme une supergravité N=1 dans un espace-temps à 11 dimensions (c'est d'ailleurs ainsi qu'elle a été construite) au moyen d'une réduction dimensionelle à la manière de Kaluza et Klein.

Dimensions supplémentaires

L'idée que l'espace-temps pourrait bien avoir plus de quatre dimensions n'est en effet pas nouvelle. Dès 1919, Kaluza montrait que si la relativité générale était écrite en 5 dimensions, le tenseur métrique pouvait se décomposer en un tenseur métrique à 4 dimensions, un vecteur à 4 dimensions et un scalaire. L'intérêt de cette construction vient de ce que le vecteur a exactement le comportement du quadri-potentiel de l'électrodynamique maxwellienne: autrement dit, l'électrodynamique apparaît comme le "reflet" de la gravitation dans la cinquième dimension. La théorie de Kaluza présente pourtant deux inconvénients apparents: les deux couplages sont très différents et la cinquième dimension n'est pas franchement manifeste. En fait, Klein a montré en 1927 que les deux inconvénients disparaissaient simultanément si la cinquième dimension était refermée sur elle-même avec un rayon de courbure R: le couplage de l'électrodynamique *est* alors le couplage gravitationnel divisé par R^2, et si R est assez petit il est normal de ne pas voir directement la cinquième dimension. Numériquement R est de l'ordre de la longueur de Planck, donc tout va bien.

En fait, tout ne va pas si bien, car cette dimension supplémentaire pose quelques problèmes:

- Pourquoi cette dimension a-t-elle une courbure beaucoup plus prononcée que les autres?
- Pourquoi sa courbure ne varie-t-elle pas au cours du temps (sinon le rapport entre la constante de Newton et la constante de structure fine varierait)?
- Les constantes observées ne sont pas fondamentales et leurs relations dépendent d'une propriété inobservable: la "compactification" de la cinquième dimension.
- Etant compacte, cette dimension induit la quantification des excitations des degrés de liberté internes: celles-ci se manifestent en 4 dimensions comme des particules dont la masse est un multiple de la masse de Planck. Pourquoi ne les observe-t-on pas?

Dans le cas de la supergravité, on introduit 7 dimensions supplémentaires. Cela ajoute une difficulté supplémentaire: comment se compactifient ces 7 dimensions, car il existe une infinité de façons de le faire, et les résultats en 4 dimensions sont extrêmement différents selon le choix fait. Comment est sélectionné le "bon choix"? On pourait "compactifier" indépendamment chaque dimension supplémentaire en un cercle (l'ensemble formant alors un tore à 7 dimensions), ou bien les compactifier en une sphère à 7 dimensions (laquelle possède une foule de propriétés mathématiques intéressantes), ou une sphère "écrasée", ou "parallélisée", etc. Chaque symétrie interne de cet espace compact à 7 dimension se manifeste à 4 dimensions comme une symétrie de jauge, dont les bosons de jauge sont les 28 bosons de spin 1 de la supergravité N=8 en 4 dimensions. Par conséquent beaucoup d'études se sont consacrées à ce difficile problème au début des années 1980, dans l'espoir de construire une théorie *finie* (Kallosh démontra en 1980 que la théorie serait finie jusqu'à huit boucles) unifiant la gravitation et les interactions forte et électrofaible, et prédisant le spectre complet des particules élémentaires (bosons et fermions). Une difficulté, peut-être fatale, a été soulevée par Witten en 1981: la compactification d'un nombre impair de dimensions interdit l'existence de fermions chiraux en 4 dimensions (sauf à compactifier sur un orbifold, une variété "anguleuse" dont les points sont identifiés par une symétrie discrètes).

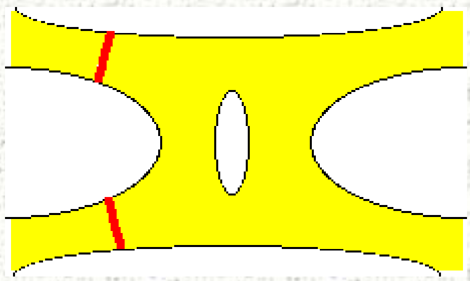
L'espoir était si grand que Hawking inaugurait en 1980 sa chaire à Cambridge par une conférence annonçant la fin prochaine de la physique (apparemment identifiée par lui à ses aspects les plus théoriques), dès que les détails de la supergravité en 11 dimensions et sa réduction en 4 seraient établis.

Cordes

En 1984, une remarquable campagne médiatique convainquait la majorité des théoriciens que la supergravité était en fait bonne à jeter au panier et que l'avenir était dans les supercordes.

