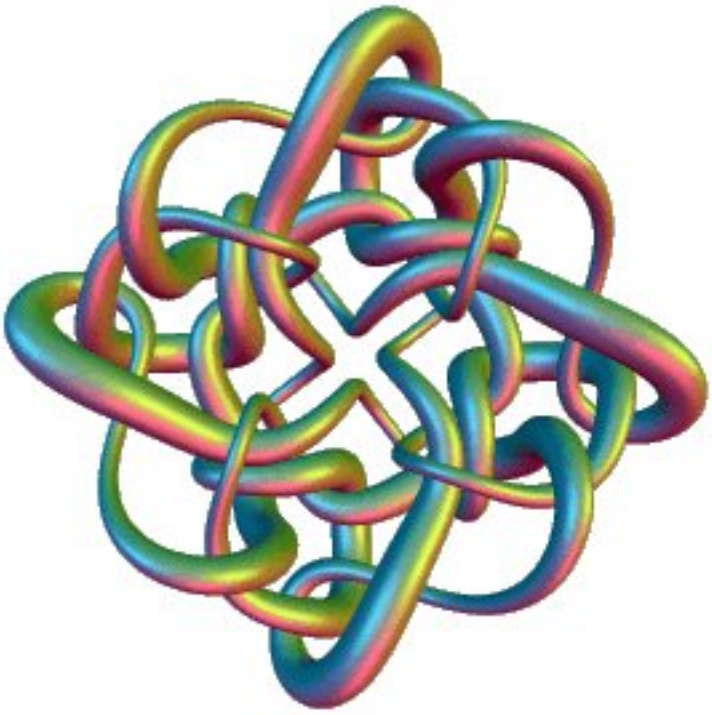


LA THÉORIE DES CORDES

Alexandre Depire
Institut Henri-Poincaré

TABLE DES MATIÈRES

1. Introduction

Le but de cette étude est d'expliquer l'historique de l'élaboration de la théorie des cordes.

A cette occasion, nous présenterons, bien sûr, le contexte dans lequel cette gestation a eu lieu. Nous serons donc amenés à parler des autres théories physiques fondamentales. Celles-ci seront expliquées dans les grandes lignes en mettant l'accent sur les aspects intéressants pour la théorie des cordes.

Je n'aborderai pas les aspects expérimentaux, c'est à dire les expériences actuelles ou futures qui pourraient confirmer ou infirmer la théorie des cordes ou départager plusieurs de ses variantes. C'est un coté très intéressant mais il n'est pas nécessaire pour expliquer la théorie des cordes.

Premièrement, je mets moins l'accent sur la recherche d'une théorie universelle mais plutôt sur l'aspect historique de la construction de la théorie des cordes. C'est à dire que je désire approfondir le "comment" plutôt que le "pourquoi". Bien qu'historique, il y a parfois chevauchement de certaines parties. Ceci est dû à la nécessité de regrouper certaines explications, sinon la description serait trop décousue et plus difficile à comprendre. Je donnerai quelques dates chaque fois que ce sera possible. Mais il faut aussi signaler que ces dernières années l'histoire s'est précipitée. Il y a eu un foisonnement extraordinaire d'idées et de développements théoriques sur la théorie des cordes, ce qui rend difficile une stricte chronologie.

Deuxièmement, je souhaitais approfondir la description de la théorie des cordes. Avouons le, j'avais surtout envie d'écrire cette étude !

2. Les pionniers

2.1. Kaluza

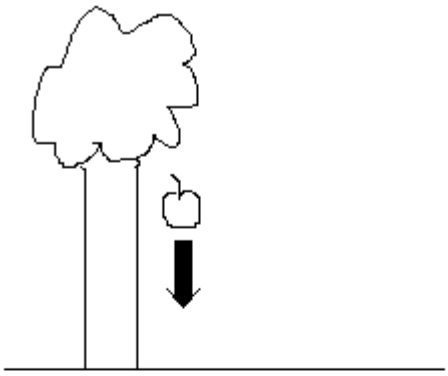
Le début de l'histoire

Nous voici au début de notre histoire en 1919.

Quelle est la situation à cette époque ? Les théories quantiques, que nous aborderons plus tard, sont seulement en gestation. Les deux théories maîtresses sont la relativité générale d'Einstein et la théorie de l'électromagnétisme de Maxwell. Ces deux théories semblaient pouvoir tout expliquer (à quelques "détails" près comme les atomes) et avaient atteint un degré de raffinement mathématique extraordinaire.

Je décrirai un peu plus loin ces deux théories. La relativité générale est la théorie qui s'applique à la gravitation. C'est à dire à la force qui fait tomber les pommes et tourner les planètes. C'est typiquement une théorie dont le domaine est les grandes échelles, disons de la taille des poussières aux galaxies.

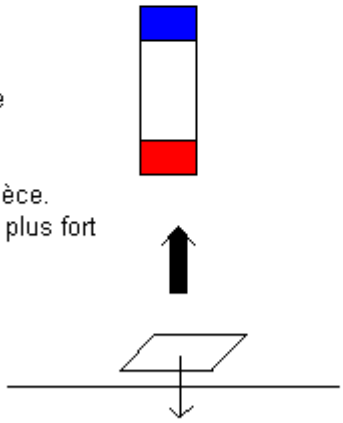
L'électromagnétisme traite, comme son nom l'indique, de l'électricité et du magnétisme. Mais c'est aussi la théorie qui s'applique à la lumière et plus généralement aux ondes radios, aux rayons X, etc. Les forces électriques et magnétiques sont très différentes de la gravitation. Cette dernière est toujours attractive. Alors que les forces électromagnétiques peuvent être répulsives, ce qui s'observe facilement avec deux aimants. Les forces électromagnétiques sont aussi considérablement plus fortes que la gravitation. Avec un simple petit aimant, on soulève facilement un morceau de fer, le soustrayant ainsi à la force de gravité qui le maintenait au sol.



L'attraction universelle de la gravitation est toujours attractive.

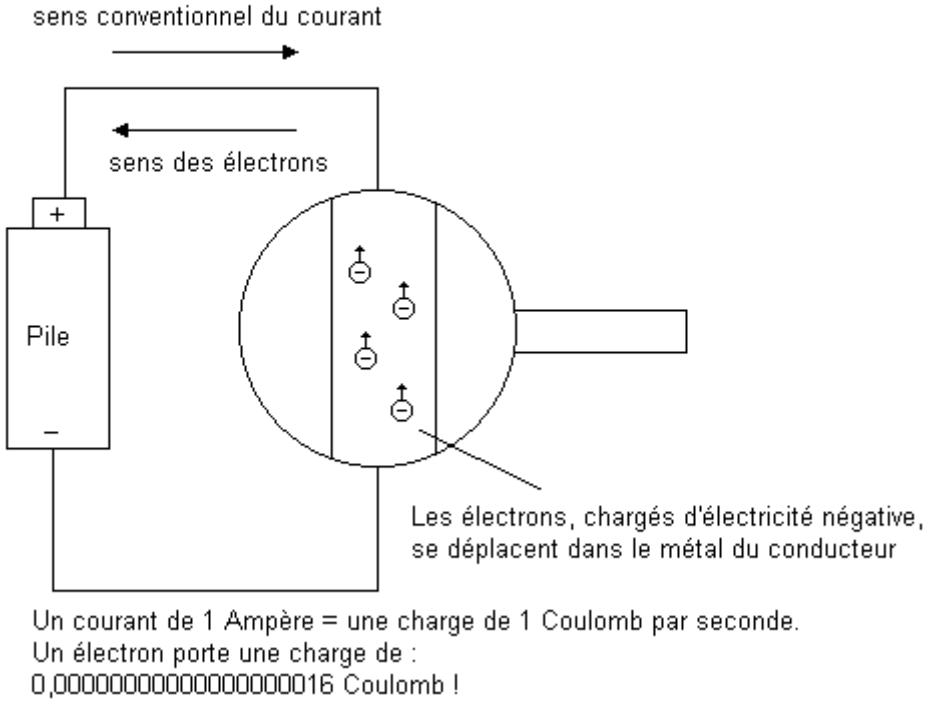
Deux aimants peuvent se repousser. Le champ électromagnétique peut être attractif ou répulsif.

La pièce métallique est attirée vers le bas par l'attraction de la gravité (due, ici, à la terre). Pourtant l'aimant soulève la pièce. Un petit aimant est beaucoup plus fort que la terre entière !



L'interaction électromagnétique est beaucoup plus forte que l'interaction gravitationnelle

Cette intensité plus grande des forces électromagnétismes explique aussi que leur domaine de prédilection s'étend à des échelles très petites. Le courant électrique, dans un fil conducteur, est du au parcours des électrons qui sont électriquement chargés. Et les électrons sont extraordinairement petits. Pour un courant de un ampère, il passe environ seize milliards de milliards d'électrons par seconde dans le conducteur !



Enfin, même au niveau de leur description mathématique, ces deux théories sont très différentes.

Le problème qui se posait à l'époque était : comment unifier les deux théories ? C'est à dire, comment obtenir une seule théorie expliquant à la fois la gravité et l'électromagnétisme ? Etait-ce seulement possible ?

La solution ne semblait pas évidente quant, en 1919, Théodore Kaluza eut une idée curieuse qui va nous entraîner très loin. Mais nous y reviendrons plus loin après avoir un peu mieux décrit la relativité générale et l'électromagnétisme.

La relativité restreinte

La relativité d'Einstein s'exprime simplement avec deux postulats :

- Les lois physiques doivent s'exprimer de la même manière pour tout observateur.
- Il existe une vitesse limite à la transmission des signaux.

Elle fut formulée au début du XXème siècle par Einstein.

A la place d'observateur on emploie souvent le terme "repère". C'est une notion mathématique mais qui exprime tout simplement qu'un observateur donné mesure (repère) les événements qui l'entourent par rapport à lui avec des règles graduées et une horloge.

Le premier postulat semble évident. Si une loi physique explique un phénomène, il est évident que cette loi est la même pour tous les observateurs. Et il est naturel de demander à ce que l'expression (mathématique) de cette loi soit également semblable (pour ce qui est de sa formulation) pour tous les observateurs.

Le deuxième postulat semble moins évident. A la fin du XIXème siècle, Michelson et Morley avaient découvert que la vitesse de la lumière (dans le vide) ne dépendait pas de la direction (ce que l'on

croyait à l'époque, je ne m'étendrai pas sur ce sujet). Puis on découvrit que plus généralement cette vitesse était identique pour tout observateur.

Ce postulat est donc d'origine expérimentale. Il devrait donc, en toute rigueur s'exprimer :

- Il existe une vitesse (égale à celle de la lumière dans le vide) identique pour tout observateur.

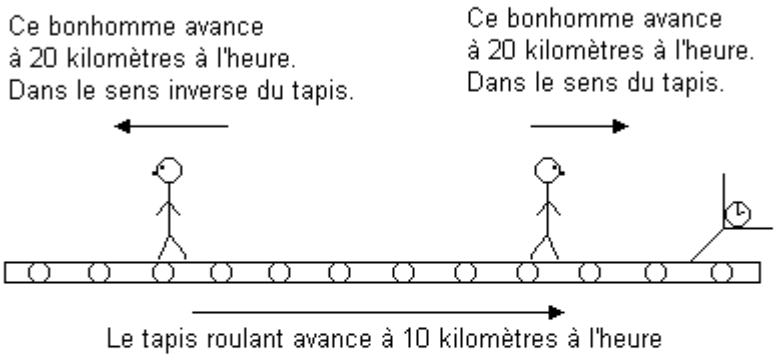
Que cette vitesse soit aussi une vitesse limite impossible à dépasser n'est pas très important (et c'est aussi une conséquence du reste de la théorie).

Ces deux postulats ne semble pas extraordinaires et pourtant, que de conséquences !

Limitons-nous d'abord à la relativité restreinte. Celle-ci se limite à des observateurs qui se déplacent l'un par rapport à l'autre à vitesse constante. On n'aborde pas le cas des accélérations. Le problème est donc plus simple mais ici aussi les conséquences sont extraordinaires.

Prenons tout simplement l'addition des vitesses. Supposons que je sois dans un train. Ce train roule à 50 kilomètres par heure. Je me déplace dans le train, par exemple dans le même sens que le train, à la vitesse de 5 kilomètres par heure. Alors, je me déplace par rapport au sol à la vitesse de 55 kilomètres par heure, c'est à dire la vitesse du train ajoutée à ma propre vitesse.

L'addition des vitesses

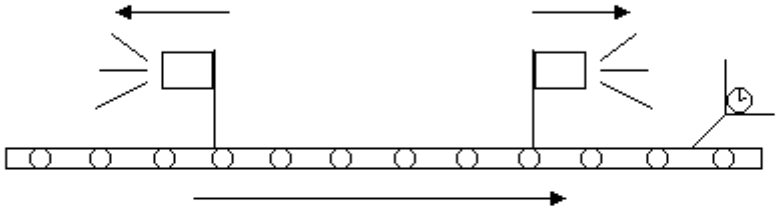


L'observateur au sol voit le bonhomme de droite avancer vers la droite à 30 kilomètres à l'heure (20 + 10).
Et le bonhomme de gauche à 10 kilomètres à l'heure vers la gauche (20 - 10).

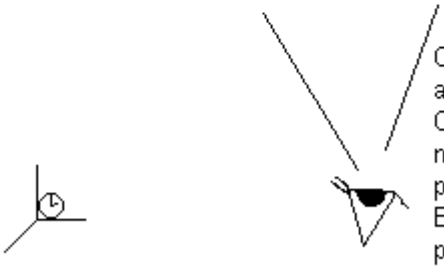
Mais pour la lumière c'est faux !

Addition des vitesses aves la lumière

La lumière est émise par des spots placés sur le tapis roulant. La lumière est émise à gauche et à droite avec une vitesse de 300.000 kilomètres par seconde (mesurée sur le tapis).



Le tapis roulant avance à 200.000 kilomètres par seconde



C'est la même expérience que tout à l'heure avec les bonhommes sur le tapis roulant. On s'attend à ce que l'observateur au sol mesure une vitesse de 500.000 kilomètres par seconde à droite et 100.000 à gauche. Eh non, il mesure toujours 300.000 kilomètres par seconde ! Clairement, à ces vitesses très élevées l'addition classique des vitesses n'est plus valable. Il se passe quelque chose de "spécial".

Comment expliquer ce paradoxe ? Il faut évidemment revoir notre règle d'addition des vitesses. Mais la vitesse est la mesure du déplacement d'une certaine distance au cours d'un certain temps. Donc, il faut aussi revoir nos notions de distances et de temps !

Cela se montre aisément. Pour en savoir plus, je conseille de lire LA MESURE DU TEMPS et LA MESURE DES DISTANCES.

Récapitulons simplement brièvement les conséquences de la relativité restreinte.

1. Le temps n'est plus absolu. Pour un observateur immobile (au repos), le temps d'un objet en mouvement semble s'écouler plus lentement. Ce phénomène est appelé dilatation du temps.
2. L'espace n'est plus absolu. Pour un observateur au repos, un objet en mouvement semble être plus court dans le sens du déplacement. Ce phénomène s'appelle la contraction des longueurs.
3. L'addition classique des vitesses (on dit galiléenne) n'est plus valable. Les règles permettant de passer des mesures de la position et du temps pour un observateur à un autre, s'appellent les transformations de Lorentz. Elles permettent d'obtenir la bonne formule pour les vitesses.
4. L'énergie et la masse c'est la même chose. Lorsqu'un système perd de l'énergie il devient plus léger ! Bien entendu ce phénomène ne s'observe pas aisément car un kilogramme de matière équivaut à nonante millions de milliards de Joules ! Pour le constater, il faudrait détruire entièrement ce kilogramme de matière. Ce qui ne pourrait se faire qu'avec de l'anti-matière (les physiciens observent ce phénomène avec les particules tel que des électrons et des anti-électrons). Mais l'effet est déjà sensible (et mesurable) avec l'énergie nucléaire.
5. Pour un observateur au repos, un objet qui se déplace a une énergie plus élevée. Ce qui est normal ! C'est tout simplement l'énergie cinétique qui peut se convertir sous d'autres formes lors d'un impact par exemple, et tout le monde sait qu'il faut de l'énergie pour mettre un objet

en mouvement. Mais à cause de (4) l'objet a aussi une masse plus grande. Et lorsque l'objet approche la vitesse de la lumière, pour l'observateur au repos, sa masse (et son énergie) augmente de plus en plus vite. A la limite, si l'objet atteignait la vitesse de la lumière sa masse deviendrait infinie ! Il faudrait, pour qu'il atteigne cette vitesse, lui communiquer une énergie infinie ce qui explique que cette vitesse ne peut pas être dépassée. Cela implique aussi que la lumière est sans masse propre (sinon sa masse apparente, à la vitesse de la lumière, serait infinie).

La relativité générale

La relativité générale se propose d'étudier ce qui se passe si le premier postulat est valable pour tout observateur. Donc même pour des observateurs qui ne se déplacent pas à vitesse constante. C'est à dire des observateurs qui accélèrent, ralentissent, tournent, … Cette fois la situation est nettement plus compliquée. Mais on peut raisonner intuitivement et comparer les accélérations avec la gravité. En effet, lorsque l'on lâche un objet celui-ci tombe et sa vitesse augmente de plus en plus (jusqu'à ce qu'il heurte le sol ou soit freiné par l'air, bien sûr). C'est pourquoi on parle d'accélération de la pesanteur. Sur terre (au niveau du sol) un objet qui tombe est accéléré de 10 mètres par seconde au carré. Ce qui signifie qu'à chaque seconde qui passe sa vitesse augmente de 10 mètres par seconde.

Expérimentalement on constate que la masse inerte (qui s'oppose à la mise en mouvement d'un objet) et la masse pesante (la masse responsable de l'attraction gravitationnelle) sont toujours identiques. C'est le principe d'équivalence. D'où l'idée d'identifier accélération et gravitation.

La comparaison conduit à une hypothèse étonnante : l'espace (et le temps qui est indissociable en relativité) est courbe ! Voir par exemple LA RELATIVITE GENERALE et LA COURBURE DE L'ESPACE-TEMPS.

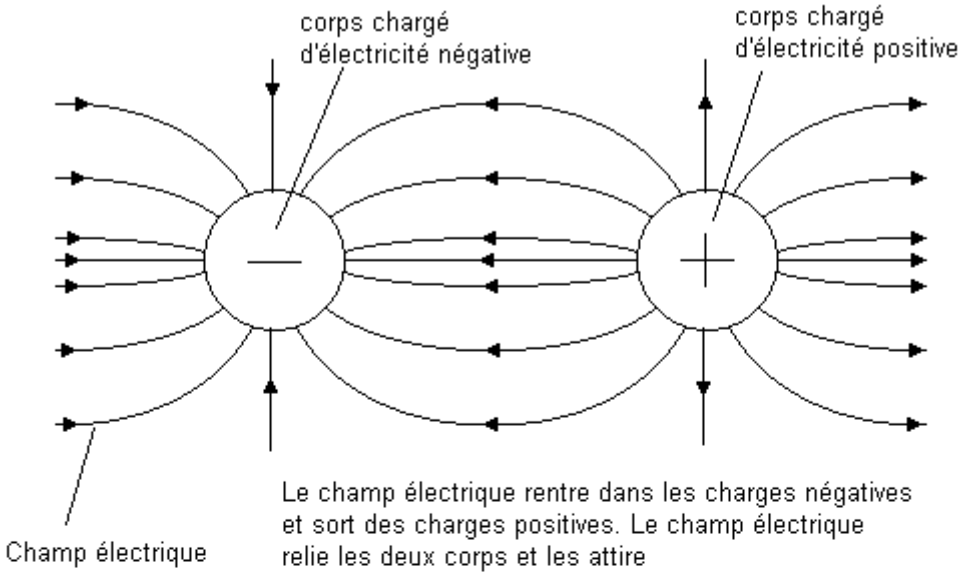
Voici résumé quelques conséquences (non exhaustives) de la relativité générale :

1. L'espace et le temps sont courbe.
2. La courbure influence la trajectoire des objets qui suivent (lorsqu'ils se déplacent librement) le chemin le plus court appelé géodésique (dans un espace plat c'est évidemment une droite).
3. La matière (et donc l'énergie) donne la courbure à l'espace-temps. Plus un objet est massif, plus cette courbure devient importante. La courbure ne devient notable que pour des objets très massifs (comme les étoiles). Mais même une très légère courbure comme celle provoquée par la terre suffit à expliquer le mouvement de la lune !
4. La gravité est une conséquence de cette courbure
5. Les rayons lumineux sont déviés par les étoiles. Ce qui fut constaté la première fois en observant les étoiles proches du soleil (proches sur la voûte céleste) grâce à une éclipse totale.
6. Les planètes ne décrivent pas de parfaites ellipses mais dérivent légèrement au cours du temps. Ce qui peut se mesurer pour la planète mercure.
7. Pour un observateur extérieur, le temps près d'un objet très massif (une étoile à neutrons par exemple) semble s'écouler plus lentement.

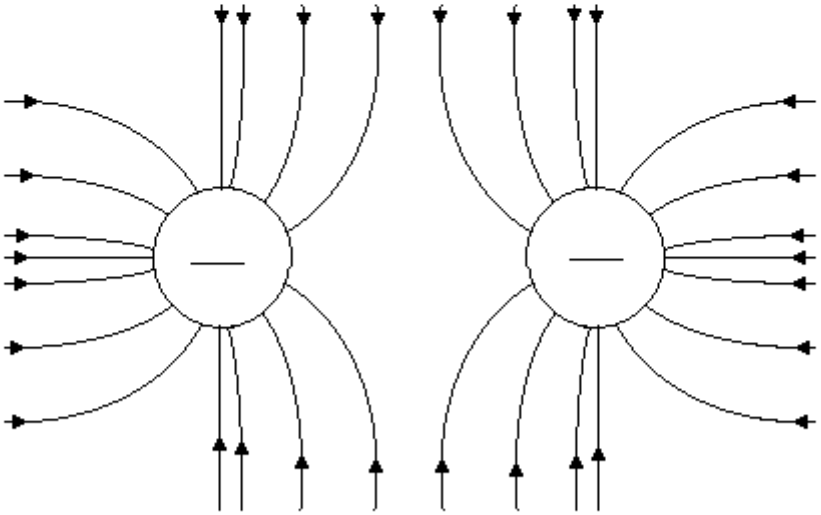
L'électromagnétisme

La théorie de l'électromagnétisme fut établie par Maxwell à la fin du XIXème siècle. Elle fut l'aboutissement du travail d'un grand nombre de prédécesseurs qui avaient étudié l'électricité et le magnétisme.

Lorsqu'un objet est chargé électriquement il émet autour de lui un champ électrique. Lorsque deux objets sont chargés électriquement ils s'attirent ou se repoussent selon que leurs charges sont de signes opposés ou égaux.