Pourquoi les supercordes, et d'où venaient-elles? Il s'agit de la théorie d'un objet fondamental étendu: une particule est, pour le théoricien, un objet ponctuel, un point sans extension, sans dimension, et c'est d'ailleurs là qu'il faut chercher l'origine des infinis des théories (renormalisables ou non). Un objet intrinsèquement étendu, une ligne à une dimension spatiale, pourrait peut-être éliminer ces divergences infinies. On imagine donc ces cordes comme des objets à une dimension se déplaçant dans un espace-temps statique (en général de Minkowki) à D dimensions. De tels objets peuvent avoir deux formes: cordes ouvertes aux extrémités libres, ou cordes fermées en anneau. Ces cordes ont des modes de vibration quantifiés que l'on peut associer à des "particules" de masse et de spin donnés, une corde correspondant ainsi à une infinité de "particules".

Ces cordes se déplaçant dans l'espace-temps tracent une surface d'univers en 2 dimensions. Cette surface d'univers est régulière et peut être décrite par une métrique en 2 dimensions hab. L'interaction de deux cordes se traduit tout simplement par le fait que leurs surfaces d'univers se rejoignent, fusionnent, avant de se séparer éventuellement plus tard. Elles peuvent d'ailleurs fusionner à nouveau, donnant ainsi lieu à des graphes à une boucle:



Il est important de remarquer que la position spatio-temporelle du vertex d'interaction, ponctuelle pour une théorie décrivant des particules, est étendue pour une théorie décrivant des cordes. Cela va considérablement améliorer la renormalisabilité de la théorie.

On peut ainsi concevoir soit une théorie ne contenant que des cordes fermées, soit une théorie contenant à la fois des cordes ouvertes et des cordes fermées (car une corde ouverte engendre inévitablement des cordes fermées).

L'intérêt des cordes pour la théorie quantique de la gravitation vient de ce que **le premier mode d'excitation d'une corde fermée est une "particule" de masse nulle et de spin 2, qu'il est donc tentant d'identifier au graviton**.

La dynamique de l'évolution de ces cordes dans l'espace-temps est déterminée par l'action que l'on choisit pour elle. La solution la plus naturelle pour une corde libre est de prendre une action proportionnelle à la surface d'univers qu'elle balaie, en parfaite analogie avec l'action d'une particule libre proportionnelle à la longueur de sa ligne d'univers:

$$S = T \text{ Intégrale}[\ d^2\sigma\ \{\det(\ daX^\mu\ dbX^\nu\ g_{\mu\nu})\}^{1/2}\ ]$$

Les coordonnées σ repèrent un point sur la surface d'univers, dont les coordonnées spatio-temporelles sont X^μ. La constante de proportionnalité T a pour dimension l'inverse du carré d'une longueur (ou le carré d'une masse) et correspond classiquement à la **tension** de la corde. C'est elle qui fixe les échelles de longueur (et donc de masse) de la corde et de ses modes d'excitation, et on lui affecte comme valeur (le carré de) la masse de Planck.

Sous cette forme (dite de Nambu-Goto), la théorie est difficile à quantifier à cause de la racine carrée du déterminant. Une forme équivalente, due à Polyakov, fait intervenir la métrique $h^{ab}(\sigma)$ sur la surface d'univers:

$$S = T \text{ Intégrale}[\ d^2\sigma\ \{\det(h)^{1/2}\ h^{ab}\ d_aX^\mu\ d_bX^\nu\ g_{\mu\nu}\}\ ]$$

Sous cette forme, l'action est identique à celle d'une théorie décrivant D champs scalaires $X^\mu(\sigma)$ en 2 dimensions. On peut donc y appliquer tout l'arsenal des modèles sigma non-linéaires et des théories des champs conformes.

Quel spectre d'excitations obtient-on? Comme pour les modes de vibration d'une corde de violon, il faut préciser les conditions aux limites (aux bords de la corde). Pour une corde fermée, on choisit bien naturellement des conditions périodiques $X^\mu(\sigma)= X^\mu(\sigma+2\pi)$. Pour une corde ouverte, on a longtemps choisi des conditions de Neumann $dnX^\mu(0)=dnX^\mu(\pi)=0$ (traduisant l'absence d'un flux d'impulsion par les extrémités de la corde). Depuis quelques années, le choix de conditions de Dirichlet $X^\mu(0,\pi)$=constante est préféré (les extrémités de la corde sont obligées de se déplacer sur des hypersurfaces, ce qui viole l'invariance de Poincaré mais ouvre la porte aux D-branes).