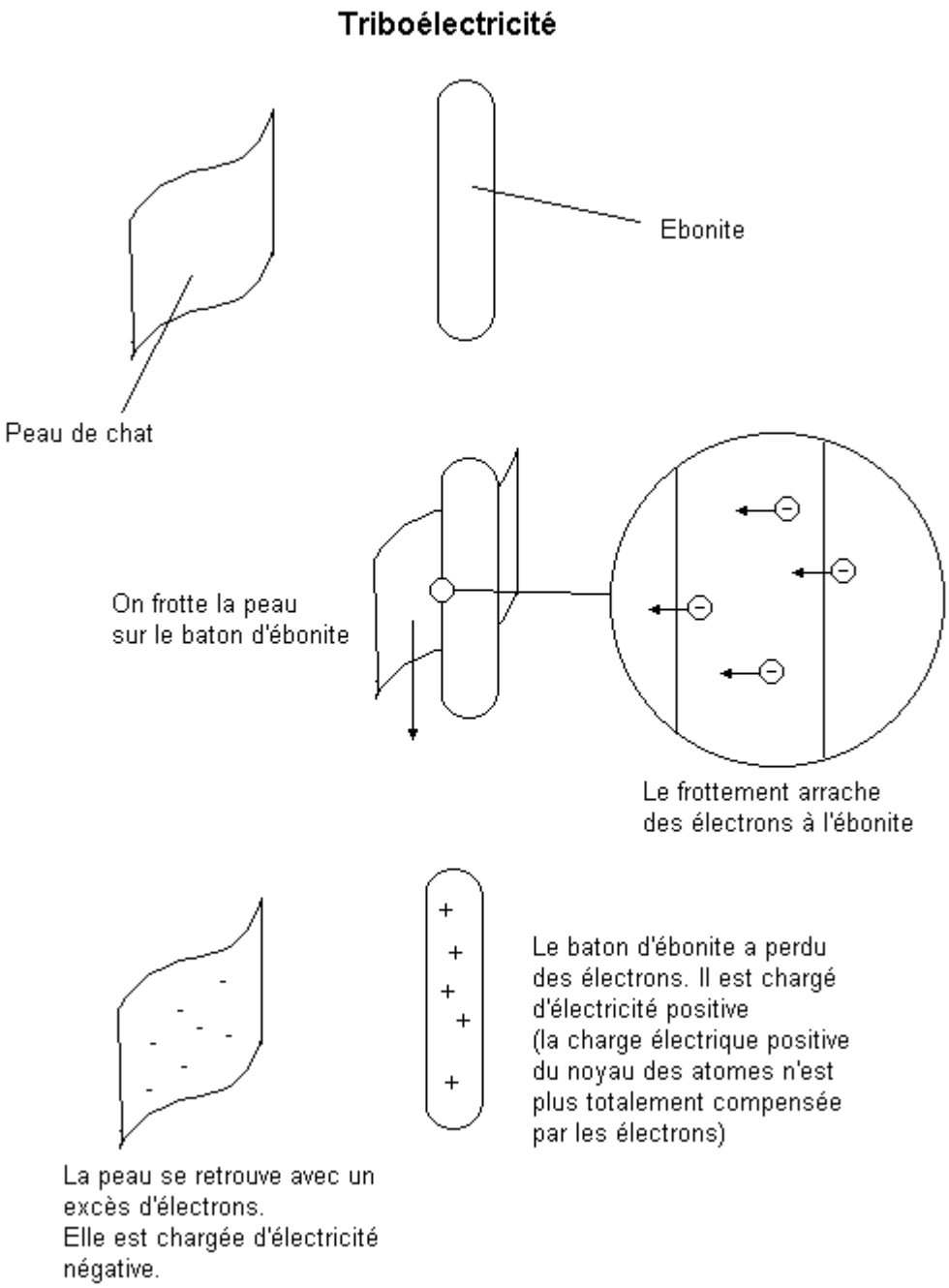


Le champ électrique rentre dans les charges négatives et sort des charges positives. Le champ électrique relie les deux corps et les attire

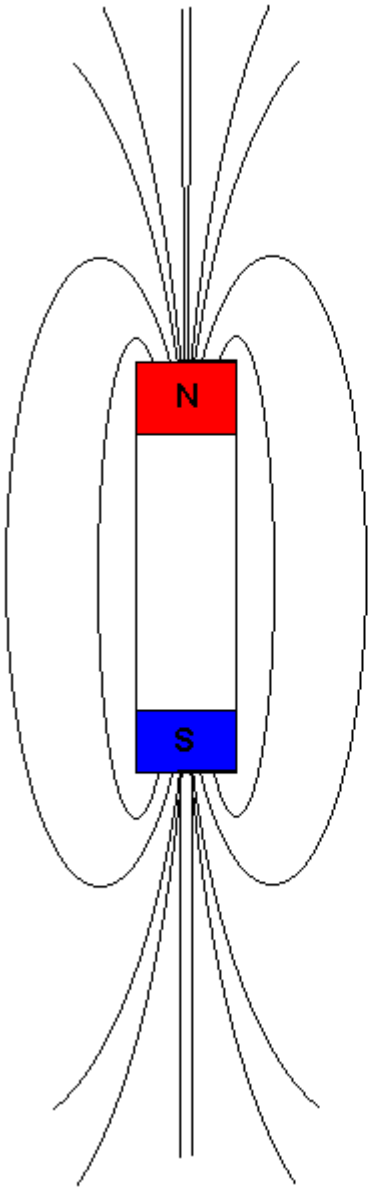


Si les deux charges ont le même signe.
Le champ électrique ne peut joindre les deux corps et les charges électriques sont repoussées.

En pratique, les charges électriques négatives sont produites par les électrons, ou plus exactement par un excès d'électrons. De même, les charges positives sont dues à un déficit en électrons.



Le champ magnétique peut être provoqué par un aimant.

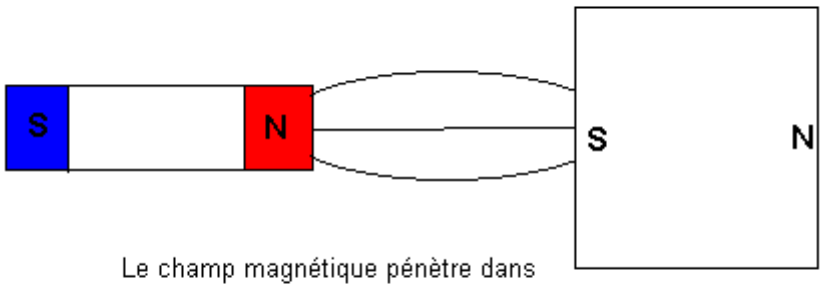


Un aimant possède un pôle nord et un pôle sud (analogue des charges électriques positives et négatives) et émet un champ magnétique

Tout comme il y a des charges électriques positives et négatives, les aimants ont un pôle sud et un pôle nord. Les pôles de même type se repoussent, ceux de type différent s'attirent. Avec la particularité que les pôles d'un aimant ne peuvent pas être isolés contrairement aux charges.

Lorsque l'on place un métal ferromagnétique (comme le fer) dans un champ magnétique il prend lui-même une aimantation, c'est pourquoi il est attiré par l'aimant.

Aimantation induite

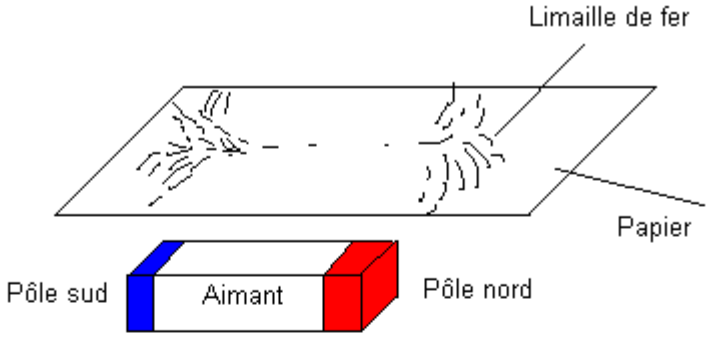


Le champ magnétique pénètre dans le fer. Il devient lui même un aimant. Le pôle nord attire le pôle sud tout comme les charges électriques positives attirent les négatives

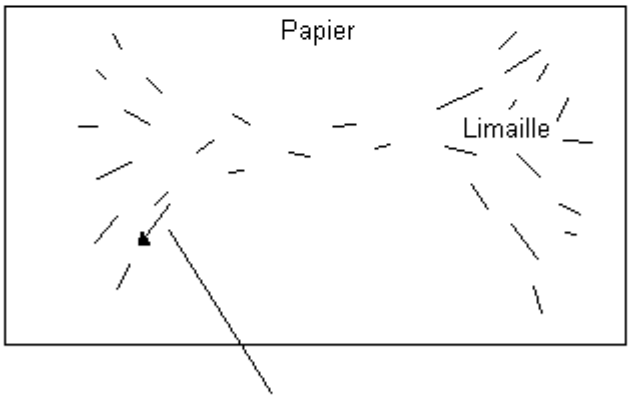
Mais nous n'arrêtons pas de parler de "champs". Qu'est-ce que c'est ? Un champ est une propriété physique (électrique, magnétique ou autre) qui prend une valeur en tout point de l'espace (un peu comme un fluide).

Lorsque cette valeur est un simple nombre, on parle de champ scalaire. Par exemple, la température est un nombre (en degrés Celsius par exemple) définit en chaque point. La température forme donc un champ scalaire.

Cette valeur peut être plus compliquée qu'un simple nombre. Par exemple, elle peut être un vecteur. Un vecteur est comme une petite flèche avec une grandeur et une direction. Les champs électriques et magnétiques sont de ce type, ce qui se vérifie aisément en disposant de la limaille de fer au-dessus d'un aimant.

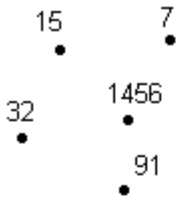


Les champs électriques et magnétiques sont invisibles à l'oeil nu. Mais on peut les visualiser par exemple en mettant de la limaille sur une feuille de papier au dessus d'un aimant. La limaille s'alligne selon les lignes du champ magnétique.

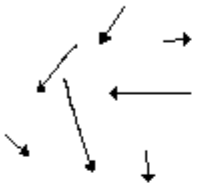


Le champ a une intensité et une direction qui suit la limaille. On le représente en ce point par une petite flèche de grandeur égale à l'intensité. C'est un vecteur.

Une suite de vecteurs peut former une trajectoire appelée ligne de champ. Il est plus facile de représenter un champ avec ces lignes de champs.

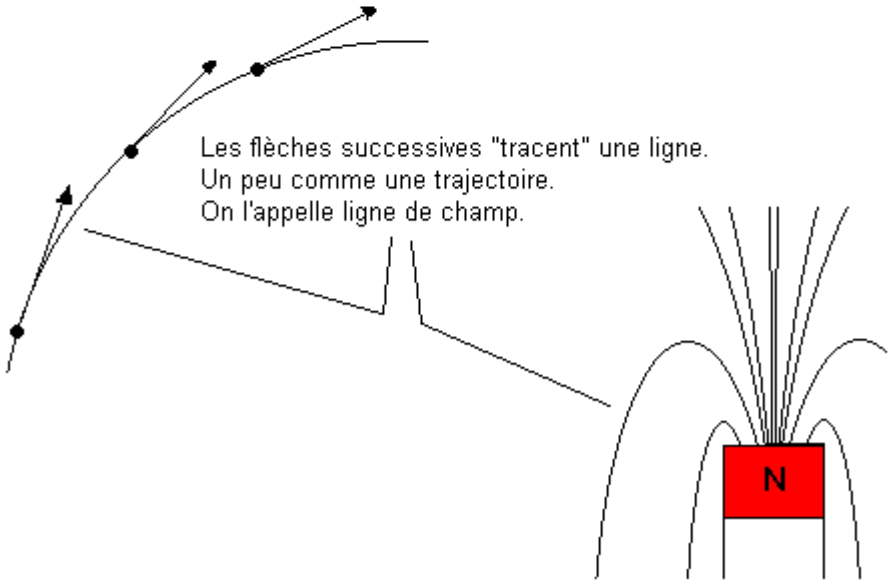


Champ "scalaire". Il a une valeur en chaque point.
Un exemple simple est la température (qui varie de point en point).

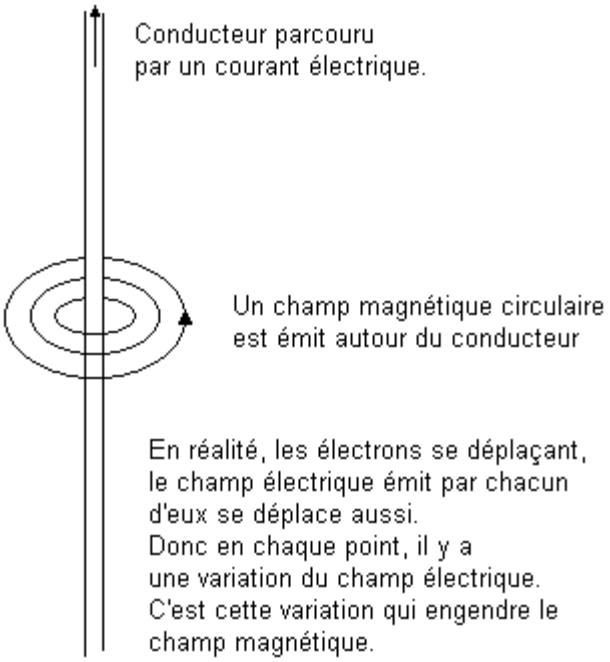


Champ "vectoriel". En chaque point il y a une valeur qui peut être représentée par une flèche d'une certaine grandeur et d'une certaine direction.
Un exemple simple est la vitesse d'un fluide. En chaque point le fluide a une vitesse donnée.

En réalité, il y a une valeur en TOUS les points.
Mais on ne peut en dessiner que quelques unes pour une question de place.

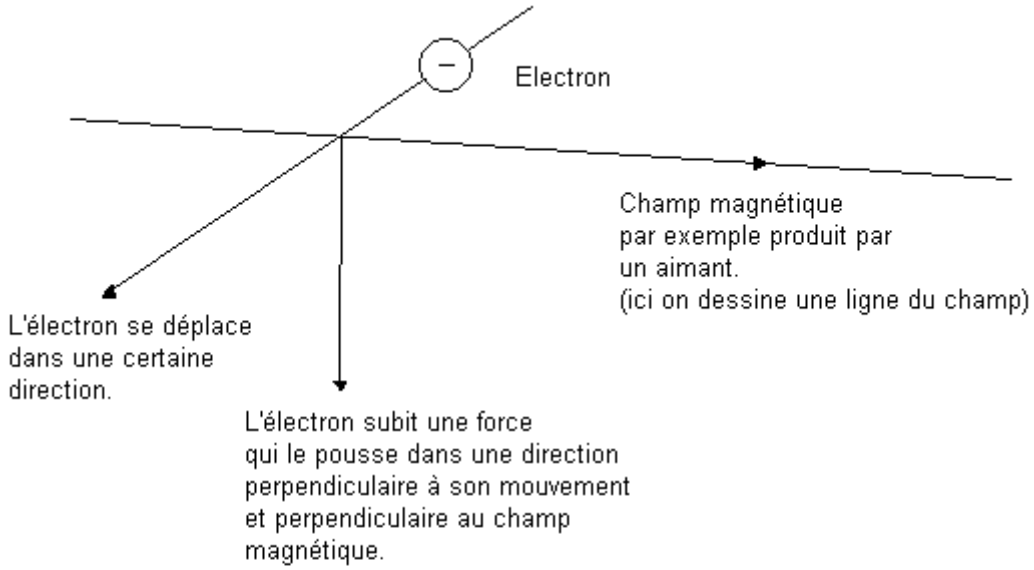


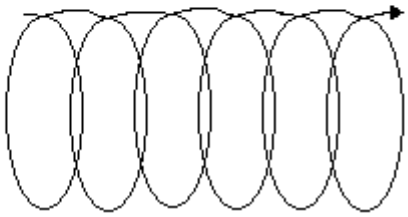
Lorsque l'on étudie l'électricité et le magnétisme on constate rapidement qu'ils ne sont pas indépendants. Par exemple un courant électrique provoque un champ magnétique. Nous avons vu qu'un courant électrique est un simple flux d'électrons. Et effectivement, toute charge électrique en mouvement engendre un champ magnétique (en fait tout champ électrique variable engendre un champ magnétique).



De même, un champ magnétique variable provoque l'apparition d'un champ électrique. Ce champ électrique agit sur les particules chargées comme les électrons et peut provoquer l'apparition d'un courant dans un conducteur. C'est sur ce principe que fonctionne une dynamo.

Enfin, lorsqu'une charge électrique se déplace dans un champ magnétique elle subit une force perpendiculaire.





A cause de cette force constament perpendiculaire. Un électron qui se déplace dans un champ magnétique constant parcourt une spirale !

C'est sur ce principe que fonctionne un moteur électrique.

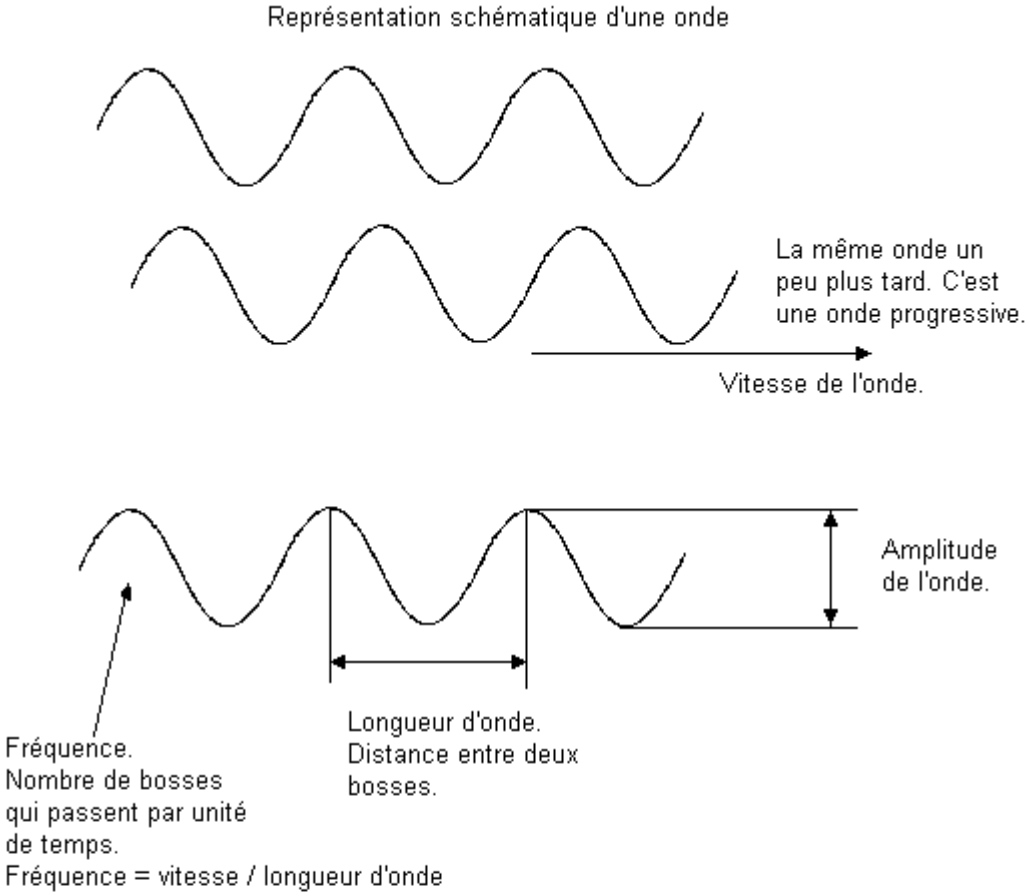
Il s'avère donc que les champs électriques et magnétiques sont intimement liés.

Maxwell réussit à unifier les deux champs dans un unique jeu d'équations remarquablement symétriques. On y voit clairement que le champ électrique et le champ magnétique jouent un rôle semblable.

Ces équations montrent d'ailleurs que les deux champs peuvent dériver d'un unique champ appelé potentiel électromagnétique. Il comporte une composante vectorielle et une composante scalaire.

Une des conséquences des équations de Maxwell est la possibilité d'avoir des ondes électromagnétiques.

Une onde est comme une vague. C'est une vibration qui se propage. Elle possède une fréquence, une vitesse et une longueur d'onde.



Il est possible d'exprimer les équations de Maxwell sous forme relativiste (la relativité restreinte). En réalité les équations sont inchangées ! En effet, les équations de Maxwell sont déjà relativistes (on peut dire qu'elles étaient en avance sur leur temps). Ceci n'a rien d'étonnant car les ondes électromagnétiques se propagent à la vitesse de la lumière. A cette vitesse, la relativité est reine et une théorie correcte ne pouvait être que relativiste.

On peut toutefois exprimer les équations à l'aide des notations mathématiques relativistes (appelées notations covariantes). Sous cette forme les équations deviennent incroyablement simples et compactes (une seule équation extrêmement courte).

Formulées de cette manière, les champs électriques et magnétiques s'écrivent comme un champ unique appelé bien évidemment champ électromagnétique. C'est un champ tensoriel. Les tenseurs

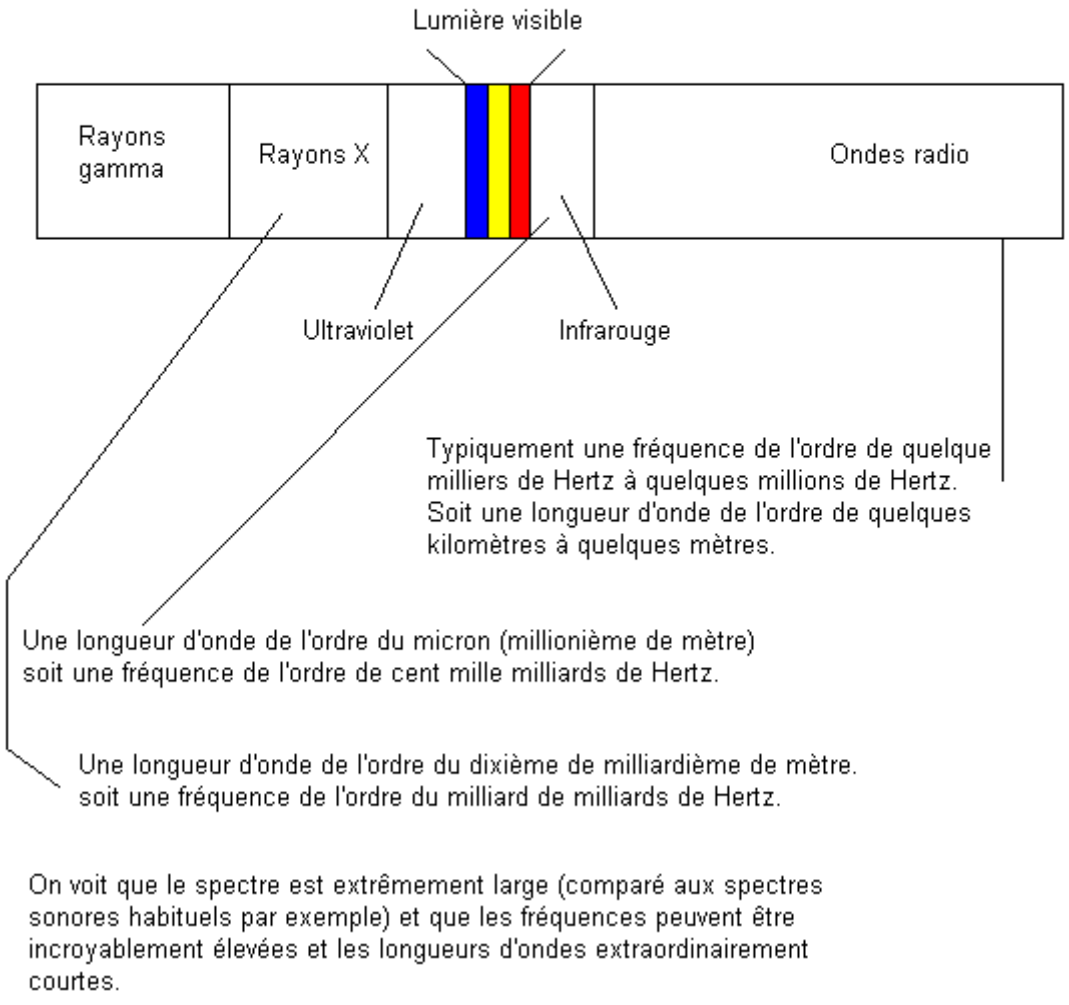
sont des objets mathématiques un peu plus compliqué que les vecteurs. Ils sont par rapport aux vecteurs ce que les vecteurs sont aux scalaires. Un scalaire n'a besoin que d'un nombre. Un vecteur a besoin de quatre nombres (en relativité, ceci est lié au fait que l'espace temps à quatre dimensions : trois d'espaces et une de temps). Et un tenseur a besoin de 16 nombres.

Le tenseur électromagnétique est un peu particulier (mathématiquement c'est un tenseur "antisymétrique"). Il y a des restrictions sur ses valeurs qui impliquent qu'il se comporte un peu comme un vecteur. Ce n'est pas étonnant puisqu'il est en réalité composé de deux champs vectoriels !

Selon leurs fréquences, les ondes électromagnétiques se manifestent comme de la lumière, des ondes radios, etc.

Le spectre électromagnétique

Selon la fréquence les ondes électromagnétiques
nous apparaissent sous différentes formes
(mais il s'agit toujours bien des mêmes !)



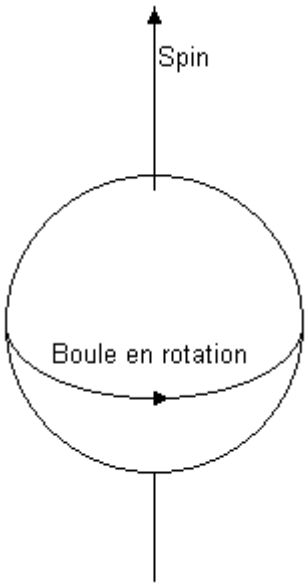
On voit que le spectre est extrêmement large (comparé aux spectres sonores habituels par exemple) et que les fréquences peuvent être incroyablement élevées et les longueurs d'ondes extraordinairement courtes.

La description de l'électromagnétisme et de la gravité (relativité générale) montre bien les grandes différences entre les deux.

Le spin

Dans la suite, nous aurons souvent l'occasion de parler du spin, le moment est donc venu d'expliquer ce que c'est.

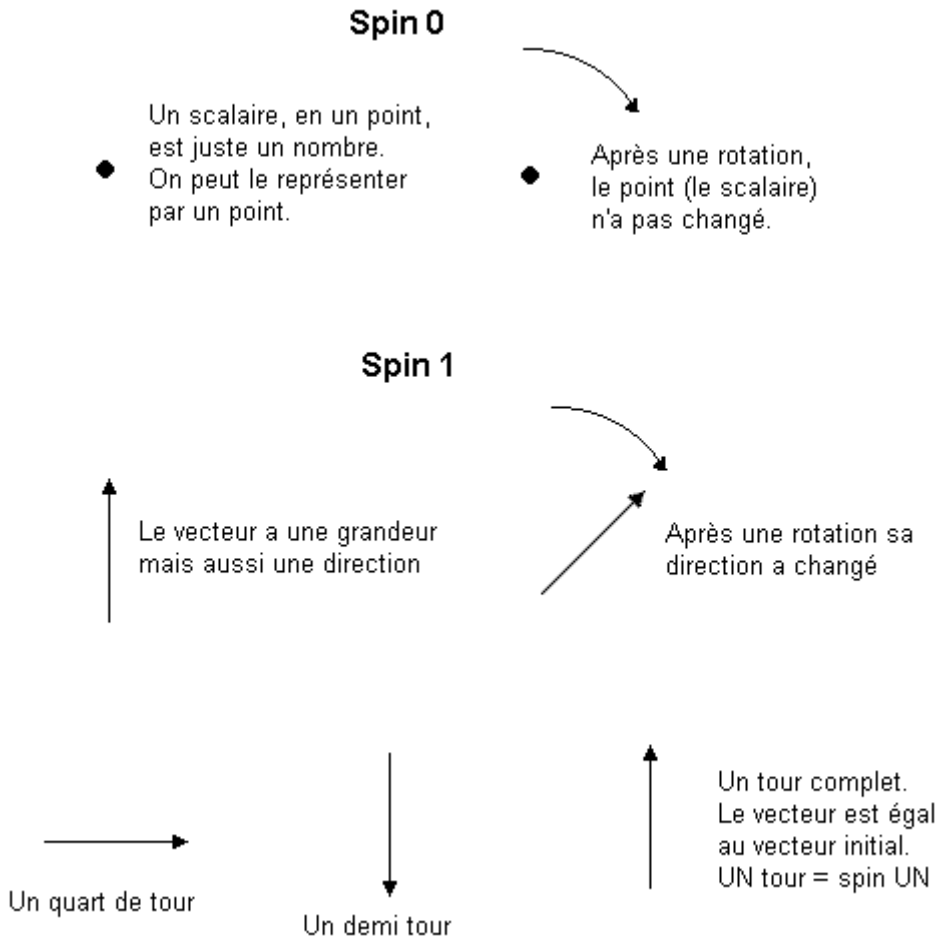
Un objet tel qu'une particule (ou une boule) peut être en rotation. On dit qu'il possède un spin.



Mathématiquement, on caractérise le spin par les propriétés d'un système physique lorsqu'on effectue une rotation.

Prenons un champ scalaire. En un point donné sa valeur est un simple nombre. Lorsqu'on effectue une rotation un point reste un point et un champ scalaire n'a pas de spin. On dit que son spin vaut zéro.

Pour un champ vectoriel, la situation est différente. Comme un vecteur a une direction, une rotation change sa direction. Lorsque l'on effectue un tour complet, le vecteur revient à sa direction originelle. On dit qu'un champ vectoriel a un spin égal à un.



Le champ électromagnétique étant de type vectoriel, il a un spin égal à un.

Lorsque l'on étudie la relativité générale, on constate que les équations autorisent aussi des ondes. Elles sont appelées ondes gravitationnelles et sont des vagues dans l'espace-temps ! Ces ondes peuvent aussi être décrites par un champ (champ de gravitation). Lors d'une rotation, il suffit d'un demi-tour pour que les valeurs de ce champ reviennent à leurs valeurs initiales. Ceci ne peut se produire qu'avec des tenseurs et le champ de gravitation est effectivement un champ tensoriel. On dit que le champ de gravitation a un spin égal à deux.

Notons que tous les champs tensoriels n'ont pas un spin égal à deux. Par exemple le champ électromagnétique est un champ tensoriel de spin 1.

On peut aussi avoir un spin 3 (un tiers de tour), etc. Nous en retoucherons un mot plus tard.

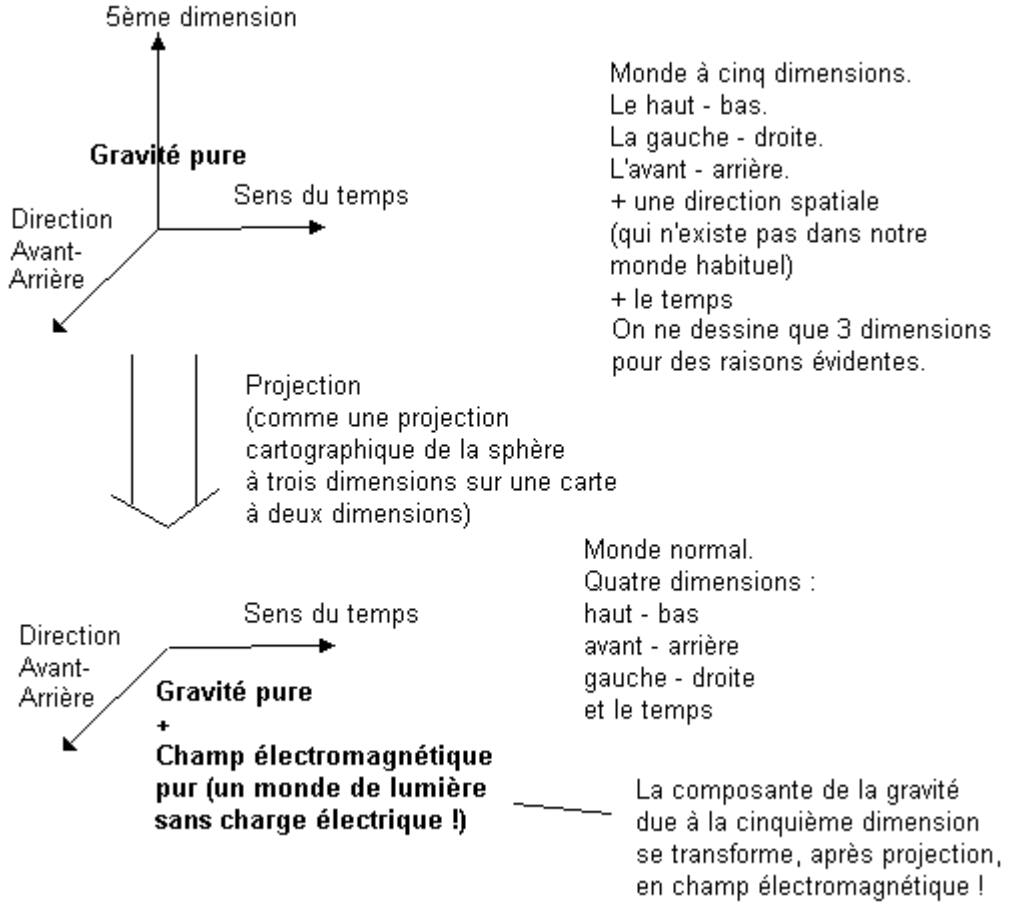
La théorie de Kaluza

Revenons à notre histoire. Quel est donc l'idée que Kaluza a eu ?

Il est parti de la relativité générale. Il a d'abord simplifié le problème en ignorant la matière. Les équations de la relativité générale sont alors un peu plus simples. Elles décrivent un monde de "pure" gravitation. Un monde remplit d'ondes gravitationnelles.

Mais il ne se contenta pas de cela. Il imagina un monde à cinq dimensions ! C'est à dire quatre dimensions d'espace et une de temps. Bien sûr le monde réel n'a que trois dimensions d'espace (de gauche à droite, d'avant en arrière et de haut en bas), mais nous y reviendrons.

Et là, quelque chose d'extraordinaire se passe. Lorsque l'on détaille les équations de cette gravité à cinq dimensions, puis que l'on ignore la dépendance à la dimension supplémentaire, on voit brusquement apparaître la relativité générale habituelle, à quatre dimensions, et le champ électromagnétique !



Kaluza avait donc montré que non seulement le champ de gravitation et le champ électromagnétique pouvaient être unifiés mais qu'en plus ce dernier champ n'était que la conséquence du champ de gravitation (à cinq dimensions) !

Mais la théorie de Kaluza pose tout de même quelques problèmes. Enumérons-les.

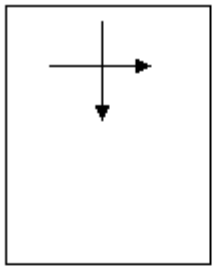
1. Pourquoi supprimer arbitrairement la dépendance à la cinquième dimension ? Bien sûr le monde réel a quatre dimensions, mais on est partit d'un monde à cinq dimensions, toutes sur le même pied d'égalité. Pourquoi changer cela en cours de route ?
2. Si le monde a cinq dimensions, comment se fait-il que nous n'en percevions que quatre ? Où est la cinquième ?
3. On a omit de signaler que lors du calcul on voit apparaître non seulement le champ de gravitation et le champ électromagnétique mais en plus on voit apparaître un champ scalaire (appelé dilaton). Sa présence, à l'époque, était gênante car ce champ n'était pas attendu ! De plus, dans le monde qui nous entoure, nous n'observons pas un tel champ.

2.2. Klein

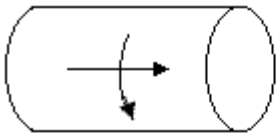
Une étape supplémentaire fut franchie par Oskar Klein en 1926. Celui-ci se demanda si certaines dimensions ne pouvaient pas être enroulées sur elles-mêmes.

Comment est-ce possible ? Tout d'abord, étant donné que l'espace-temps est courbe, il n'est pas étonnant d'avoir une dimension très courbée. Mais à quoi cela correspond-t-il ?

Pour simplifier, raisonnons d'abord sur un cas beaucoup plus simple. Imaginons un espace à deux dimensions, comme une feuille. Imaginons ensuite qu'une des deux dimensions s'enroule sur elle-même.

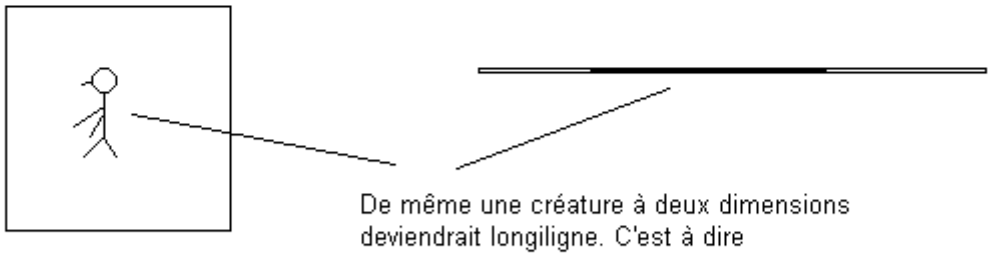


Une feuille a deux dimensions.



Une des dimensions enroulées.
Le résultat est un tube mais il
a toujours deux dimensions
(deux directions sont nécessaires
pour repérer les points sur la
surface)

Vu de très très loin, le tube prend l'aspect d'un fil. On ne perçoit plus
qu'une seule direction. C'est à dire une seule dimension.
La dimension enroulée est imperceptible.

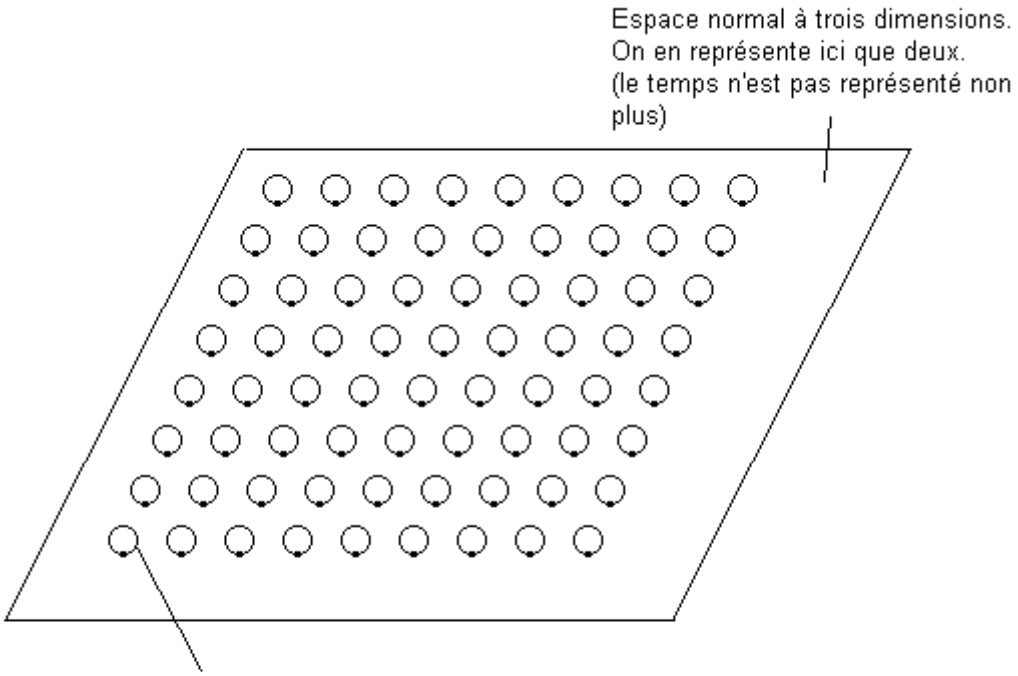


De même une créature à deux dimensions
deviendrait longiligne. C'est à dire
elle devient une créature à une dimension
incapable de percevoir la deuxième
dimension qui la compose.

Le monde réel est à quatre dimensions
(trois d'espace) et la théorie de Kaluza - Klein
en rajoute une cinquième.
C'est un peu plus difficile à dessiner !
Mais le principe est le même.

Comme on le voit, si la dimension est enroulée sur une très petite distance, vu de loin ou vu par quelqu'un beaucoup plus grand que cette distance, l'espace prend l'apparence d'un fil à une seule dimension. L'enroulement peut donc rendre cette dimension inobservable ou difficile à observer.

Comment cela se traduit-il pour notre espace ordinaire à trois dimensions ? Il faut imaginer qu'en plus de la hauteur, de la largeur et de la longueur il existe une direction supplémentaire, dans une quatrième dimension d'espace, mais qui est enroulée sur une très petite distance. Si nous pouvions nous déplacer dans cette direction nous reviendrions presque immédiatement à notre point de départ car l'enroulement est très serré. En chaque point de l'espace ordinaire il existe une direction supplémentaire formant une petite boucle. On peut facilement le dessiner pour un espace "normal" a deux dimensions plus une dimension enroulée.



On suppose donc que la cinquième dimension est une dimension d'espace enroulée sur une très petite distance. Typiquement, la distance considérée est de l'ordre de la longueur de Planck : 10^{-35} mètre, c'est à dire 0,00000000000000000000000000000000001 mètre ! Nous reviendrons plus tard sur cette distance et son explication.

L'étape suivante est identique à Kaluza. On suppose une pure gravité à cinq dimensions et on résout les équations. C'est un peu plus compliqué à cause de cette dimension enroulée et du fait que l'on prend en compte la totalité des dimensions. La cinquième dimension n'est plus éliminée (en projetant sur un espace-temps à quatre dimensions) comme avec Kaluza.

Qu'est-ce que cela apporte ?

1. La théorie est plus "saine" puisqu'on n'élimine plus arbitrairement une des dimensions en cours de calcul.
2. La cinquième dimension est inobservable, en pratique, car elle est enroulée sur elle-même sur une très petite distance.

3. La théorie prévoit toujours quatre champs : la gravitation, le champ électromagnétique et le champ scalaire (dilaton). La gravitation et le champ électromagnétique se manifestent dans les quatre dimensions "normales".
4. La théorie explique la quantification de la charge. C'est à dire qu'elle explique pourquoi la charge électrique est toujours observée comme un multiple entier d'une charge élémentaire. Cette charge élémentaire est très petite, c'est la charge portée par l'électron (voir par exemple le nombre d'électrons qui forment un courant de un ampère pour avoir une idée de la petitesse de cette charge).
5. Le champ scalaire joue un rôle important. Il garantit la cohérence de la théorie (si on le supprime arbitrairement, la théorie devient absurde c'est à dire inconsistante). Il intervient dans la quantification de la charge. Il est en pratique inobservable et intervient également dans la prédiction de particules supplémentaires.

Mais il reste tout de même quelque chose en plus par rapport aux théories classiques. La théorie prédit des particules supplémentaires de masse très grande. Cette masse est énorme (typiquement la masse de Planck, nous y reviendrons). Tellement grande qu'il est exclut de pouvoir créer de telles particules et donc de les observer !

2.3. La période d'oubli

Puis la théorie de Kaluza - Klein fut oubliée ou ignorée pendant environ… 40 ans !

Que s'est-il passé ?

Il faut tout d'abord bien comprendre que l'idée d'une cinquième dimension était quelque peu exotique. Cette idée n'améliorait pas la théorie de la gravitation ni l'électromagnétisme. Elle permettait seulement l'unification des deux. Or on n'accepte pas facilement une idée exotique si l'on n'a pas une raison impérieuse.

Ensuite, rappelons que la théorie de Kaluza - Klein n'est pas libre de problèmes. Le champ de dilaton était, tout particulièrement à l'époque, quelque chose de difficile à comprendre et à admettre.

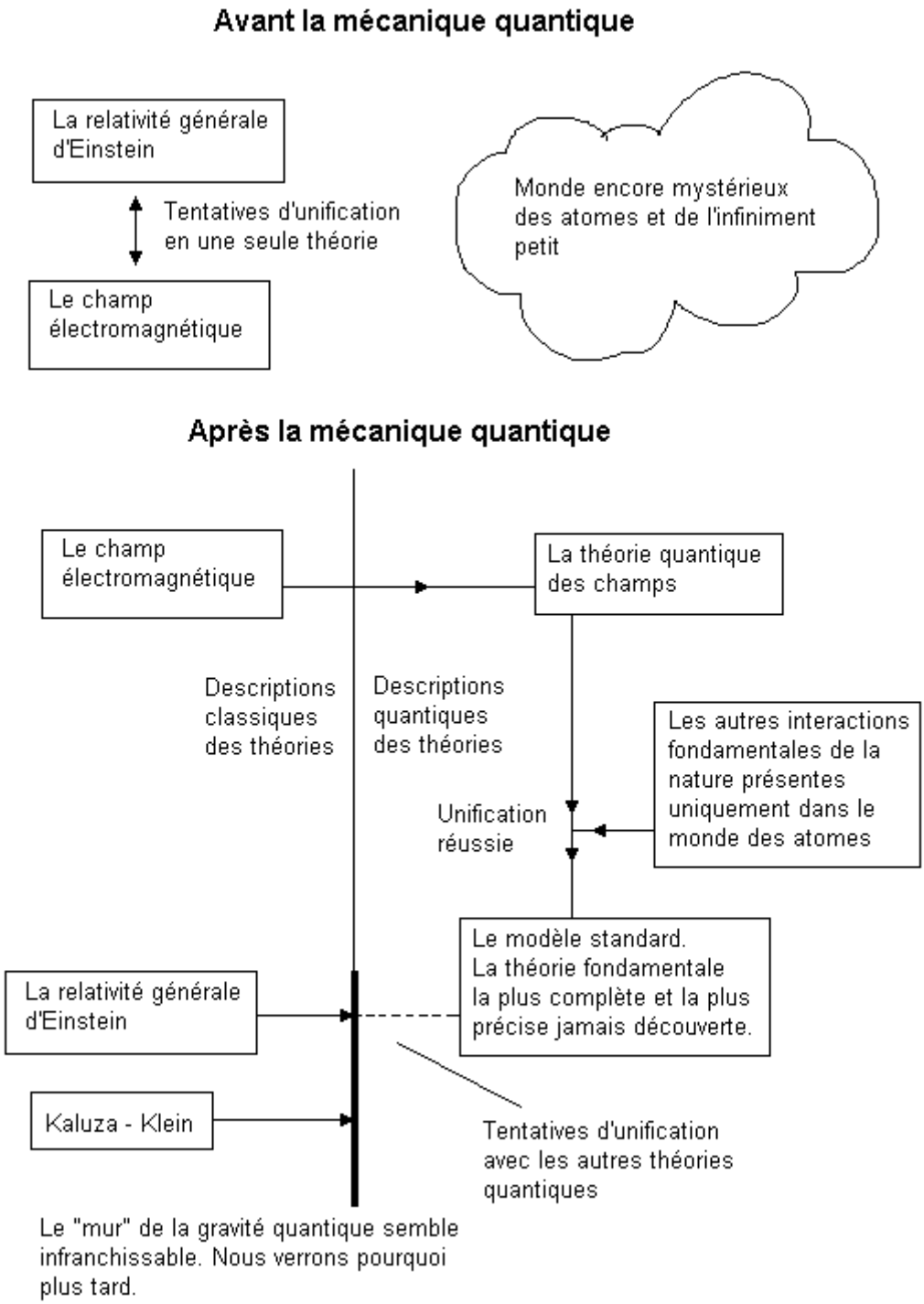
Mais le plus important ne réside probablement pas là. Après tout, on aurait pu considérer ces aspects curieux comme des artefacts mathématiques ou théoriques, sans conséquence physique réelle et pousser plus loin la théorie. En réalité, la grande coupable est la physique quantique.

Comme je l'ai dit au début du chapitre 2.1, à cette époque la mécanique quantique était en pleine gestation. Nous nous pencherons bientôt sur le monde mystérieux de la physique quantique. Celle-ci obtint très vite des succès considérables. Après avoir réussi à expliquer les comportements des atomes, les physiciens théoriciens réussirent rapidement à quantifier le champ électromagnétique.

Par la suite, la physique quantique améliora encore son palmarès en unifiant d'autres forces fondamentales (nous y reviendrons), mais le mal était déjà fait.

Tout d'abord, le "classique" n'avait plus vraiment la cote. On s'intéressait beaucoup plus aux théories quantiques. Or la théorie de Kaluza - Klein est une théorie tout ce qu'il y a de plus classique. Pire, la gravitation pose d'énormes problèmes en physique quantique (comme nous le verrons) et la théorie de Kaluza - Klein, basée sur la gravitation, n'échappe pas à ce problème.

Ensuite, puisque le champ électromagnétique était quantifié, le problème de l'unification s'était déplacé. Le but n'était plus d'unifier l'électromagnétisme et la gravitation mais plutôt d'unifier la gravitation avec l'ensemble de la théorie quantique.



Mais l'histoire n'a pas dit son dernier mot. Et les idées de Kaluza - Klein finirons par ressurgir.

3. Les théories quantiques

3.1. La mécanique quantique

Voici venu le moment de nous plonger dans les eaux profondes de la mécanique quantique. J'espère que vous n'avez pas trouvé trop difficile ce qui a précédé car la mécanique quantique est infiniment plus compliquée et plus déroutante que la relativité générale ou l'électromagnétisme !

Il va de soit que je n'expliquerai la mécanique quantique que dans les grandes lignes. Je passerai sur beaucoup de détails, en particulier les détails mathématiques (bien sûr). C'est un sujet extrêmement vaste, résultat de trois quarts de siècle de recherches théoriques par un très grand nombre de physiciens. Ses succès et ses applications sont innombrables. Un livre entier ne pourrait couvrir complètement le sujet. Nous toucherons seulement aux aspects les plus fondamentaux, nécessaires pour comprendre de quoi on parle, et aux aspects qui nous conduirons progressivement à la théorie des cordes.

La mécanique quantique est le monde de l'infiniment petit. Le monde des atomes et des particules atomiques. C'est un monde très particulier où les lois classiques, celles qui gouvernent notre quotidien, ne s'appliquent pas. Au contraire, les concepts de la mécanique quantique sont étranges et déroutants. Nous aurons l'occasion de le voir.