Dans le cas d'une corde fermée, on obtient classiquement des modes de vibration se propageant le long de la corde dans une direction (appelons-la par convention la "droite"), et d'autres modes se propageant dans la direction opposée (la "gauche"), en plus bien sûr du mouvement d'ensemble de la corde. Plusieurs méthodes de quantification (canonique dans la jauge du cône de lumière, covariante, par intégrales de chemin) ont été employées, les opérateurs de création et d'annihilation des modes conduisant à un spectre de masse $M^2=T(N-1)$. Le premier mode, N=0, correspond donc à un tachyon de masse carrée négative. Le mode suivant correspond à une particule de masse nulle (et de spin 2), et les suivants à des masses proportionnelles à la racine de la tension T.

Mais plus important encore, les états n'ont des normes positives que si la dimension de l'espace-temps est **D=26**. C'est aussi uniquement pour D=26 que la symétrie de Lorentz, brisée explicitement par un choix de jauge dans la quantification canonique, est retrouvée.

Les théories de cordes ne sont donc viables que dans des espaces-temps de dimension 26 (ou 10 pour les supercordes). Pour qu'elles aient un quelconque rapport avec la réalité observée, il faut donc imaginer que 22 (ou 6) dimensions sont repilées de façon compacte à la Kaluza-Klein. On retrouve donc les avantages et les inconvénients des dimensions supplémentaires. Par exemple une particule de masse nulle en D dimensions donne une série de particules espacées de 1/R en 4 dimensions si les D-4 dimensions compactes ont un rayon R.

Par contre, un objet étendu comme une corde fait apparaître un phénomène nouveau, la **T-dualité**. De quoi s'agit-il? Imaginons une dimension compacte circulaire de rayon R et une corde *fermée* s'enroulant autour. Cette corde a des modes de vibration (d'énergie n/R) et des modes d'enroulement (ou *winding numbers*, d'énergie mR), et son spectre présente donc une symétrie sous l'échange simultané n<->m et R<->1/R. Cette dualité entre très courtes et très longues distances est parfois interprétée comme le signe que les théories de cordes conduisent à une distance minimale dans l'espace-temps de l'ordre de la longueur de Planck.

Et les fermions?

Supercordes

Il faut ajouter à l'action de Polyakov des degrés d'excitation fermioniques:

$$S = T \text{ Intégrale}[\ d^2\sigma \ \{ \ d^aX^\mu \ d_aX_\mu \ - \ i \ \Psi^\mu \ \lambda^a d_a \ \Psi_\mu \} \ ]$$

où les Ψ^μ sont des spineurs sur la surface d'univers et des vecteurs en D dimensions, et matrices λ^a sont les matrices de Dirac en 2 dimensions. Cette action est invariante dans une transformation de **supersymétrie** $dX^\mu=\varepsilon\Psi^\mu$ et $d\Psi^\mu=-i\lambda^a d_a X^\mu\varepsilon$ (ε étant le paramètre de la transformation). On parle donc de **supercorde**.

Les spineurs Ψ^μ se décomposent eux-aussi en modes se déplaçant vers la gauche et en modes se déplaçant vers la droite. Il est possible de choisir différemment les conditions aux limites des modes gauches et droits qui peuvent être symétriques (condition de Ramond) ou antisymétriques (condition de Neveu-Schwarz). Chaque choix conduit à des "secteurs" indépendants de la théorie, et ces choix conduisent à des spectres de masse différents. On a un tachyon pour une (super)corde ouverte "Neveu-Schwarz" mais pas pour la condition de Ramond. De plus, le nombre de modes bosoniques n'est pas le même que le nombre de modes fermioniques, et on utilise une projection (due à Gliozzi, Olive et Scherk) pour obtenir une théorie supersymétrique sans tachyon. L'exigence que les états aient des normes positives fixe la dimension d'espace-temps **D=10**. Une corde fermée a alors des modes d'excitation bosoniques dans les secteurs NS-NS et R-R (dont le graviton de spin 2 et des bosons de spin 0) et des modes d'excitation fermioniques dans le secteur NS-R (gravitinos de spin 3/2, fermions de spin 1/2).

Tous ces résultats ont été obtenus dans les années 1970, quand la théorie des cordes était envisagée comme théorie de l'interaction forte (c'était une formalisation du modèle dual des résonances), avant d'être réinterprétée en 1975 par Scherk et Schwarz comme une théorie quantique de la gravitation.