Postulats de base

En physique quantique, l'état d'un système physique est représenté par un "état quantique" ou un "vecteur d'état" noté $|\alpha\rangle$ pour représenter un état α.

Les postulats de la mécanique quantique disent que les états sont "linéaires". C'est à dire que toute combinaison d'états est encore un état (mathématiquement, c'est de là que vient leur nom de vecteur). Par exemple si $|\alpha_1\rangle$ et $|\alpha_2\rangle$ sont deux états possibles, alors $|\alpha_1\rangle + |\alpha_2\rangle$ est également un état possible pour le système. Attention, les états ne sont pas de simples nombres. Ils peuvent être très compliqués.

Lorsqu'un système physique est dans un état, il peut malgré tout être dans un autre état ! Par exemple $\langle\alpha_2|\alpha_1\rangle$ est appelée "l'amplitude" d'être dans l'état α_2 si le système est dans l'état α_1.

L'amplitude n'est pas un simple nombre. C'est un nombre que les mathématiciens appellent nombre "complexe". En l'élevant au carré (plus exactement en le multipliant par son conjugué) on obtient un nombre classique. Ce nombre est interprété comme la probabilité que le système physique soit dans l'état α_2 s'il est dans l'état α_1.

Le monde quantique est donc un monde de probabilités. Et cette probabilité ne semble pas être une conséquence de notre imprécision ou de notre ignorance mais semble bien être une propriété intrinsèque de la nature.

Que représente exactement un état ? C'est un peu compliqué mais on peut prendre un cas simple. Supposons que l'on ait une particule isolée, disons un électron. Cet électron peut se situer en un endroit précis, disons la position x. Un tel état sera noté $|x\rangle$. Rappelez-vous que l'on peut superposer les états. Ainsi l'état $|x_1\rangle + |x_2\rangle$ représente une situation où l'électron peut se situer à deux endroits ! Il peut donc être dans un état quelconque $|x\rangle$. Dans ce cas, l'amplitude (et de là, la probabilité) qu'il soit situé (observé, lors d'une mesure) à la position x est $\langle x|\alpha\rangle$. Comme cette valeur peut être donnée pour chaque position, on obtient ce qui est appelé une fonction d'onde. C'est une

fonction qui donne une valeur (une amplitude) pour chaque position. On la note traditionnellement ψ ou $\psi(x)$.

Historiquement, c'est Schrödinger qui trouva, à l'aide de raisonnements intuitifs, la première équation applicable à la fonction d'onde de l'électron plongé dans le champ électrique du noyau de l'atome. Elle permit une avancée considérable dans la compréhension des atomes.

Les règles que nous avons données ci-dessus autorisent déjà des choses extraordinaires. Même sans manipuler d'équations. On l'a vu avec l'électron qui peut se trouver en deux endroits.

Par exemple, un électron est-il une onde ou une particule (comme une bille dure) ? L'idée de particule souffre quelque peu puisque l'électron peut se retrouver complètement "dispersé" un peu comme une onde. Oui mais si on effectue une mesure, on observera l'électron en un endroit précis et uniquement à cet endroit, ce qui n'est pas possible pour une simple onde ! Pour en savoir plus, voir LA LUMIERE : Onde ou particule ?.

On peut modifier un état. On peut le transformer en un autre état. Cette opération se fait avec ce qui est appelé un "opérateur". Un opérateur est un objet mathématique qui effectue une opération mathématique sur l'état. Ce n'est pas toujours une simple multiplication ou une simple addition, cela peut être une opération beaucoup plus compliquée (que je ne décrirai évidemment pas ici).

Certaines valeurs d'un système sont appelées des observables. C'est à dire des valeurs que l'on peut mesurer. C'est le cas de la position d'une particule, de sa vitesse, de sa charge électrique, etc.

Les observables sont aussi représentés par des opérateurs. Par exemple l'opérateur "vitesse de la particule", appelons-le V. Si un système a une vitesse précise, alors l'opération sur l'état correspondant donne : $|v\rangle = V|v\rangle$. L'état ne change pas (en toute rigueur il peut être multiplié par un nombre, mais cela ne change pas les raisonnements) ! C'est aussi une manière de définir les opérateurs pour les observables (ce n'est donc pas un postulat mais une définition). Lorsqu'un tel état est inchangé par un opérateur il est appelé "état propre" ou "vecteur propre".

Pour un état quelconque $|\alpha\rangle$, on peut le décomposer (en utilisant la superposition des états) : $|\alpha\rangle = a_1|v_1\rangle + a_2|v_2\rangle + ... + a_3|v_3\rangle$. Qu'une telle décomposition soit toujours possible, fait partie des postulats de base sur les états.

Alors, comme chaque vitesse "pure" est un état propre, les valeurs $a_1, a_2, ..., a_3$ sont les amplitudes (et de là, les probabilités) d'observer la particule avec la vitesse $v_1, v_2, ..., v_3$. Cette suite de valeur s'appelle la quantification car dans certains états on peut très bien n'avoir que des valeurs précises et pas d'autres. On dit que la valeur est quantifiée.

Comme les opérateurs ne sont pas des nombres, ni toujours des opérations mathématiques simples, ils ne sont pas toujours "commutatif". C'est à dire que l'on ne peut pas toujours les intervertir. Si j'ai un opérateur A et un opérateur B alors on peut très bien avoir $A \cdot B \neq B \cdot A$. C'est à dire que si j'agis sur un état avec A puis avec B le résultat sera différent si j'agis avec B puis avec A. On dit dans ce cas qu'ils ne "commutent" pas. Si le résultat est identique, alors on dit qu'ils commutent. L'opération $[A, B] = A \cdot B - B \cdot A$ est appelée un commutateur. S'il est non nul, cela signifie que les opérateurs ne commutent pas.

Par exemple, en mécanique quantique, les opérateurs vitesse et position ne commutent pas ! Une situation sans équivalent classique où mesurer position puis vitesse est identique à mesurer vitesse puis position. Alors que la non-commutation implique que l'ordre des mesures peut avoir une importance en mécanique quantique.

Il est évident que je n'ai pas expliqué la forme exacte de ces opérateurs position et vitesse, mais impossible d'entrer dans tous ces détails sans faire des mathématiques. Malheureusement.

Le principe d'incertitude

On peut montrer que si deux opérateurs (observables) ne commutent pas alors il est impossible d'avoir un état ayant une valeur précise et unique pour les deux observables à la fois (cela est lié à l'ordre des mesures expliqué ci-dessus). C'est le principe d'incertitude qui est une des conséquences les plus extraordinaires de la physique quantique. Nous allons en donner quelques exemples et applications.

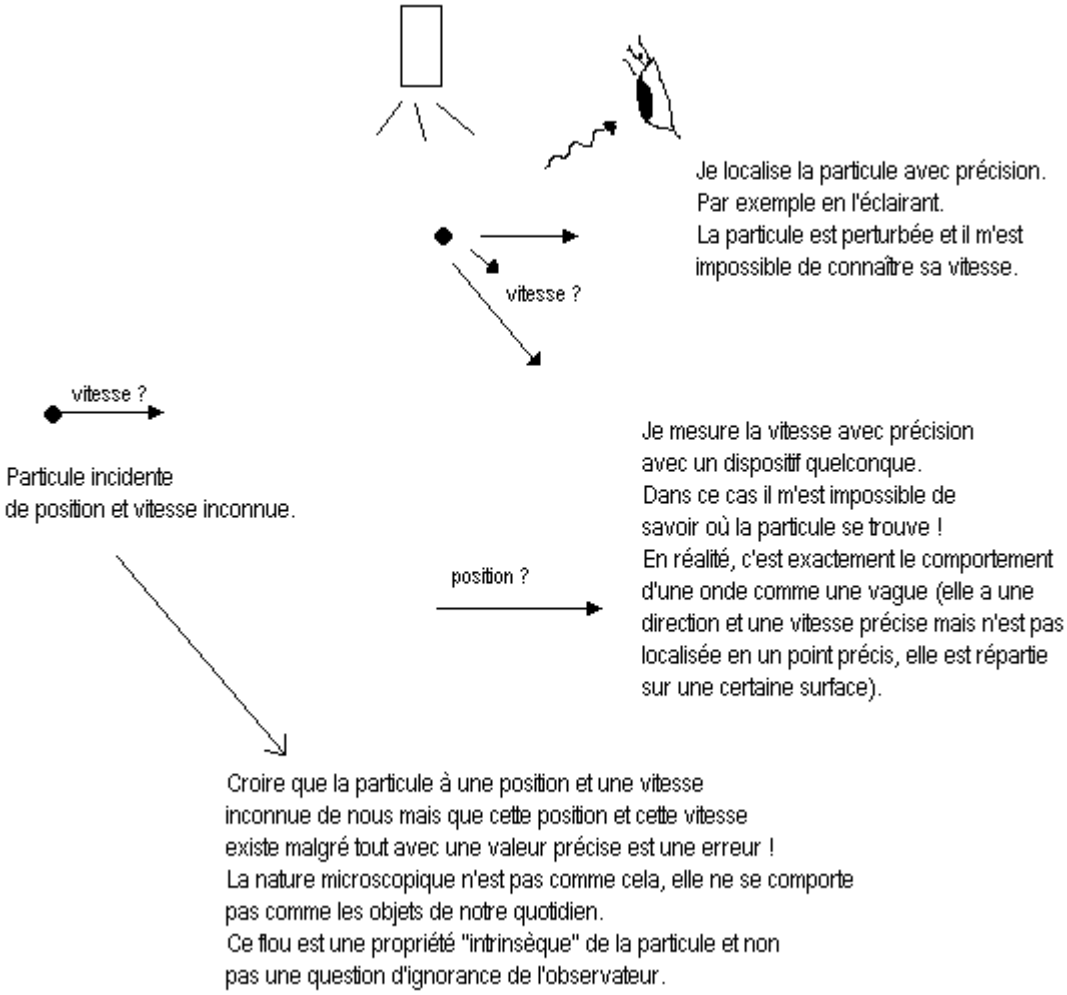
La vitesse et la position sont deux opérateurs qui ne commutent pas. Ils sont donc sujet au principe d'incertitude. Celui-ci dit qu'il est impossible de mesurer avec une précision aussi grande que voulue à la fois la vitesse et la position d'une particule. Plus précisément, si Δx est l'incertitude sur la position (dans une direction donnée) et Δv l'imprécision sur la mesure de la vitesse (dans la même direction). Alors on aura :

$$\Delta x \Delta v \geq \frac{h}{2\pi m}$$

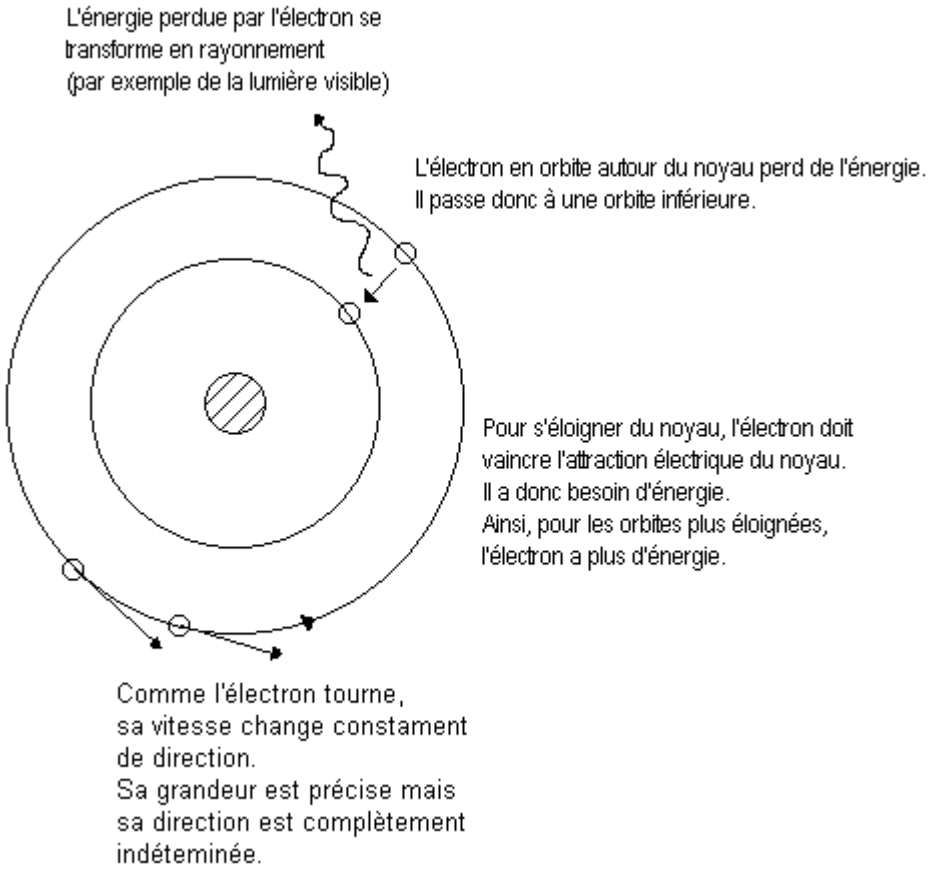
où π est le nombre pi (3.1416), m est la masse de la particule et h est une constante appelée constante de Planck. Mathématiquement, la position et la vitesse sont dites des variables "conjuguées" (au sens hamiltonien, je n'expliquerai pas ici ce que cela signifie, c'est un peu compliqué et hors sujet). Toute paire de variables conjuguées obéit à cette formule.

La constante de Planck est extrêmement petite. Cela explique que cet effet est impossible à détecter à notre échelle. Par contre, la masse des électrons est extrêmement petite aussi. Donc la fraction ci-dessus est notable pour un électron et l'effet de cette incertitude est important.

Si je mesure avec une très grande précision la position d'un électron, alors sa vitesse sera très incertaine. Inversement, si sa vitesse est mesurée très précisément, alors il peut se trouver à peu près n'importe où !

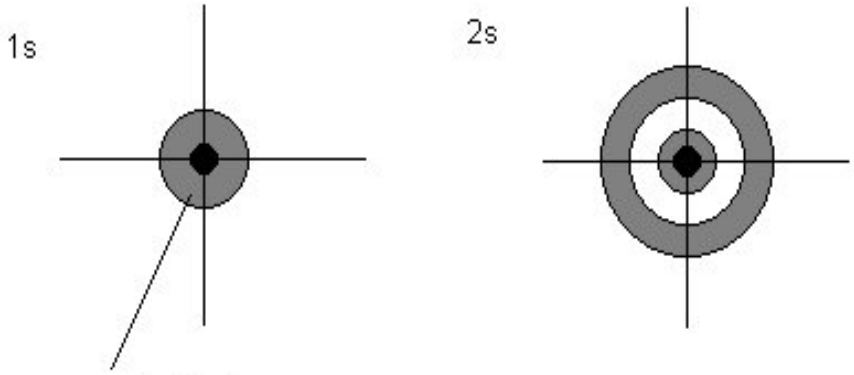


Par exemple, autour d'un atome l'électron a une énergie très précise (comme le montre l'équation de Schrödinger). Plus exactement, il existe plusieurs états d'énergies différentes (on dit des niveaux). Mais à l'équilibre l'électron se trouve sur le niveau d'énergie le plus bas. Puisque son énergie est précise, alors sa vitesse aussi. En fait, pas tout à fait, car l'électron tourne et sa vitesse change donc constamment de direction, mais elle reste constante en grandeur.

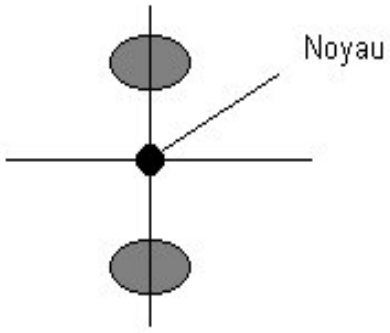


Cette vitesse précise en grandeur (mais pas en direction) fait que la position de l'électron est très incertaine autour de l'atome. Il forme une espèce de petit "nuage" autour de lui.

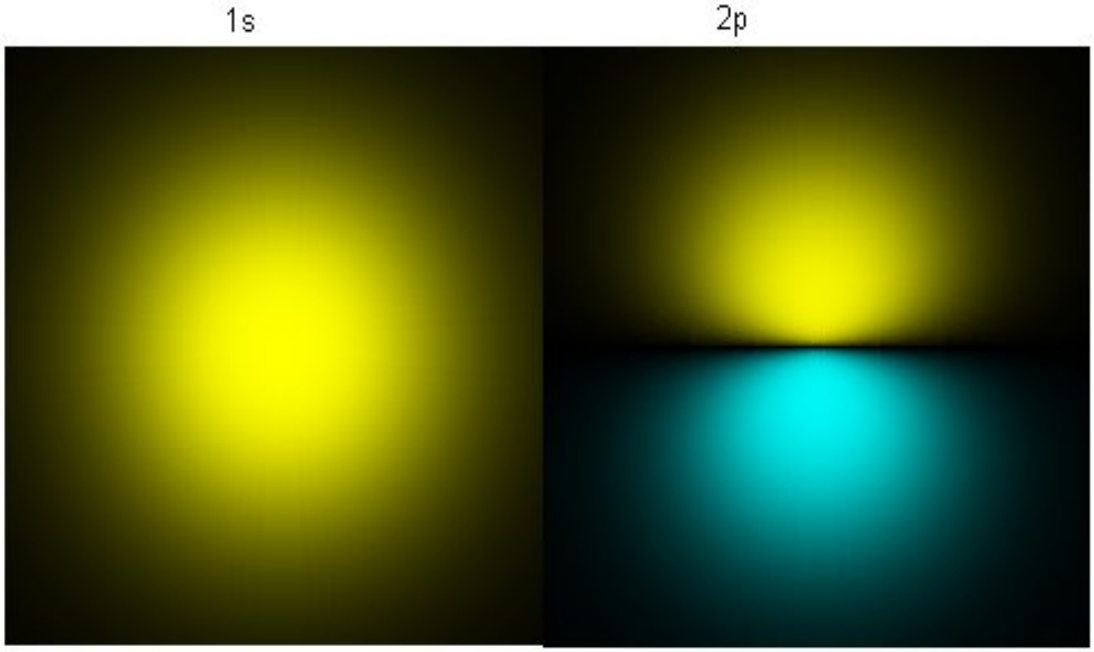
Visualisation (approximative) des fonctions d'ondes pour quelques unes des solutions (régions où l'amplitude de présence est importante).



L'électron est distribué de manière imprécise autour du noyau. Il forme une espèce de "nuage".



Vue plus précise, non pas de l'électron !, mais de sa distribution autour de l'atome



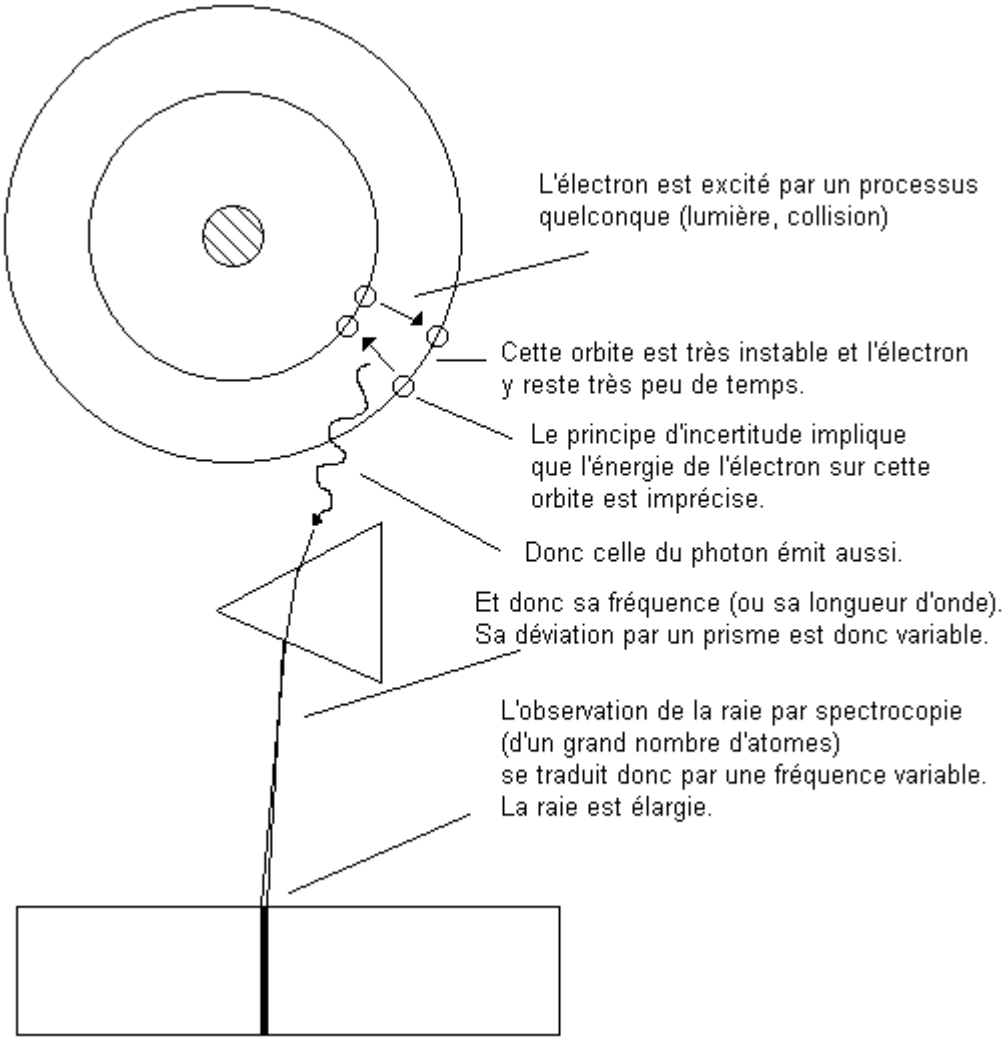
La représentation sous forme d'orbite circulaire est très imagée et moins précise que les orbitales ci-dessus (c'est le moins que l'on puisse dire !) mais cette représentation est très pratique. Disons que chaque orbite circulaire représente un niveau d'énergie de l'électron plutôt qu'une orbite réelle.

D'autres variables (opérateurs) présentent ce phénomène.

Ainsi le temps et l'énergie sont des variables conjuguées. On a donc :

$$\Delta t \Delta E \geq \frac{h}{2\pi m}$$

Que signifie en pratique cette formule ? Par exemple, si une particule reste dans un état donné pendant un temps très court (et donc forcément très précis). Alors l'énergie de cette particule sera très imprécise. Cela peut se manifester avec les atomes.



On voit que nous associons fréquence de la lumière (sa couleur, comme on l'a vu avec le spectre électromagnétique) avec l'énergie émise par l'atome. En réalité, en mécanique quantique les valeurs sont quantifiées, comme on l'a vu. Et la lumière ne peut être émise que par "paquets" d'énergie, appelés quanta ou photons. La relation entre la fréquence de la lumière est : $E = h\nu$. Où on retrouve la fameuse constante de Planck. Historiquement, ce fut la première manifestation quantique qui fut découverte (par Planck avec le rayonnement du corps noir, puis confirmée et généralisée par Einstein avec l'effet photo électrique).

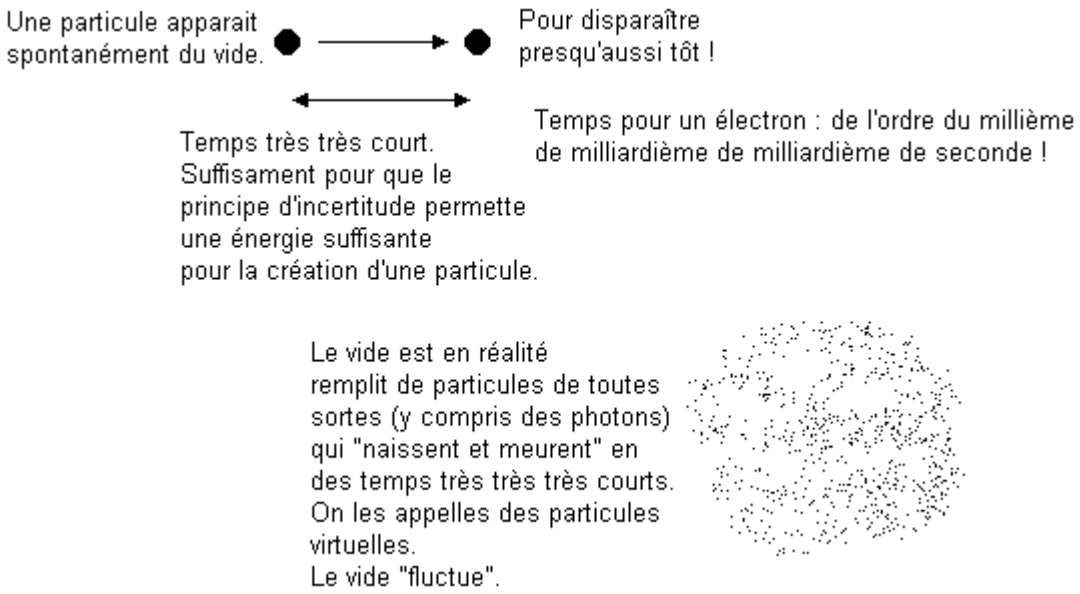
Même le spin des particules est quantifié. On trouve ainsi des particules de spin 0, 1, 2, … comme pour un champ. Classiquement, une particule peut tourner n'importe comment. Ce n'est pas le cas en mécanique quantique !