Les (super)cordes ont suscité un immense intérêt en 1984 quand Green et Schwarz ont démontré qu'il n'existait que 3 théories de cordes cohérentes (5 en y ajoutant les deux modèles hybrides des cordes hétérotiques de gross, Harvey, Martinec et Rohm en 1985). En effet, on peut ajouter aux bosons et aux fermions un indice interne de symétrie (correspondant à un groupe de jauge) pour que la théorie décrive également les interactions forte et électrofaible. Mais les symétries classiques ne sont pas automatiquement conservées au niveau quantique, ce que l'on appelle une anomalie. Fort heureusement, il est possible que les anomalies dues à certains des champs de la théorie soient compensées par celles dues à d'autres champs, rendant la théorie quantique globalement invariante. Cela implique bien entendu un choix particulier du type et du nombre de ces champs, et cela n'est en particulier possible que pour certaines symétries de jauge et certaines dimensions d'espace-temps (26 s'il n'y a que des bosons et 10 avec des fermions). Il n'y a en fait que 2 groupes de jauge possibles, SO(32) et E(8)xE(8), d'où le tableau suivant:

	Type I	Type II A	Type II B	Hétérotique 1	Hétérotique 2
Type de corde	ouverte+fermée	fermée	fermée	fermée	fermée
Supersymétrie	N=1 chirale	N=2 non-chirale	N=2 chirale	N=1 chirale	N=1 chirale
Groupe de jauge	SO(32)	-	-	E(8)xE(8)	SO(32)

Les seuls fermions Ψ^{μ} des **cordes hétérotiques** (=hybrides) se déplacent vers la gauche. Vers la droite, on n'a donc que 10 bosons X^{μ} alors qu'il en faut 26 pour la cohérence de la théorie, et on ne peut pas ajouter 16 bosons quelconques que l'on aurait aussi à gauche rendant la théorie incohérente: on ajoute alors 32 fermions sans indice spatio-temporel mais avec un indice interne à 32 valeurs. Selon les conditions aux limites imposées à ces fermions (toutes périodiques ou antipériodiques, ou moitié-moitié), le groupe de jauge correspondant est SO(32) ou E(8)xE(8) [et non SO(16)xSO(16)!].

Les états de masse nulle de ces théories hétérotiques se répartissent en deux groupes:

- un super-graviton en D=10 dimensions, donnant par réduction à 4 dimensions, graviton, gravitinos et des bosons de masse nulle (dilaton).
- les bosons et fermions de jauge du groupe de symétrie interne.

Remarquons bien que les symétries de jauge des supercordes ne sont **pas** les symétries internes de l'espace compact comme dans les théories de type Kaluza-Klein. Quelle compactification de 10 à 4 dimensions choisir? prendre une forme simple comme une sphère ou un tore conduit à trop de supersymétries en 4 dimensions. On choisit alors des variétés, dites de Calabi-Yau (des milliers d'exemples en sont connus) ou des orbifolds (des variétés dont les points sont identifés par une symétrie discrète).

Le petit nombre de théories viables à D=10 dimensions semble donc conduire à une presque-infinie variété de modèles en 4 dimensions.

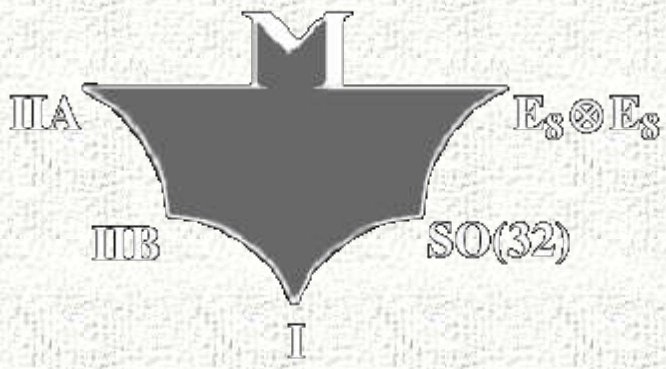
Dualités et M-théorie

Il est progressivement apparu au cours des années 1990 que les 5 théories des cordes n'étaient peut-être que différents aspects d'une même théorie sous-jacente, appelée M-théorie en 1995 par Witten et Townsend (M pour *Mother*, *Matrix*, *Mystery*...). L'origine de cette spéculation est l'existence de dualités reliant les différents modèles.

Nous avons vu que des cordes fermées pouvaient s'enrouler autour d'une dimension compacte de rayon R et qu'il existait alors une dualité, la T-dualité, entre les modes de vibration (énergie en 1/R) et les modes de vibration (énergie en R). En y regardant de plus près, il apparaît que la T-dualité relie en fait les modes de vibration de la corde de type IIA aux modes d'enroulement de la corde de type IIB, et réciproquement. Une T-dualité relie aussi les deux types de cordes hétérotiques.