Pour voir quelques autres effets et conséquences de la mécanique quantique, je conseille de voir Mécanique quantique et relativité où l'on parle de deux phénomènes qui semble (en apparence) contredire la relativité.

Les fluctuations du vide

Une conséquence fantastique découle de l'incertitude sur le temps et l'énergie et de la relativité. Supposons que je prenne le vide. Supposons aussi que je regarde ce qui se passe en un endroit donné pendant un temps très court. Alors, le principe d'incertitude nous dit que l'énergie de cet état (le vide !) est très imprécise. Or la relativité dit que l'énergie c'est aussi de la masse, donc des particules. Donc, pendant ce temps très court des particules peuvent apparaître spontanément du vide ! On les appelle des particules virtuelles car elles disparaissent très vite (sinon la durée de leur existence ne serait plus très courte).

Donc le vide doit être remplit de plein de "fluctuations". Des particules de toute nature doivent apparaître et disparaître constamment.



Ces fluctuations ne sont pas quelque chose d'imaginaire car leurs effets sont bien réels ! Pour en savoir un peu plus, voir Les fluctuations du vide.

Les bosons et les fermions

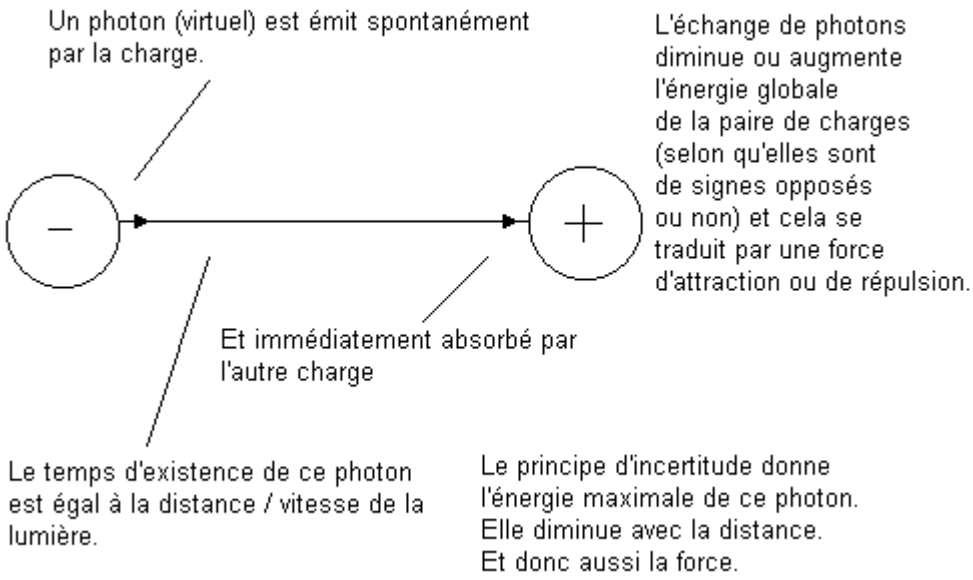
Le monde contient deux types de particules. Les particules de spin entier (0, 1, 2, …) appelés les bosons et les particules de spin demi-entier (1/2, 3/2, …) appelés les fermions.

Par exemple les photons sont des bosons car ils ont le spin 1 (ce qui est identique à la valeur du spin du champ électromagnétique ce qui n'est pas étonnant). Et les électrons sont des fermions car ils ont le spin 1/2.

Que signifie un spin demi-entier ? On a vu que le spin caractérisait les propriétés sous les rotations. Ainsi, pour un champ de spin 1, les vecteurs reprennent leur position initiale après un tour. Pour le spin deux, un demi-tour suffit. Pour un spin 1/2, il faut effectuer deux tours ! Voilà qui est extraordinaire et n'a à nouveau aucun équivalent classique. Quel objet classique pourrait bien avoir la

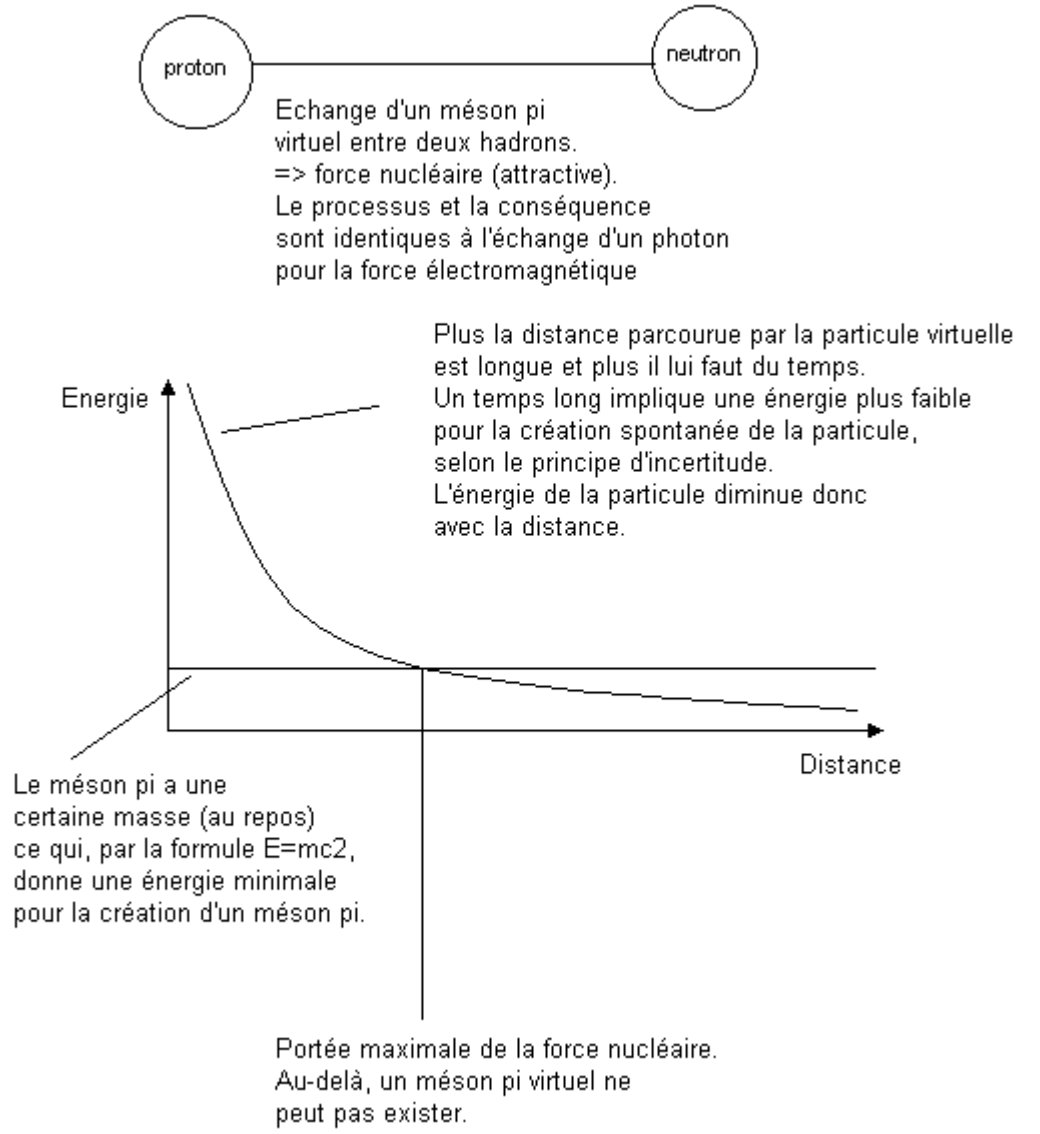
propriété de ne pas retrouver son état original après un tour complet ? Aucun ! Cette particularité à des effets curieux comme d'empêcher deux électrons d'être exactement dans le même état (autour du même atome, avec la même énergie, etc.) alors que pour des bosons c'est possible (dans un rayon laser, tous les photons ont même direction, même fréquence, etc.).

La théorie montre que la matière habituelle (les atomes) est constituée de fermions tandis que les bosons sont responsables des forces. Par exemple, le photon est responsable de la force électromagnétique.



Si le boson vecteur de force est sans masse propre, alors la "portée" de la force est infinie. Sinon, la force a une portée limitée. Par exemple, le photon n'ayant pas de masse propre (comme on l'a vu), l'électromagnétisme agit à grande distance. Par contre, les forces nucléaires qui dérivent de l'interaction forte (nous en parlerons plus tard) et qui ont un vecteur massif n'agissent qu'à courte distance : au sein du noyau atomique.

Le lien entre masse du boson et portée de la force est lié au principe d'incertitude sur le temps et l'énergie. Le boson qui est échangé est comme une particule virtuelle, il est créé (émis par une particule) puis détruit (absorbé par l'autre). Cet échange prend un certain temps puisqu'il ne peut pas dépasser la vitesse de la lumière. Le principe d'incertitude dit que l'incertitude sur l'énergie est d'autant plus faible que ce temps est grand. Or pour créer une particule massive, il faut au moins l'énergie correspondant à sa masse propre (on a vu que la masse c'est de l'énergie). Donc, pour une particule massive le temps d'échange est limité et donc aussi la distance. Par contre, pour un photon pas de limite. Si la distance est grande, l'énergie "disponible" est faible mais un photon peut avoir une énergie aussi petite que l'on veut, il suffit que sa fréquence soit petite. Cela explique aussi qu'à grande distance la force électromagnétique est moins grande puisque les énergies en jeu sont plus faibles.



3.2. La théorie des champs

Pour traiter de la création (ou de la destruction) des particules, les formalismes que nous avons vus tel la fonction d'onde ne conviennent pas. On a vu que la fonction d'onde décrivait l'amplitude pour une seule particule.

On peut avoir des états, et de là une fonction d'onde plus complexe, pour deux particules. Mais il est assez difficile de manipuler des états à nombre de particules variables. Pire encore, je peux avoir un état qui décrit un système avec une particule et un autre état qui décrit le système avec deux particules. Je peux superposer les états et avoir un état qui est à la fois à une et deux particules !

De plus les théories doivent être relativistes car la création des particules, transformation d'énergie en masse, est un effet typiquement relativiste (on parle ici, bien sûr, de relativité restreinte). Ce qui complique un peu les choses.

La solution est venue des champs. L'idée étant : et si on pouvait représenter, par exemple, les électrons par un champ. Tout comme on l'a vu pour la lumière qui est composée du champ électromagnétique en physique classique et de photons en physique quantique. D'où l'idée de

représenter classiquement toutes les particules par des champs puis de passer à la physique quantique. Est-ce que ça marche ? Oui, et même très bien !

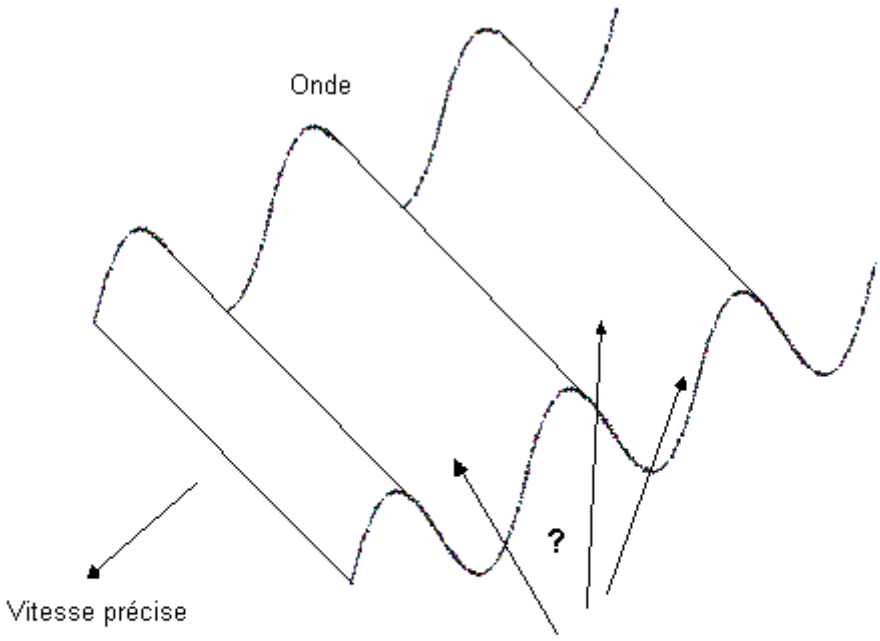
La théorie quantique des champs fut élaborée dans les années trente mais, même maintenant, c'est encore un sujet actif de recherches théoriques.

Les champs en mécanique quantique

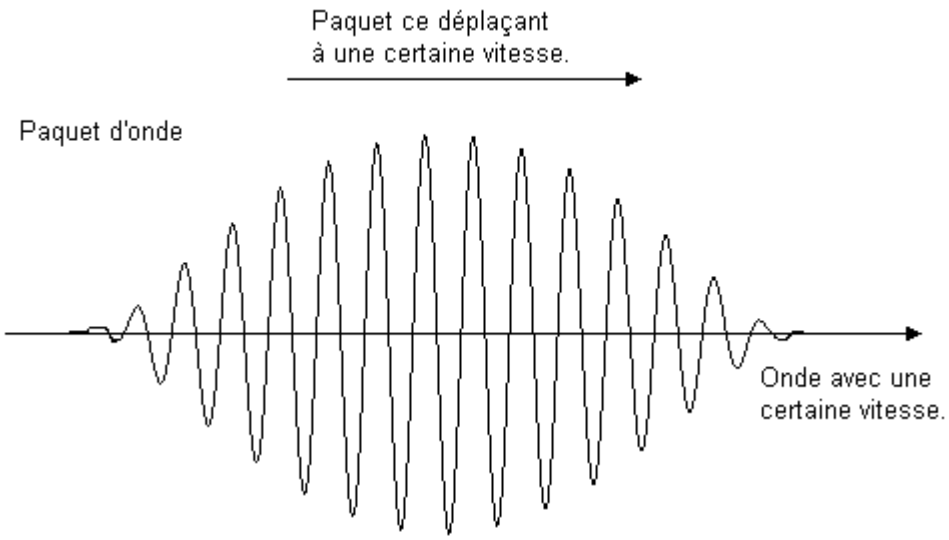
Les champs possèdent les propriétés requises. Par exemple, ils peuvent être linéaires ce qui entraîne des mécanismes de superposition comme pour les états.

On quantifie le champ en remplaçant les variables décrivant le champ par des opérateurs. On recherche les variables conjuguées (au sens mathématique de ce terme) et on exprime que les opérateurs ne commutent pas. Et le tour est joué.

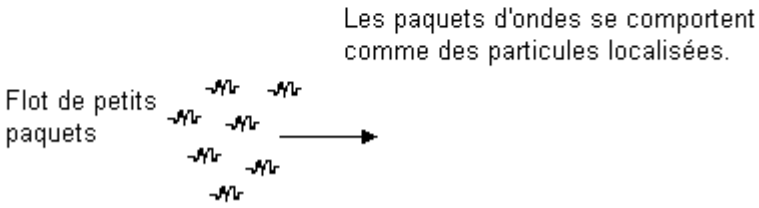
Un champ peut avoir des "modes" précis. C'est à dire avoir une fréquence de vibration précise. Après quantification, on constate que ces modes se comportent comme des états à une particule ! Comme un champ peut avoir des tas de fréquences en même temps, alors il peut correspondre à des états avec plusieurs particules. Il est même possible de calculer des opérateurs dit de création (ou de destruction) qui modifient un état en lui ajoutant une particule. Ils jouent un grand rôle dans la théorie et, on le devine aisément, dans les champs en interaction où des particules peuvent être créées.



L'onde se comporte comme la particule
avec une position incertaine.
Cela explique la dualité onde - corpuscule
des objets quantiques.



Le calcul montre que la vitesse du paquet (vitesse de groupe) est différente de la vitesse de l'onde (vitesse des "bosses" ou vitesse de phase).



Quantifier un champ pose parfois quelques difficultés techniques. C'est le cas du champ de l'électron car sa nature de fermion, c'est à dire son spin 1/2, entraîne quelques complications. Pour information, le champ "classique" de l'électron se décrit avec une équation appelée équation de Dirac et qui avait été imaginée au début de la mécanique quantique pour essayer de trouver une formulation relativiste de l'équation de Schrödinger.

C'est encore pire pour le champ électromagnétique. Sa nature "purement" relativiste (à cause de sa vitesse égale à la vitesse de la lumière) entraîne de grosses difficultés (sur lesquelles je ne m'étendrai pas ici) mais qui peuvent être surmontées au prix de quelques complications mathématiques.

Le vide

Comment se représenter le vide pour un champ ? C'est simplement l'état minimal. Il a l'énergie minimale et correspond à zéro particules. Lorsqu'on agit sur cet état avec l'opérateur de destruction on obtient zéro (et non pas l'état du vide). C'est normal, comment détruire une particule dans un état qui n'en contient aucune ?

Mais cet état du vide est un peu particulier. Les règles de non-commutation entraînent que certaines variables ne peuvent s'annuler. On a vu que deux opérateurs qui ne commutent pas sont tel que $A \cdot B \neq B \cdot A$. Mais si A s'annule (ou plutôt l'application de A sur l'état physique), alors on a forcément $A \cdot B = 0 = B \cdot A$. Donc A ne peut pas s'annuler. L'état du vide ne contient aucune particule mais il reste "quelque chose" !

Si l'on analyse l'état du vide plus en profondeur, on constate que les modes de vibrations ne s'annulent pas tout à fait. On constate rapidement qu'ils correspondent à des fluctuations du vide. Les fameuses particules virtuelles !

Comme il y a une infinité de modes (car il y a une infinité de fréquences possibles), alors le vide est remplit d'une infinité de fluctuations.

Quelle est l'énergie du vide ? Si on fait le calcul de manière abrupte, on obtient l'infini ! En réalité ce n'est pas surprenant puisqu'il y a une infinité de fluctuations. Chaque mode de vibration apporte une petite contribution et il y a une infinité de modes possibles.

Mais ce résultat n'a pas vraiment de sens. On ne peut pas observer l'énergie du vide. On peut seulement en observer les variations comme dans l'effet Casimir. Pour observer ou "extirper" l'énergie du vide il faudrait le détruire entièrement ce qui semble plutôt difficile !

Lorsque l'on mesure l'énergie d'un état, disons à une particule, on le fait toujours par comparaison avec l'énergie du vide. Dans ce cas, les énergies des états deviennent finie. Bien sûr, il est possible d'exprimer cela mathématiquement de manière précise avec disparition des infinis (on ne fait pas bêtement l'infini moins l'infini). Techniquement, on exprime les opérateurs en fonctions des opérateurs de création et destruction et on place ces derniers à droite (appelé ordre normal). Curieux que cela rende les résultats finis, n'est-ce pas ? Mais il est difficile d'expliquer pourquoi sans mathématique.

Les champs en interaction

Pour un champ simple tel que le champ électronique ou le champ électromagnétique, on a vite fait le tour. On peut juste décrire des états avec un nombre donné, et constant, de particules. On peut aussi effectuer quelques statistiques intéressantes sur les fluctuations du vide. Mais c'est à peu près tout. Mais les choses deviennent beaucoup plus intéressantes dès que l'on considère plusieurs champs ensemble. Par exemple, le champ électronique et le champ électromagnétique. C'est aussi beaucoup plus réaliste puisque des charges électriques émettent des champs électromagnétiques comme on l'a vu. Dans ce cas, les champs peuvent interagir entre eux et provoquer la création de particules, en particulier de photons car leur énergie pouvant être aussi faible que l'on veut, il est facile d'en créer.

Les champs peuvent aussi interagir avec eux même. De tels champs sont appelés champs en auto interaction.

Typiquement, dans les équations qui décrivent les champs, on rajoute simplement un terme qui décrit l'interaction puis on quantifie.

Malheureusement, si la situation est plus intéressante elle est aussi beaucoup plus compliquée. Et, sauf dans quelques cas simples, les équations ne peuvent pas être résolues de manière exacte. On a donc besoin de méthodes approchées. La plus connue et la plus puissante est fournie par la théorie des perturbations qui fut élaborée dans les années quarante.

La théorie des perturbations

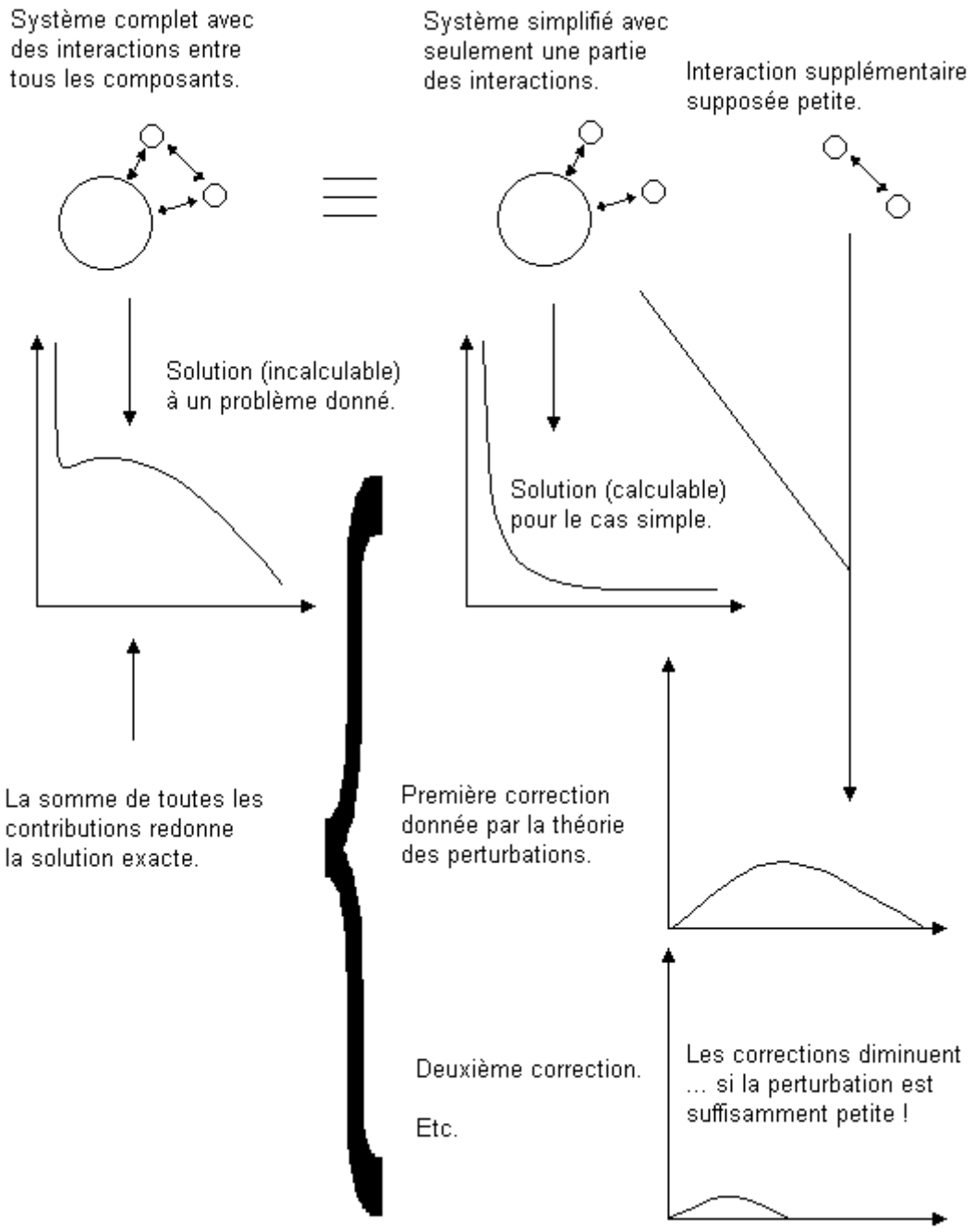
Le principe de cette théorie est simple. On considère un système physique réaliste mais insoluble de manière exacte. Puis on regarde un système physique plus simple, qui lui ressemble et que l'on sait résoudre de manière exacte. On considère alors la différence entre les deux systèmes comme une "perturbation" que l'on sait traiter mathématiquement de manière approchée.

Regardons cela d'un peu plus près. On a vu que le champ électronique et le champ électromagnétique étaient relativement simples et ils peuvent être l'objet de calculs exacts. L'interaction entre les deux, c'est à dire le terme d'interaction ajouté aux équations, est alors considéré comme une perturbation au système sans interaction. Mathématiquement, le terme d'interaction est caractérisé par une valeur, appelée constante de couplage, qui est une image de la force de l'interaction. La force électromagnétique est encore assez modérée et la perturbation peut généralement être considérée comme assez petite. Cette constante, dans le cas de l'électrodynamique (c'est le nom donné au champ électromagnétique + le champ électronique en interactions), est appelée constante de structure fine (pour des raisons historiques que je ne détaillerai pas ici) et vaut 1/136, c'est à dire environ sept millièmes. Nous appellerons cette constante k.

Appelons le résultat d'un calcul sur le système complet (champ électronique + champ électromagnétique + interaction entre les deux) R_{tot} . Appelons le résultat du même calcul sur le système simplifié (champ électromagnétique + champ électronique, sans interaction) R_0 . Ce résultat peut être calculé exactement mais pas le précédent. La théorie des perturbations fournit des outils mathématiques qui permettent d'obtenir une relation du genre :

$$R_{tot} = R_0 + kR_1 + k^2 R_2 + ...$$

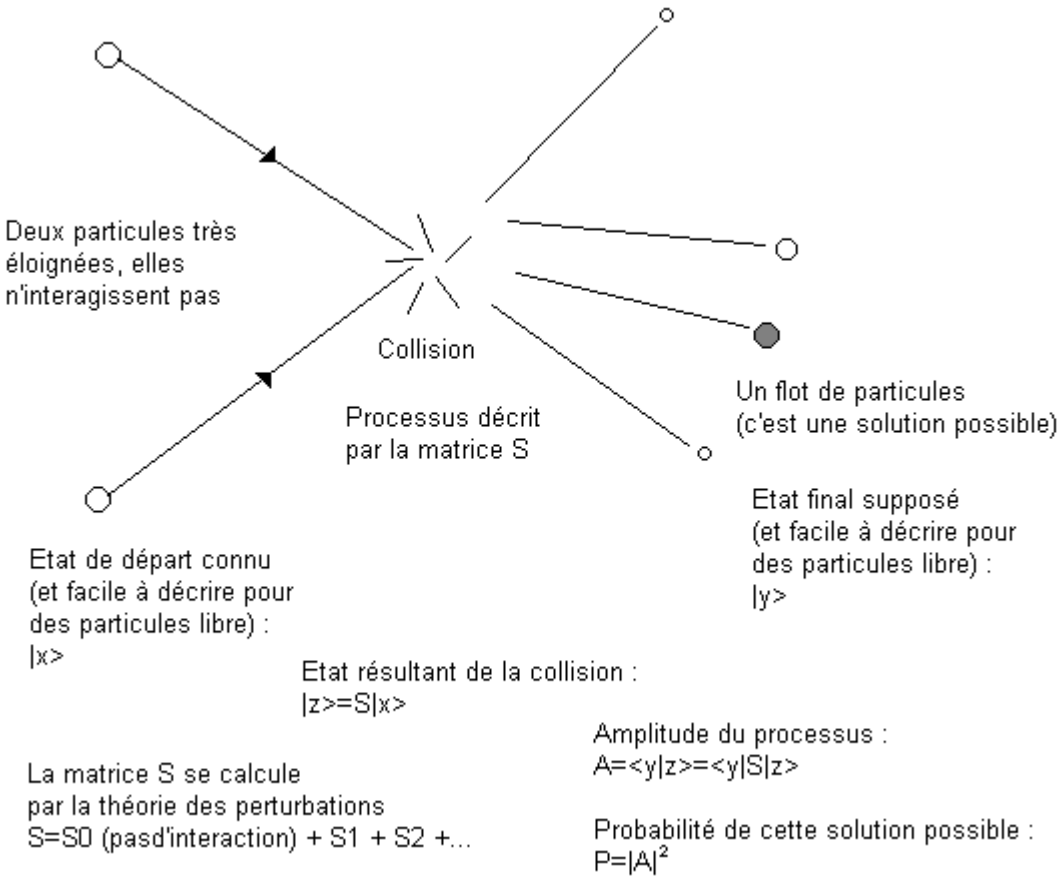
Les termes successifs R_1, R_2 , … sont des corrections successives de plus en plus précises. Les outils mathématiques donnent ce qu'il faut pour les calculer. Plus on calcule de termes et plus le résultat sera précis. Bien sûr, il faut que k soit suffisamment petit pour que les corrections successives ci-dessus deviennent de plus en plus petite. Par exemple, avec la constante de structure fine, le premier terme donne une correction de l'ordre du pour cent, le deuxième terme une correction de l'ordre du dix millième, etc.



La théorie des collisions

Une approche particulièrement fructueuse de la théorie des perturbations est son application à la théorie des collisions.

On considère initialement des particules bien séparées. Celles-ci peuvent être décrites par les champs sans interaction (c'est une approximation). Puis ces deux particules se rapprochent et entrent en collision. Dans cette phase, les champs sont en interaction. Puis les particules s'éloignent et finissent par être suffisamment loin pour être à nouveau considérées comme bien séparées et décrites par des champs sans interaction. Les champs initiaux et finaux, sans interaction, sont appelés des champs libres.



On peut caractériser la situation entrante par des champs libres à, par exemple, deux particules. Appelons cet état $|in\rangle$. De même l'état sortant sera $|out\rangle$. Ce dernier état permet une description simple (puisque ce sont des champs sans interaction) mais il reste inconnu puisque nous ne connaissons pas le résultat de la collision. Si on le connaissait, il permettrait de calculer tout ce que nous voulons savoir. Par exemple, si l'état $|3\rangle$ caractérise un état à trois particules (dont une a été créée) avec des directions et des vitesses bien précises, alors on peut calculer l'amplitude (et la probabilité d'avoir ce cas) : $\langle 3|out\rangle$.

La collision peut être calculée comme le résultat d'un opérateur qui transforme l'état entrant en état sortant : $|out\rangle = S|in\rangle$. L'opérateur S est souvent appelé matrice S car on travaille fréquemment dans une représentation mathématique des opérateurs sous forme matricielle (des tableaux de nombres).

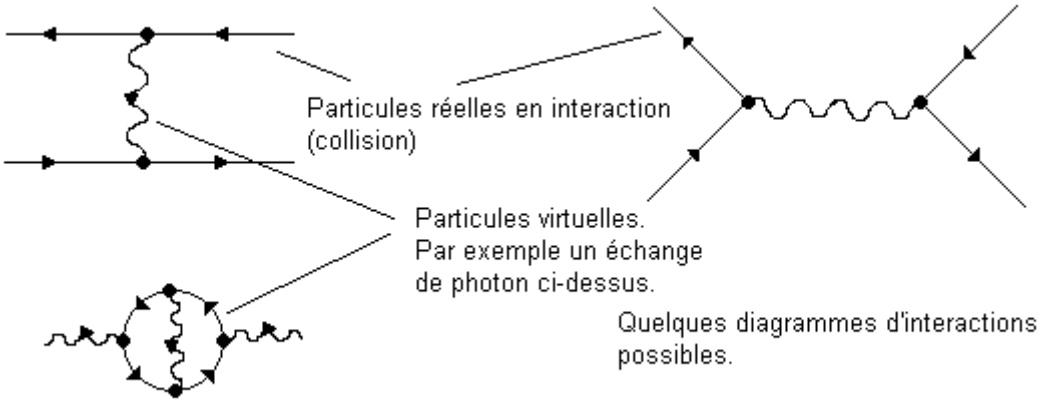
Le problème est alors de calculer cette matrice S qui contient tout ce que nous voulons savoir sur le processus de collision. Notons que s'il n'y avait pas de collision, c'est à dire si les particules s'évitaient ou s'il n'y avait pas d'interaction, on aurait alors $|out\rangle = |in\rangle$. L'opérateur qui ne change pas les états : $|\alpha\rangle = I|\alpha\rangle$ est appelé opérateur identité (par analogie avec la multiplication par le nombre 1 qui ne change rien). La théorie des perturbations permet alors de calculer une expression du type :

$$S = I + S_0 + S_1 + ...$$

Diagrammatique

Il existe une extraordinaire expression graphique de cette équation. Elle fut élaborée par Feynman. Cette méthode permet de tracer des diagrammes (appelés diagrammes de Feynman) et s'appelle donc méthode diagrammatique.

Un diagramme représente un processus de collision. On a deux particules qui s'amènent et qui repartent avec dans le processus des créations de particules.

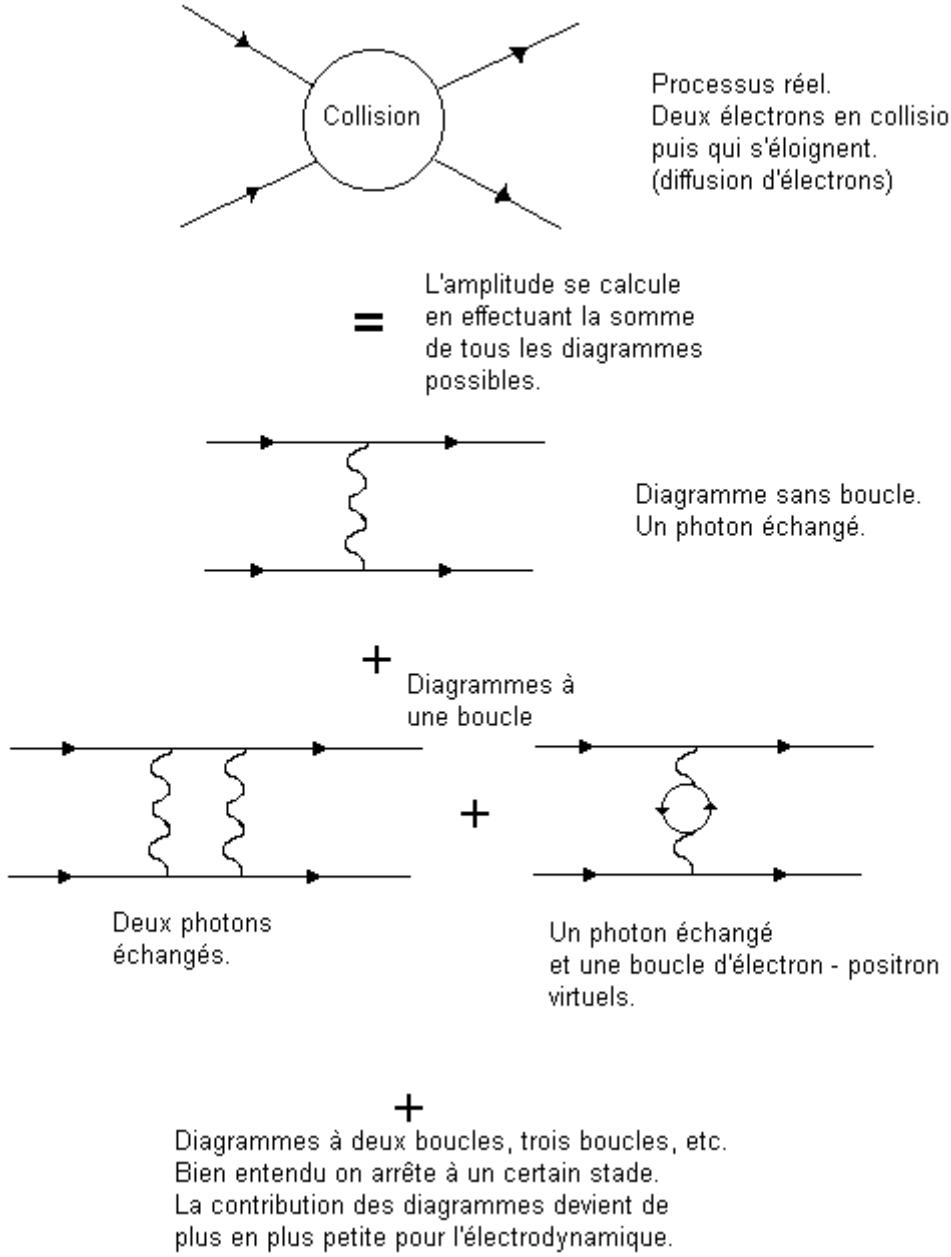


Ces diagrammes caractérisent des collisions mais peuvent aussi caractériser des propagations (une seule particule avant et après, comme le dernier ci-dessus) ou des désintégrations (une particule au départ, plusieurs après).

Il est possible d'établir une correspondance entre les diagrammes et les termes successifs calculés avec la théorie des perturbations.

Les règles de construction des diagrammes sont appelées règles de Feynman. Elles donnent les contraintes dans la construction des diagrammes (par exemple un électron ne peut pas être créé tout seul car la charge électrique est toujours conservée). Mais elles donnent aussi la manière d'associer le diagramme et les calculs (elles donnent des expressions mathématiques aux "branches" et aux "nœuds" du diagramme et la méthode pour les regrouper). Ces règles mathématiques sont généralement assez simples.

Dans l'expression pour la matrice S, les termes successifs correspondent aux nombres de boucles dans le diagramme.



Pour résoudre un problème de collision donné, on trace tous les diagrammes (jusqu'à une certaine taille, le nombre de diagrammes grandit très vite). Puis on les traduit en équations grâce aux règles de Feynman, on additionne le tout et le tour est joué, on a trouvé l'expression (approchée) de la matrice S. Si le résultat n'est pas suffisamment précis, il suffit de calculer plus de diagrammes.

3.3. La renormalisation

Le mécanisme de la renormalisation que nous allons présenter fut élaboré à la fin des années quarante et dans le courant des années cinquante. Certains résultats mathématiques rigoureux ne furent même obtenus que dans les années soixante. Il a encore subit des prolongements depuis et n'a certainement pas fini de nous réserver des surprises.

Les divergences

Le problème vient de l'apparition dans les calculs de nombres infinis. C'est plutôt gênant quand on calcule des probabilités. La probabilité maximale est évidemment de 100 %. Ce serait déjà ennuyeux de trouver plus, mais alors là, l'infini, c'est carrément catastrophique !

Le problème est même très sérieux. Quand on calcule les diagrammes pour l'électrodynamique, on constate que les seuls qui ne posent pas de problème sont ceux sans boucle ou avec une boucle. Dès que l'on a deux boucles, patatras, les infinis font leur apparition.

D'où vient le problème ? Précisons d'abord ce que nous entendons par "valeurs infinies". Les calculs consistent en la résolution d'intégrales sur les fréquences. C'est à dire que l'on a des fonctions mathématiques, et l'on effectue une opération mathématique (l'intégration) sur un paramètre qui est la fréquence. Le calcul devant s'effectuer en utilisant toutes les fréquences possibles de 0 à l'infini. Les petites fréquences sont celles des ondes radios et infrarouges, comme on l'a vu, et les hautes fréquences les ultraviolets.

On peut effectuer un calcul en limitant les fréquences. Par exemple on peut effectuer le calcul des intégrales de la valeur v1 à la valeur v2. Puis, dans la formule obtenue, on doit en principe faire tendre la valeur v1 vers zéro et la valeur v2 vers l'infini pour obtenir une valeur (par exemple l'amplitude ou la probabilité d'un processus). Et c'est là que les problèmes apparaissent. Dans la plus part des cas, le résultat devient infini. Plus la valeur de v1 approche de zéro et plus la valeur de v2 approche de l'infini, plus la valeur devient grande. On dit que l'intégrale est divergente. Si elle reste finie, elle est dite convergente.

Par abus de langage ont dit que le diagramme est divergent. Lorsque le résultat devient infini quand v1 tend vers zéro on dit que le diagramme est divergent infrarouge. De même lorsqu'il devient infini pour v2 tendant vers l'infini, on dit que le diagramme est divergent ultraviolet. On parle aussi de singularité infrarouge ou ultraviolette.

Les divergences infrarouges

Les divergences infrarouges sont les moins graves. Leur explication physique est simple. Les photons de petite fréquence ont aussi, nous l'avons vu, une petite énergie. Par conséquent pour une énergie donnée, on peut avoir un très grand nombre de photons infrarouges (on dit aussi des photons "mous"). A la limite, le nombre de photons peut tendre vers l'infini lorsque la fréquence tend vers zéro tout en conservant une énergie totale finie.

Mais vouloir tenir compte de tous les photons de fréquence aussi petite n'a pas de sens. Plus un photon a une énergie faible et plus il est difficile à observer. Des photons de très petite fréquence (des ondes radios dans les très grandes longueurs d'onde) échappent à la détection des appareils. On peut donc conserver une valeur minimale pour v1 sans changer les résultats observés. Et avec cette valeur minimale, sans la faire tendre vers zéro, les résultats deviennent fini. Il n'y a plus de divergence infrarouge.

Les divergences ultraviolettes

Une divergence ultraviolette est plus grave car les photons de grande fréquence ont aussi une grande énergie. Si leur nombre tend vers l'infini, l'énergie aussi ! Cette fois il ne s'agit plus d'une difficulté expérimentale mais bien d'un problème théorique, et grave de surcroît.

C'est d'autant plus ennuyeux que les résultats à une boucle donnent des résultats finis et corrects du point de vue expérimental (bien que moyennement précis puisque l'on a peu de diagrammes).

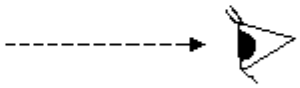
Pourquoi cela dérape-t-il juste après ? D'où vient cette anomalie ?

Dans la théorie des champs, nous avons expliqué que l'on part des équations décrivant le champ. Ce sont des équations classiques où les particules sont considérées comme classiques. Les particules entrent dans les équations via leurs caractéristiques : masse, charge électrique, etc. Ces particules sont appelées particules "nues" car elles peuvent être isolées. Un état classique à une seule particule représente une particule isolée.

Puis on passe à la quantification qui a pour conséquence de faire apparaître les fluctuations du vide. En particulier une particule ne peut plus être isolée.

Electron "nu" idéal.
Celui de la théorie classique et qui sert de base dans la théorie quantique.

Electron "nu" entouré de fluctuations.
Un nuage de particules virtuelles.



Un observateur extérieur, c'est à dire en réalité un appareil de mesure, ne "voit" que l'électron "habillé".
Lui seul a une réalité physique.

Autour de la particule il y a un nuage de particules virtuelles qui interagissent avec la particule "réelle". L'ensemble est appelé particule "habillée".

En réalité, seule la particule habillée a un sens physique réel, car on ne peut jamais enlever cet "habillage". Les seules particules que l'on peut observer, mesurer, sont les particules habillées.

Lorsque l'on mesure la masse ou la charge des particules, on mesure en réalité la masse ou la charge de la particule habillée. Celle-ci est forcément différente de l'hypothétique particule nue. Les paramètres de la particule nue sont appelés paramètres nus, par exemple la masse nue (on dit aussi masse "barre"). Ceux de la particule habillée sont appelés paramètres réels ou ne portent pas de qualificatif (masse réelle, masse habillée ou tout simplement masse tout court). Pour bien faire la distinction nous emploierons le terme de paramètres réels.

Supposons que l'on parte de masse nue, charge nue, … finis. On calcule alors les masses réelles, charges réelles,… en utilisant la théorie et en particulier les diagrammes. On obtient alors des résultats infinis comme nous l'avons vu. Mais dans la réalité c'est l'inverse qui se passe ! Les valeurs réelles sont finies et c'est les paramètres nus qui devraient être infinis. Pour eux ce n'est pas gênant puisque les particules nues n'existent pas physiquement. Les paramètres nus peuvent être considérés comme un intermédiaire mathématique pour décrire la théorie.

Voilà donc l'origine de ces infinis. On a considéré que les particules nues étaient les particules réelles et avaient des propriétés finies alors qu'en réalité ces particules nues n'existent pas et devraient (si jamais elles existaient) avoir des propriétés infinies.

Tout cela explique, intuitivement, que l'apparition des divergences ne se produise qu'à partir des diagrammes présentant une certaine complexité puisque ceux-ci représentent des particules "un peu plus habillées".

Maintenant que nous avons compris ce qui se passe, il nous faut trouver une méthode pour obtenir des résultats corrects.

Pour résoudre le problème nous allons procéder en trois étapes.

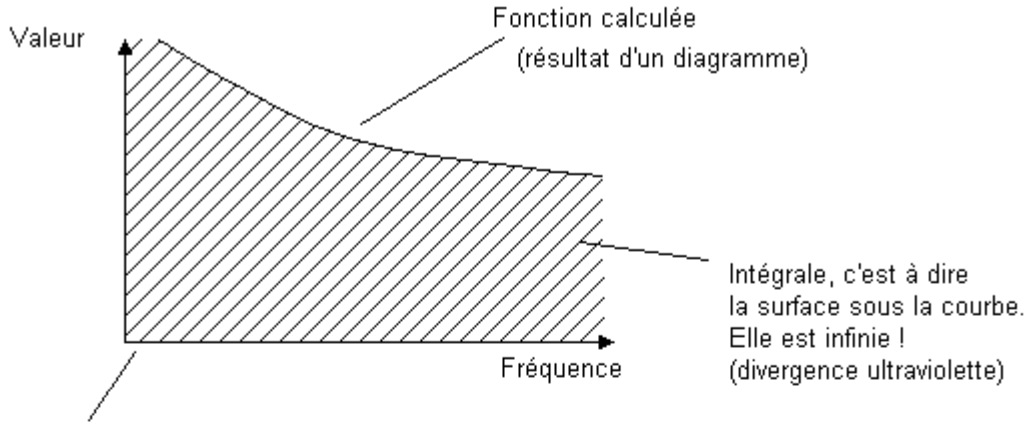
1ère étape : la régularisation

La première étape s'appelle la régularisation.

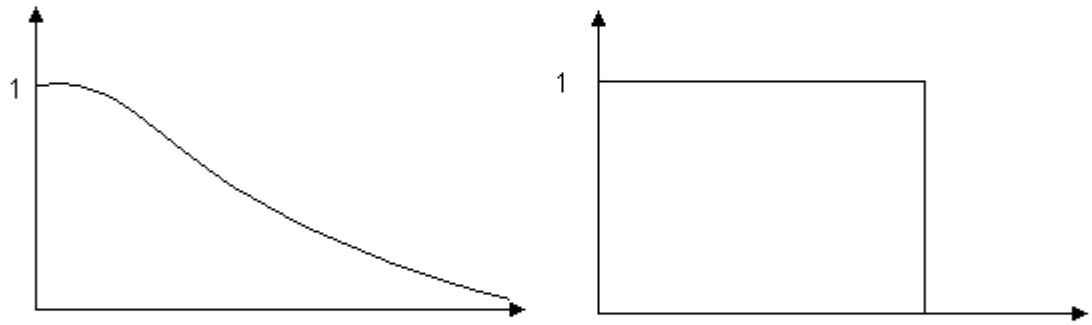
Cela n'a pas de sens de travailler avec des nombres infinis. Des intégrales qui divergent sont mathématiquement sans signification et on peut obtenir des résultats aberrants.

On va donc veiller à d'abord ne manipuler que des résultats finis. Il existe pour cela une énorme variété de méthodes. Aucune n'est universelle. Chacune présente des avantages et des inconvénients.

Une première méthode consiste à effectuer une "coupure". C'est à dire que l'on coupe les fréquences pour ne conserver que les plus petites. Techniquement, on peut faire comme expliqué au début, on garde la valeur de v2 finie. Il existe des méthodes moins "brutales" comme amortir les valeurs de grande fréquence.