Il existe aussi une autre dualité, appelée **S-dualité**, qui relie une théorie dans la limite des faibles couplages (limite perturbative) à une autre théorie dans la limite des forts couplages (limite non-perturbative). Une telle S-dualité relie la corde de type I à la corde hétérotique de groupe de jauge SO(32). Quant à la corde de type IIB elle est auto-S-duale.

On peut donc chaîner les théories de cordes, le lien entre le type IIA et la corde hétérotique E(8)xE(8) étant assuré par la mystérieuse M-théorie:



Il se pourrait bien que la M-théorie ne soit autre que la supergravité à 11 dimensions, rejetée cavalièrement en 1984, moyennant un passage par les D-branes. Mais il existe de nombreux avis divergents. Vafa a ainsi défendu en 1996 une F-théorie en 12 dimensions.

P-branes et D-branes

Les brillants résultats de la théorie des cordes ont bien sûr conduit les théoriciens à explorer les possibilités d'objets de dimensions supérieure à 1: des membranes (objets à 2 dimensions) et plus généralement des p-branes (objets à p dimensions). L'équivalent des surfaces d'univers a alors p+1 dimensions, et la puissance des théories conformes à 2 dimensions n'est plus applicable. La théorie quantique diverge à nouveau. Il a cependant été découvert que les p-branes pouvaient coexister avec des cordes dans un espace-temps à D dimensions. Mieux, si les extrémités d'une corde ouverte sont assujetties à se déplacer à la surface d'une p-brane, une riche phénoménologie émerge. On parle alors d'une **D-brane** (D comme Dirichlet) à la suite de Polchinski (1995).

On peut ainsi considérer que notre univers à 4 dimensions n'est qu'une D-brane se déplaçant dans un univers à 10 dimensions. Fermions usuels et bosons de jauge de Su(3)xSu(2)U(1) seraient des modes d'excitations de cordes ouvertes à 10 dimensions dont les extrémités, attachées à notre D-brane, s'y déplacent en nous apparaissant sous forme de particules. Le graviton correspond, lui, à une corde fermée et il n'est pas attaché à la D-brane: il se déplace en 10 dimensions. La conséquence est que l'intensité des forces de gravitation, ressentie à 4 dimensions, est beaucoup plus faible que ce qu'elle est "réellement" en 10 dimensions.

Une conséquence de cette hypothèse est que les dimensions supplémentaires ne sont pas nécessairement très petites, de l'ordre de la longueur de Planck naïve, et pourraient même être d'ordre *millimétrique* (la limite actuelle des tests de la loi de gravitation en 1/r2 de Newton).

Rien n'interdit a priori l'existence de plusieurs D-branes, ce qui se traduirait expérimentalement par l'existence de plusieurs univers analogues au nôtre mais n'interagissant que gravitationnellement avec nous, à travers l'espace entre les D-branes (ou *bulk*). L'origine de la matière noire?

Revenons à la M-théorie. La supergravité en 11 dimensions pourrait être compactifiée sur un 2-brane en tore. Ce tore apparaît comme une corde fermée si on néglige son épaisseur, et cette corde a les caractéristiques de la corde de Type IIA, confirmant une partie de l'hypothèse selon laquelle la M-théorie est la supergravité. La supergravité pourrait aussi donner la supercorde hétérotique E(8)xE(8) si une des 11 dimensions était compactifiée en un segment: on aurait alors deux univers de matière, chacun avec un groupe de symétrie E(8) n'interagissant que gravitationnellement.

Bilan

La théorie des cordes a permis des percées spectaculaires:

- L'existence nécessaire de cordes fermées, et donc d'un graviton, dans toute théorie de cordes implique l'existence de l'interaction gravitationnelle.
- La présence de fermions entraîne la supersymétrie de la théorie, qui a de nombreux avantages théoriques.
- La cohérence quantique (absence d'anomalies) entraîne la présence de groupes de symétrie interne contenant SU(3)xSU(2)xU(1). les cordes impliquent l'existence des interactions forte et électrofaible.
- Tous les modèles de cordes découlent d'une théorie unique, la M-théorie.
- Le calcul de la relation entre entropie et masse d'un trou noir a pu être mené à terme.

Il reste quelques menus problèmes:

- L'espace-temps "réel" a 10 ou 11 dimensions, et le passage à 4 dimensions reste problématique.
- Il entraîne en particulier la présence de particules de masse nulle (dilatons, modules...) et de particules très massives ("pyrgons"...).
- L'espace-temps est un espace-temps fixe, indépendant de son contenu, et en général plat, continu et commutatif.

Retour au sommaire