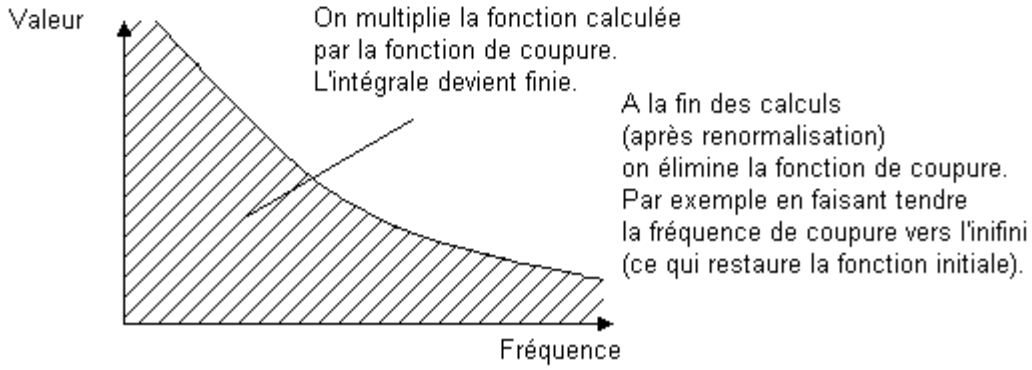


Si la surface diverge ici,
c'est une divergence infrarouge.
Il suffit d'ignorer cette partie
(qui est due à l'aspect irréaliste
de la mesure de photons
de fréquence aussi petite)



Fonctions de coupure possibles.
Il suffit qu'elles amortissent les hautes
fréquences (mais on leur cherche aussi
d'autres qualités, voir le texte)

Fréquence de coupure.



Une autre méthode consiste à faire varier le nombre de dimensions de l'espace-temps. C'est la régularisation dimensionnelle. Attention, il n'y a pas d'interprétation physique à cette opération. C'est une pure astuce mathématique. D'ailleurs on utilise même des valeurs fractionnelles pour le nombre de dimensions (4,5 dimensions par exemple) ! On peut montrer que les résultats deviennent

convergents, donc fini, pour des dimensions suffisamment élevées (pour l'électrodynamique). Dans un monde à cinq dimensions, les équations de l'électrodynamique quantique sont convergentes.

Il existe bien d'autres méthodes (régularisation de Pauli - Villars par exemple). Quels sont les critères recherchés dans ces méthodes ?

Le premier critère est bien entendu la simplicité des calculs. Les deux méthodes ci-dessus sont, par exemple, relativement simples. Mais, selon le problème physique considéré, l'une ou l'autre méthode peut s'avérer plus simple.

Une caractéristique cruciale est la consistance de la méthode. Il ne faut pas que le résultat final dépende de la régularisation qui n'est qu'une astuce intermédiaire pour manipuler des quantités mathématiquement significatives. Si la méthode que j'utilise dépend d'un paramètre, par exemple v2 ci-dessus, il ne faut pas que ce paramètre se retrouve dans le résultat final.

Il est souvent préférable que la relativité soit conservée. Certaines méthodes (comme la coupure de fréquence) brisent explicitement l'invariance relativiste. Cela peut parfois s'avérer gênant.

Enfin, il est évident que la méthode doit pouvoir s'appliquer ! Par exemple, la régularisation dimensionnelle n'est pas utilisable en présence de fermions (pour des raisons techniques que je ne développerai pas ici). Donc en présence d'électrons cette méthode est inutilisable.

2ème étape : la renormalisation

Maintenant que nous avons des résultats finis, convergents, nous pouvons passer au mécanisme appelé renormalisation.

Ce mécanisme consiste à réexprimer les résultats non plus en fonction des paramètres nus (qui sont en réalité infinis) mais en fonction des paramètres réels finis. Il existe plusieurs méthodes. L'une d'entre elle, la plus simple mais la moins rigoureuse, consiste à isoler les termes divergents dans les calculs et à les enlever. Cette méthode est peu rigoureuse car il y a plusieurs manières d'isoler les parties divergentes !

Une méthode plus rigoureuse consiste à modifier les masses et les charges en leur ajoutant des "contre termes". On dit donc que la masse nue, par exemple, est égale à la masse réelle plus un contre terme (potentiellement divergent puisque la masse nue est infinie).

$$m_0 = m + \delta m$$

Le suffixe 0 est mis pour indiquer qu'il s'agit de la masse nue.

On calcule la masse ou la charge, par exemple avec des diagrammes (c'est un tantinet plus compliqué que cela, mais c'est le principe) et on pose l'égalité du résultat avec la masse réelle ou la charge réelle.

Cela modifie les formules qui deviennent alors convergentes. La procédure consiste donc à réexprimer les formules en fonction des paramètres réels finis. Techniquement, les parties divergentes de la formule sont compensées par les parties divergentes provoquées par le contre terme :

Formule originale (divergente) = Formule convergente + des divergences ultraviolettes

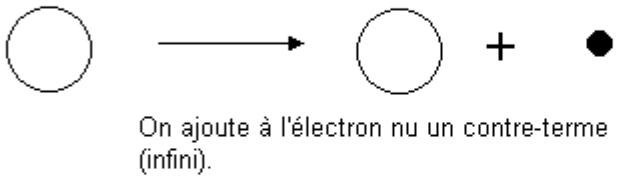
Après introduction des contre termes, on obtient :

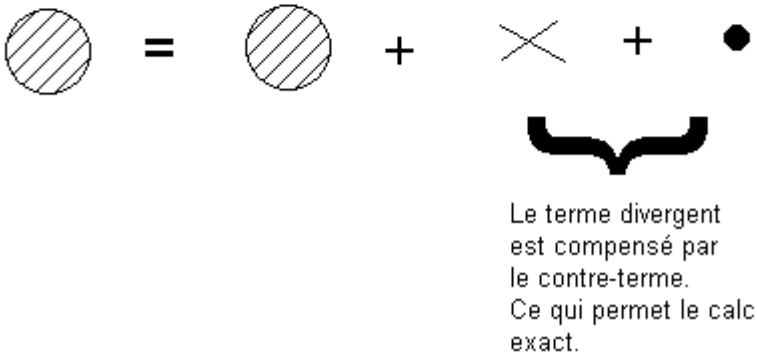
Formule originale (avec des contre termes) = Formule originale (sans contre termes) + des termes dépendant des contre termes.

Le résultat final est alors

Formule finale = Formule originale (avec des contre termes) = Formule convergente + des divergences ultraviolettes + des termes dépendant des contre termes.

Et les deux dernières parties se compensent (après identification des paramètres calculés et des paramètres réels) rendant la formule convergente.





Ce genre de manipulation avec des valeurs potentiellement infinie (masse nue, contre terme) peut faire peur. Mais n'oublions pas qu'avec la régularisation tous ces termes sont en réalité finis et ont une signification mathématique précise.

3ème étape : fin de la régularisation

La dernière étape est simple, elle consiste à enlever la régularisation. Comme les formules sont devenues convergentes, les résultats finaux sont finis. Par exemple, on fait tendre v2 vers l'infini dans la méthode de coupure, mais le résultat lui-même reste fini.

Toute cette procédure peut sembler artificielle et pourtant cela marche très bien, et les résultats sont corrects (expérimentalement) ! Cela cache forcément quelque chose de plus profond, nous verrons d'ailleurs plus tard (interactions fortes) que la renormalisation conduit à un mécanisme physique extrêmement fécond (le groupe de renormalisation).

Théorie des perturbations

La méthode précédente peut être appliquée avec la théorie des perturbations. On travaille ordre par ordre. C'est à dire que l'on travaille comme suit :

- on calcule tous les diagrammes à l'ordre de deux boucles.
- On effectue la procédure de renormalisation. Les paramètres (masses, charges, …) sont modifiés par des contre termes. Les résultats sont finis.
- On calcule tous les diagrammes à l'ordre de trois boucles.
- Les diagrammes sont, à nouveau, infini mais on peut appliquer la même procédure. Les paramètres sont à nouveau modifié (sans que les diagrammes précédents ne deviennent divergents). Les contre termes sont simplement plus "précis". Les résultats sont à nouveaux finis (et plus précis)

La méthode consiste formellement à calculer la valeur des contre termes par la théorie des perturbations : $\delta m = 0 + \delta m_1 + \delta m_2 + ...$

L'ensemble de cette théorie a bien sûr un énorme arsenal mathématique a sa disposition. En particulier il existe des théorèmes prouvant la convergence de toutes les formules après renormalisation des paramètres. Il existe des théorèmes prouvant l'indépendance des résultats selon les méthodes utilisées (en particulier un paramètre comme v2).

Les théoriciens ont mis au point des méthodes sophistiquées applicables en théorie des perturbations. Constructions dites "explicites", récursives, méthode paramétrique, etc.

Les théories renormalisables et non renormalisables

Il est très important pour la suite de parler des théories renormalisables et non renormalisables. Qu'est-ce que c'est ?

Nous nous sommes surtout concentrés sur l'électrodynamique. Avec cette théorie la renormalisation fonctionne très bien. Mais ce n'est pas toujours le cas. On peut construire toutes sortes de théories (qui n'ont pas toujours d'applications physiques mais qui sont utiles pour améliorer les outils mathématiques) et certaines ne peuvent pas être renormalisées.

Comment est-ce possible ?

Nous avons vu que pour l'électrodynamique il suffisait de renormaliser la masse et la charge de l'électron. Pour certaines théories c'est insuffisant. En pratique, on travaille comme suit.

On trace les diagrammes à un certain ordre (par exemple tous les diagrammes à deux boucles). On regarde les diagrammes possédant des divergences ultraviolettes. Pour chacun d'eux on introduit un paramètre à renormaliser. Dans l'électrodynamique il faut introduire deux paramètres (la masse et la charge), il y a d'autres diagrammes divergents mais des mécanismes de compensation en éliminent automatiquement (c'est un mécanisme mathématique ayant en réalité une origine physique, lié aux symétries, et appelé identités de Ward) mais cela importe peu.

Puis on trace les diagrammes à l'ordre suivant. A nouveau, on a des diagrammes divergents. On regarde si les paramètres initiaux suffisent sinon on doit rajouter ne nouveaux paramètres. Et on continue ainsi de suite.

Si le nombre total de paramètres à introduire (deux en électrodynamique) pour rendre tous les diagrammes convergents (à tout ordre possible) est fini, alors la théorie est dite renormalisable.

Si le nombre de paramètres est infini alors la théorie est dite non renormalisable. C'est à dire que chaque fois que l'on rajoute des diagrammes ont doit ajouter de nouveaux paramètres à renormaliser.

Ces paramètres ne sont pas prédits par la théorie. On doit les mesurer puisque dans la procédure de renormalisation on les compare avec les paramètres mesurés. On identifie par exemple, comme on l'a vu, la masse calculée et la masse réelle (mesurée). On les appelle des paramètres libres.

Cela veut dire que dans une théorie non renormalisable il y a une infinité de paramètres arbitraires ou qui doivent être déduits de l'expérience. La théorie perd toute signification car il est toujours possible de trouver des valeurs pour cette infinité de paramètres de telle manière que les résultats collent aux résultats expérimentaux.

Il n'est bien entendu pas nécessaire de suivre toute cette procédure pour savoir si la théorie est renormalisable ou pas. Il existe des formules simples qui permettent de le savoir immédiatement (en général !), à partir de l'expression des équations des champs.

3.4. Les symétries

En physique, les symétries jouent un rôle très important. Une opération de symétrie est une opération qui change certains paramètres du système (nous allons en voir tout de suite quelques exemples). Le système est dit invariant sous cette symétrie si les résultats sont inchangés ou plus exactement si on obtient les mêmes résultats ayant subit l'opération de symétrie.

Système -> résultats R

Système après symétrie -> résultats R'

R après symétrie = R' (invariance)

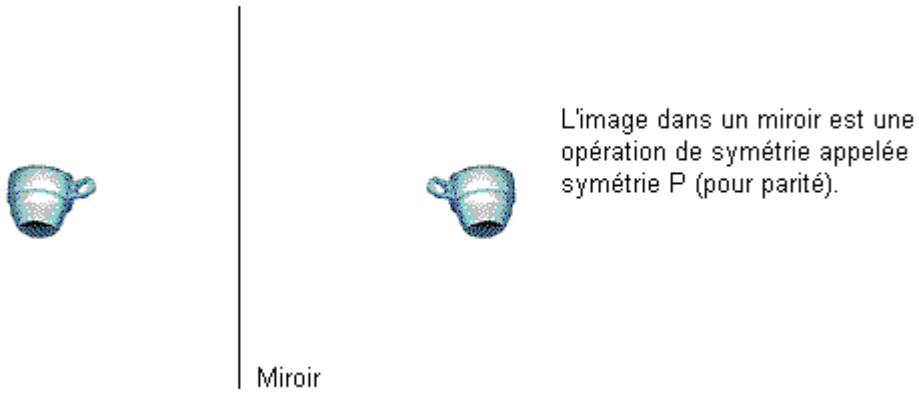
Tout cela est un peu formel, ce sera plus facile à comprendre sur des exemples.

Symétries discrètes

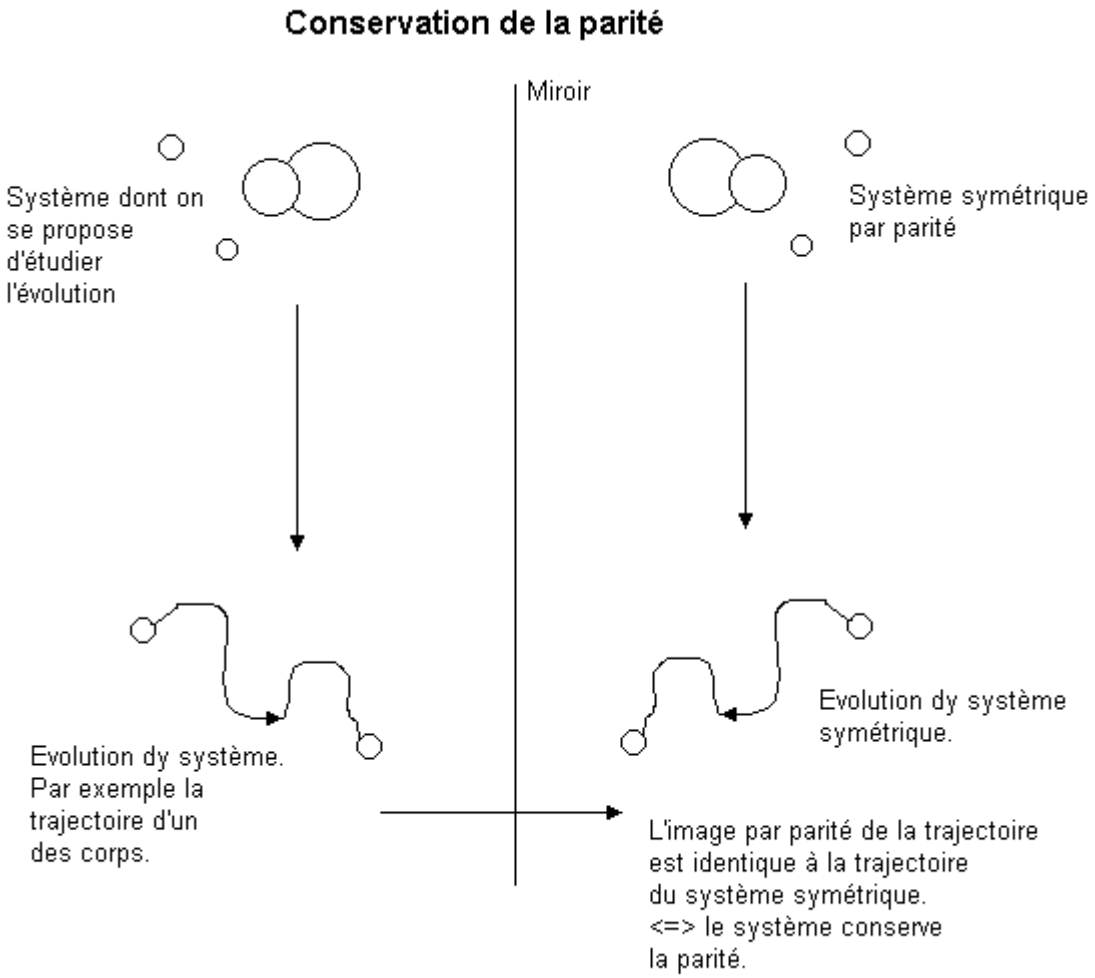
Il existe trois symétries discrètes que nous allons prendre la peine de présenter car cela en vaut la peine.

Parité

La parité est analogue à la réflexion dans un miroir. On renverse le sens de toutes les directions.

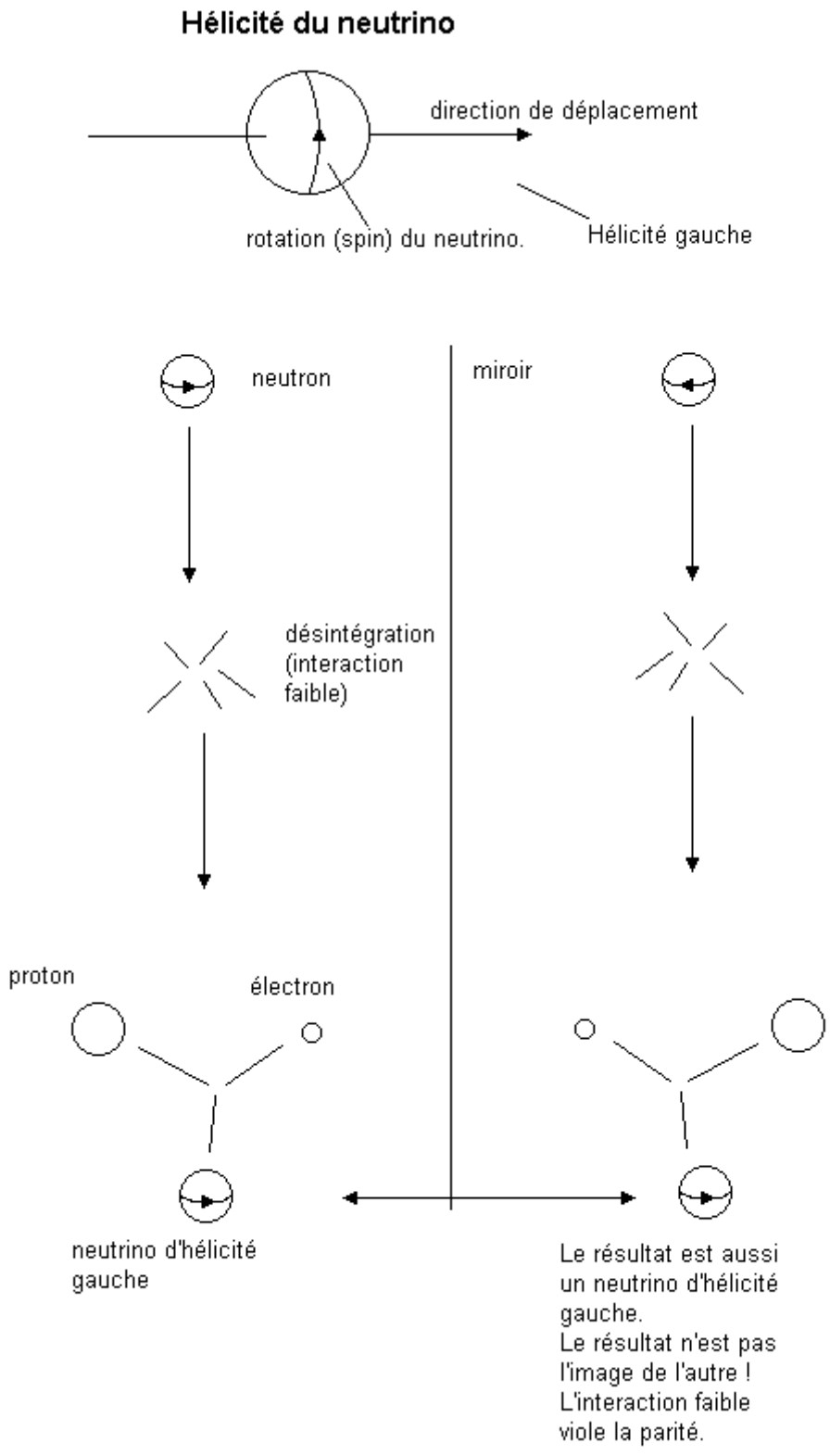


Le système sera invariant sous la parité si les lois physiques sont identiques dans le miroir. Par exemple, l'électrodynamique est invariante sous la parité.



Prenons un système donné avec un électron. Les équations vont prévoir une trajectoire pour cet électron. Prenons le même système vu dans un miroir. On peut lui appliquer les mêmes équations. La trajectoire prévue est alors bien l'image de la première, le système est invariant sous la parité.

Il peut sembler évident qu'il doive en être ainsi. Mais c'est faux ! Il existe des phénomènes physiques qui ne sont pas invariant par parité. Par exemple, il existe une particule appelée neutrino qui a la curieuse propriété de tourner toujours dans le même sens. Si on la regarde par l'arrière, on la voit tourner dans le sens inverse des aiguilles d'une montre (ont dit que son hélicité est gauche).

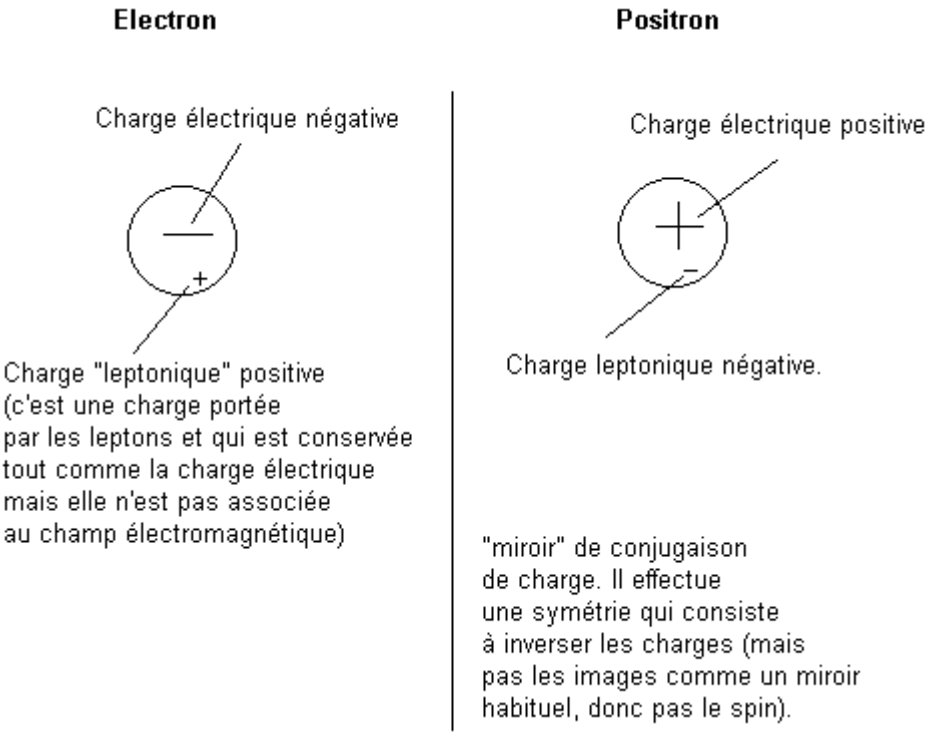


Après inversion par parité, on s'attendrait à ce que la rotation d'un neutrino soit inversée (hélicité droite). Mais le monde n'est pas ainsi ! Il n'existe que des neutrinos d'hélicité gauche. Donc, même après parité, les neutrinos prévus par la théorie sont encore des neutrinos d'hélicité gauche. La théorie (et la nature) viole la parité. C'est une conséquence de l'interaction faible qui viole la parité (nous parlerons plus tard de cette force).

Comme pour toute opération, en mécanique quantique on peut la caractériser par un opérateur mathématique agissant sur les états physiques. Pour l'opérateur on le nomme P et il agit sur un état en le transformant en le même état inversé par parité.

Conjugaison de charge

La conjugaison de charge consiste à inverser toutes les charges des particules. Pour l'électrodynamique cela revient à remplacer les charges négatives par des positives.



A nouveau, l'électrodynamique est invariante sous cette opération. La théorie prévoit le même comportement pour les électrons négatifs que pour des électrons positifs. En fait, les électrons positifs sont les positrons, l'anti-particule de l'électron. Nous en reparlerons un peu plus loin.

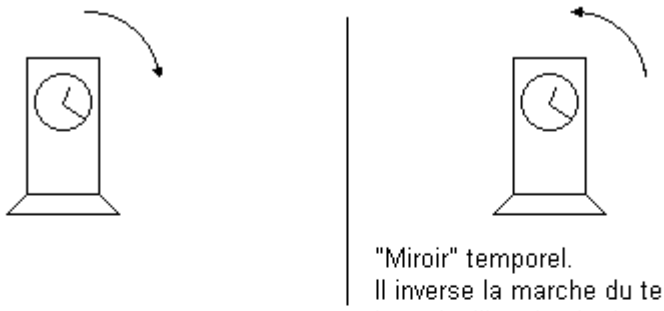
L'opérateur pour la conjugaison de charge s'appelle C et il change, par exemple, les électrons en positrons.

Le neutrino possède une anti-particule appelée, bien évidemment, anti-neutrino. Mais les lois de l'interaction faible ne respectent pas non plus cette symétrie ! Par contre, elles sont invariantes sous la combinaison des deux, c'est à dire sous l'opérateur . Si on effectue une opération de conjugaison de charge suivie d'une parité, le système est invariant aussi pour les neutrinos et l'interaction faible.

On a longtemps cru que cette double opération état toujours invariante. C'est à dire qu'elle était une symétrie conservée pour toutes les lois de l'univers. Mais on a découvert des particules, les mésons K, qui violent cette symétrie . C'est une propriété de l'interaction forte dont nous parlerons également plus tard.

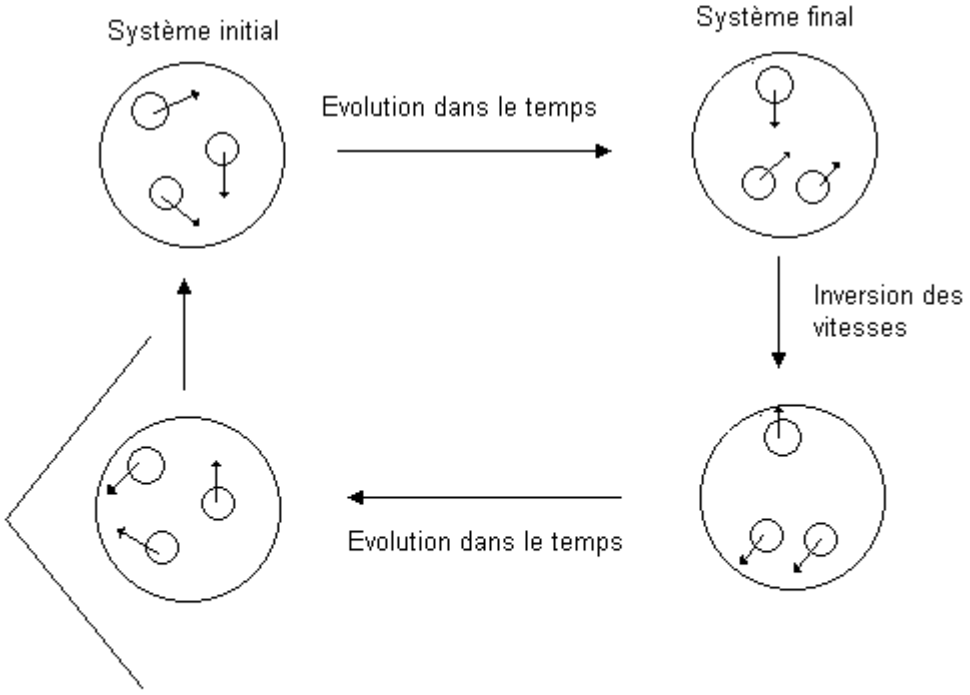
Renversement du temps

La dernière symétrie consiste à renverser le temps.



"Miroir" temporel.
Il inverse la marche du temps.
Les aiguilles des horloges tournent en sens inverse.

Puisqu'après la symétrie tout tourne "à l'envers", pour un système physique, prendre l'opération de symétrie T revient à prendre l'état final et à inverser les vitesses.



Si, après inversion des vitesses, on retrouve l'état initial, alors le système est invariant par renversement du temps.

Au début on a un système dans l'état $|\alpha\rangle$, il évolue au cours du temps et se transforme en l'état $|\beta\rangle$.

Prenons maintenant l'état $|\beta\rangle$. Renversons le sens du temps. Appliquons les équations et voyons si l'on retombe bien sur l'état $|\alpha\rangle$. Si oui, le système est invariant par renversement du temps.

L'électrodynamique est invariante par renversement du temps et l'opérateur est nommé T.

Théorème PCT

Les théoriciens ont réussi à démontrer mathématiquement (ce qui est une très belle réussite) que dans un monde quantique relativiste, la combinaison des trois symétries était toujours respectée. C'est à dire que le monde est invariant sous la combinaison des trois opérateurs PCT.

Cela implique que si une des trois symétries est non respectée, alors forcément la combinaison des deux autres doit également être violée pour pouvoir rétablir l'invariance de la combinaison des trois symétries.

On a vu que l'interaction forte violait la symétrie . Cela signifie qu'elle doit obligatoirement violer la symétrie T. Le comportement des mésons K n'est pas invariant par renversement du temps !

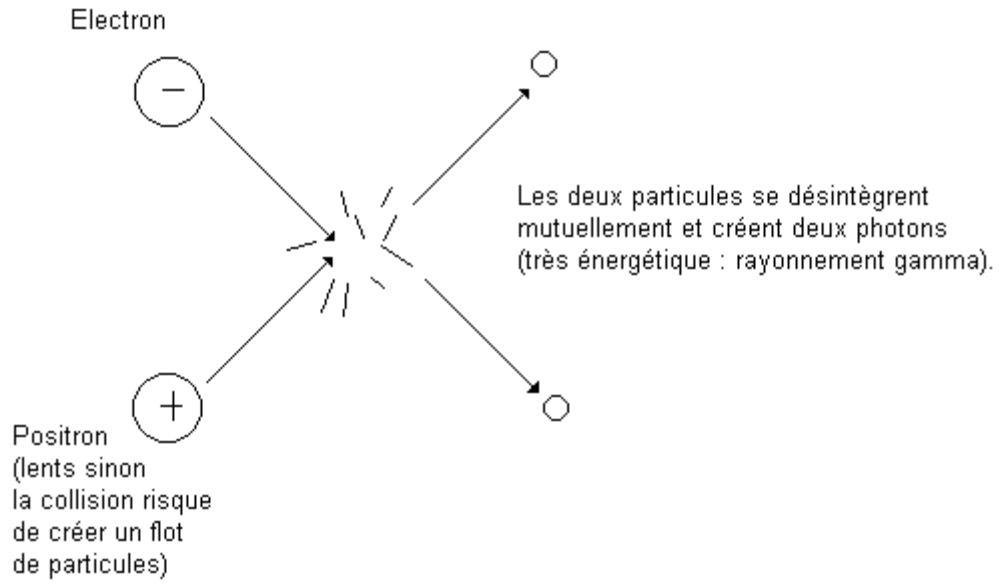
Faut-il voir là l'origine de la flèche du temps ? C'est à dire du fait que le temps semble s'écouler dans un sens privilégié ? Non, probablement pas. Du moins pas directement. Cette asymétrie dans le sens du temps est un effet statistique (décrit par la thermodynamique qui dit que le sens du temps se fait toujours de l'ordre vers le désordre) alors que cette violation de la symétrie T par les interactions fortes est mineure. Toutefois, si l'on recherche l'origine de cette asymétrie dans l'univers, on finit par réfléchir à l'univers tout entier et à sa création ! On en vient à se demander pourquoi à l'origine l'univers était très ordonné et à la fin de l'univers (il sera) très désordonné. L'origine de cette asymétrie dans la nature globale de l'univers pourrait avoir une origine fondamentale qui n'est pas nécessairement étrangère avec celle constatée dans l'interaction forte.

Anti-Matière

Le théorème PCT a des conséquences extraordinaires. Il implique notamment l'existence de l'anti-matière. C'est à dire que pour toute particule il doit exister une anti-particule avec la même masse mais avec toutes ses charges opposées (jusqu'ici nous n'avons vu que la charge électrique). Ainsi nous avons déjà parlé du positron, de l'anti-neutrino, mais il y a aussi l'anti-proton, l'anti-neutron,… Toutes particules qui ont réellement été observées.

Parfois, pour certaines particules qui ne portent aucune charge (ni électrique, ni autre), la particule est sa propre anti-particule. C'est le cas du photon. L'anti-photon est identique au photon, c'est la même particule.

L'univers est essentiellement composé de matière. Si une partie de l'univers était fait d'anti-matière cela ne passerait pas inaperçu. En effet, lorsque de la matière rencontre de l'anti-matière, il y a annihilation totale. Les deux se désintègrent totalement pour former des photons.



La quantité d'énergie dégagée est formidable. Rappelez-vous la relativité restreinte et le rapport entre masse et énergie.

A la zone de contact matière / anti-matière dans l'univers, on aurait un formidable dégagement d'énergie qui serait clairement visible.

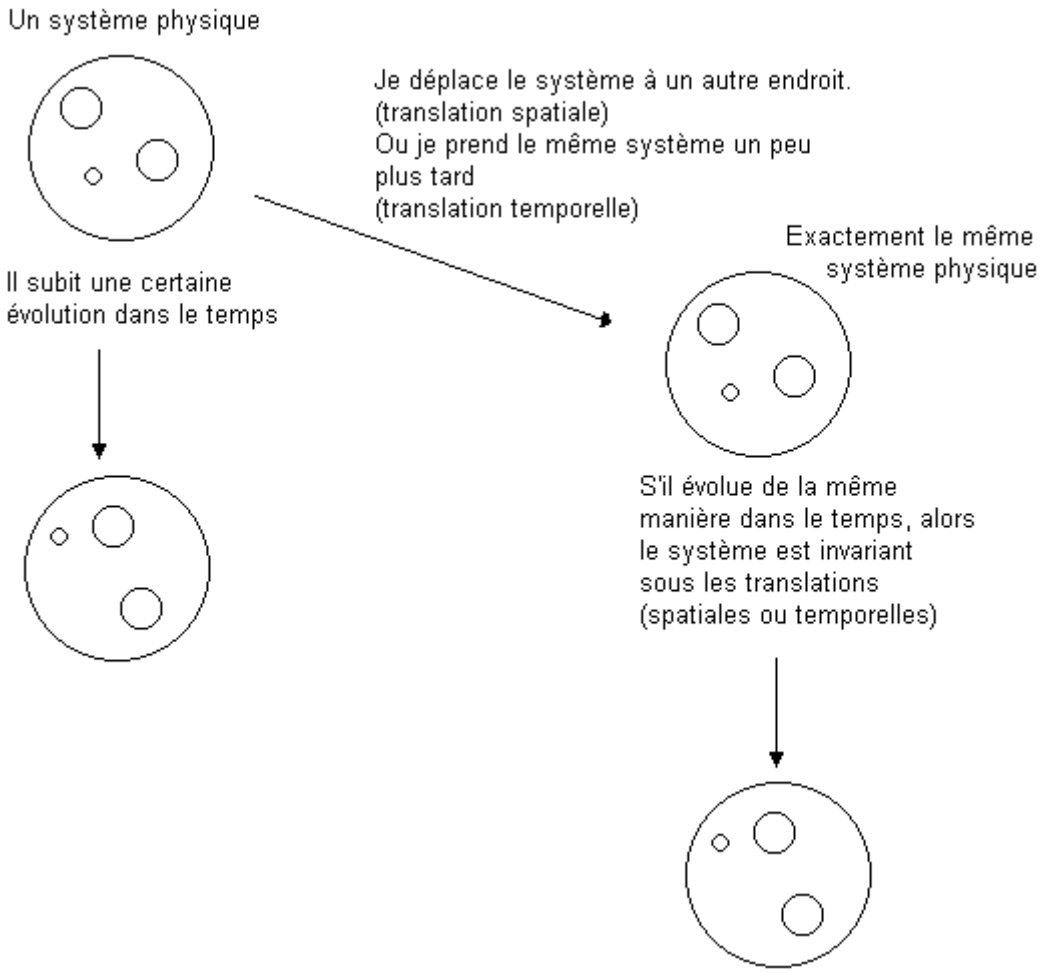
De plus, la terre est bombardée de rayons cosmiques. Ce sont des jets de particules qui viennent du soleil, mais aussi d'autres étoiles et même des confins de l'univers. Or les rayons cosmiques sont composés en quasi-totalité de matière.

Le fait, que la symétrie CP ne soit pas respectée, a des conséquences extraordinaires sur la matière et l'anti-matière. Cela implique que dans les réactions, l'un des deux peut être légèrement favorisé. Cela pourrait expliquer pourquoi l'univers est composé uniquement de matière. Je ne vais pas donner un cours sur le Big Bang, mais disons que celui-ci prévoit qu'à une époque reculée l'univers était plus petit, plus dense et plus chaud. Dans cette soupe de particules très énergétiques les réactions de collisions étaient nombreuses et la matière et l'anti-matière étaient en équilibre (à cause des collisions il y avait autant de création de particules que de désintégrations). Pour finir, toute la matière et l'anti-matière composant cette soupe s'est refroidie et la matière s'est désintégrée avec l'anti-matière. Il est juste resté un peu de matière, qui formera un peu plus tard les étoiles et les galaxies, le reste formant un bain d'ondes électromagnétiques, de photons. Les physiciens ont calculé qu'il suffisait d'un petit excès d'un milliardième de matière pour expliquer l'aspect et la composition de l'univers actuel. C'est très peu et il pourrait s'expliquer par la violation de la symétrie CP.

Symétries continues

Les symétries discrètes portent ce nom parce qu'il s'agit de symétries "tout ou rien". Mais il y a d'autres possibilités. On peut avoir des symétries continues, c'est à dire qui dépendent continûment d'un ou plusieurs paramètres.

Prenons un exemple simple : les translations spatiales. Supposons un système donné avec des particules. Les particules vont alors évoluer en suivant certaines trajectoires. Prenons maintenant le même système mais déplacé. Laissons le, à nouveau, évoluer et regardons les trajectoires. Si ce sont les mêmes trajectoires mais déplacées de la même manière, alors le système est invariant par translation.



La translation spatiale dépend de trois paramètres. Je peux déplacer mon système dans le sens gauche - droite d'une certaine quantité, mais aussi dans le sens avant - arrière et dans le sens haut - bas.

Il existe d'autres opérations de ce type. Par exemple les rotations et les translations temporelles. Pour une translation temporelle on reprend simplement le même système dans le même état mais à un moment différent.

Les équations de l'électrodynamique sont invariantes sous toutes ces opérations de symétrie.

Il est possible de montrer qu'à chaque symétrie continue invariante correspond une valeur conservée. Intuitivement cela semble logique. S'il y a quelque chose qui fait que l'évolution du système est invariante sous certaines transformations, alors il semble logique de pouvoir trouver une quantité qui est également invariante sous l'évolution.

Ainsi, à l'invariance sous les translations temporelles correspond la conservation totale de l'énergie. Si l'on prend le système (dans son ensemble), alors son énergie totale est conservée au cours du temps.

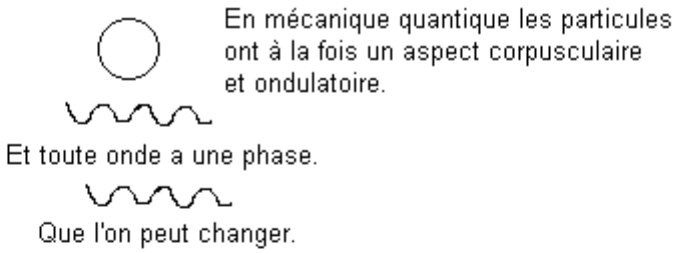
De même, à l'invariance sous les translations spatiales correspond la conservation de l'impulsion totale (pour une particule, l'impulsion correspond à sa masse multipliée par sa vitesse).

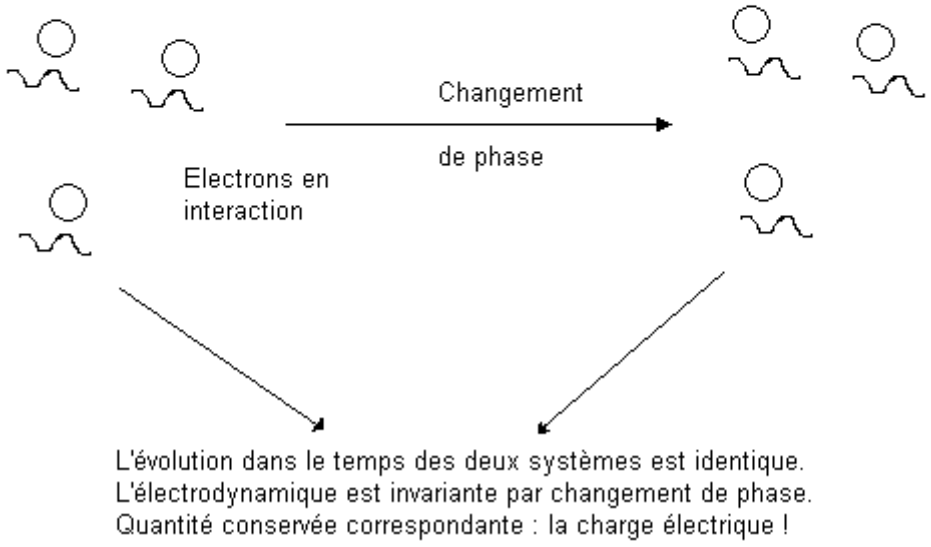
Symétries internes

Il existe d'autres types de symétries continues. On les appelle symétries internes par opposition aux précédentes qui sont des symétries de type géométrique.

Ces symétries ne sont pas faciles à visualiser justement à cause de cette différence. Lorsque l'on regarde les équations, on constate qu'il peut exister des opérations, analogue aux translations géométriques, mais sur des paramètres internes (au lieu des variables positions et temps). C'est des symétries tout à fait analogue. Mais au lieux de modifier des variables tel que la position ou le temps, elles modifient d'autres variables dans les équations.

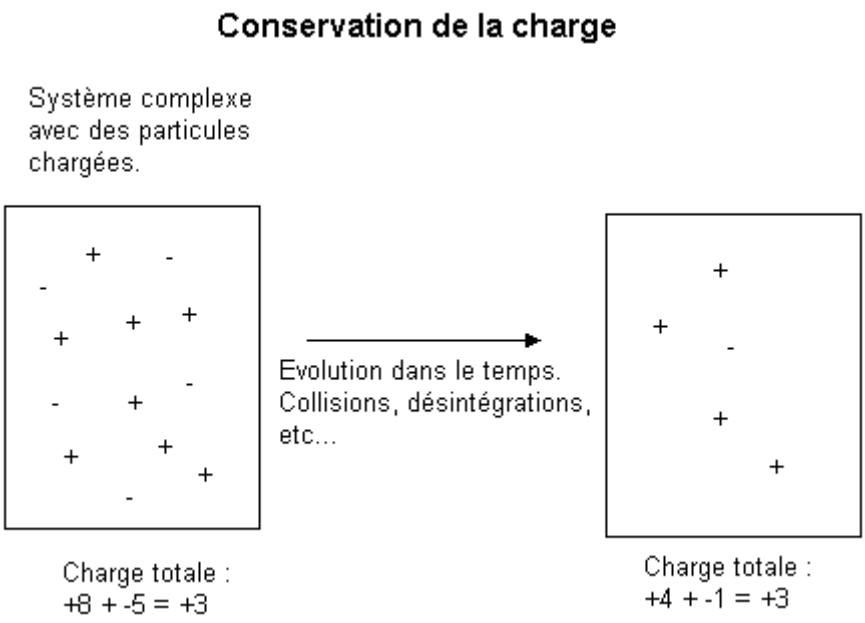
Prenons par exemple le champ électronique. On constate que les équations sont inchangées lorsque l'on modifie l'état des électrons en changeant leur "phase". Nous avons vu que les ondes (et donc les champs) pouvaient être caractérisées par une fréquence, une longueur d'onde et une phase.



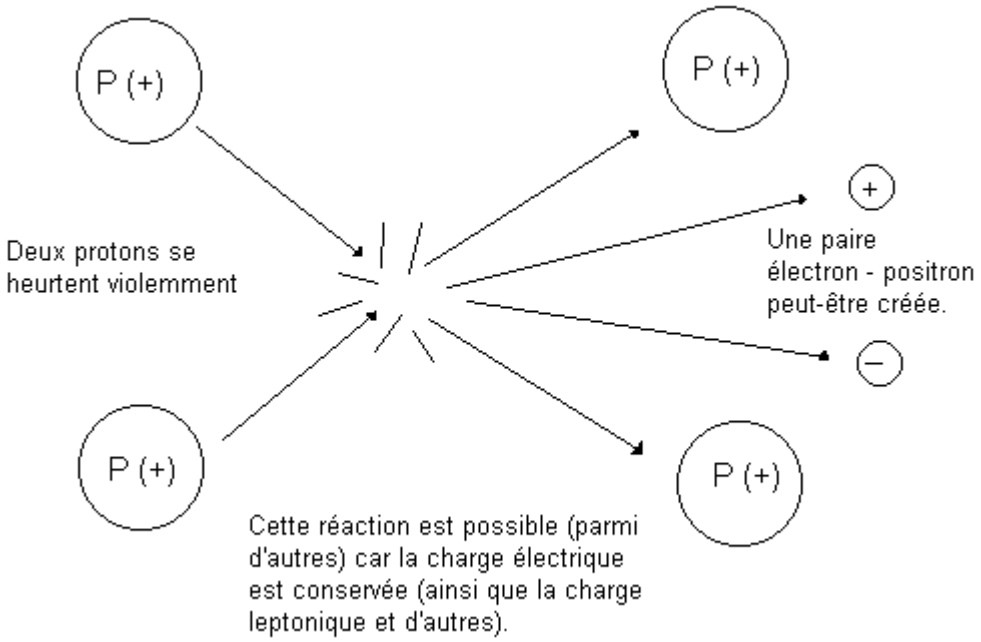


Tout comme les autres symétries continues, les invariances aux symétries internes conduisent à des quantités conservées. Cela a été généralisé grâce au théorème de Noether. Ces quantités conservées sont appelées dans ce cas des "charges". On montre aussi qu'à toute charge, on peut associer un "courant" conservé.

Dans le cas de l'électrodynamique, la symétrie au changement de phase correspond à la conservation de la charge électrique et du courant électrique.

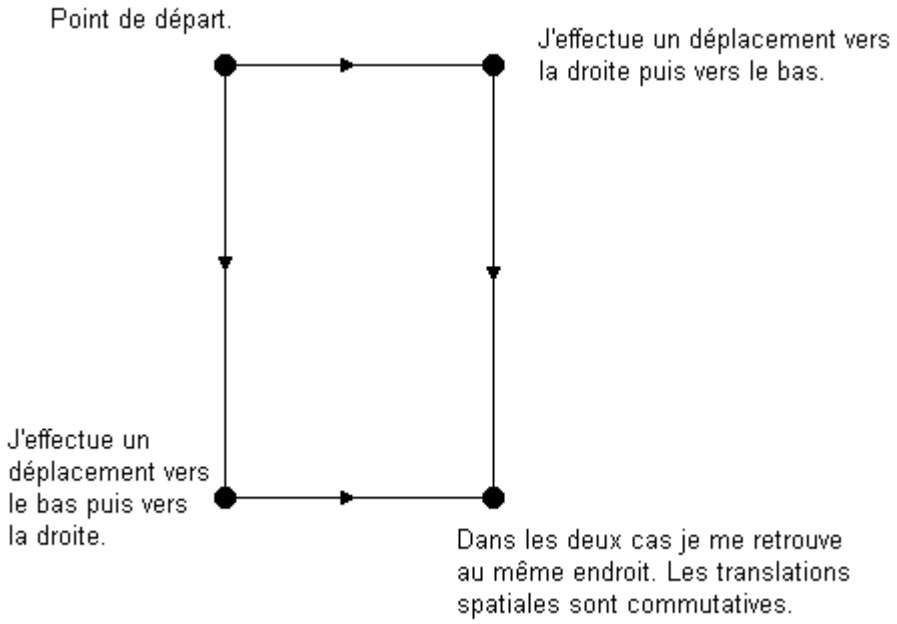


Rappelez-vous que dans les diagrammes j'avais dit que les règles de Feynman imposaient un certain nombre de contraintes. Effectivement, les diagrammes doivent respecter la conservation des charges. Un électron, porteur d'une charge négative, ne peut pas apparaître comme ça, tout seul, du néant. Par contre une paire particule - anti-particule peut toujours être créée puisque leurs charges sont opposées et se compensent. Seul l'énergie peut-être violée par les particules virtuelles, dans les diagrammes, puisque nous savons que cette énergie est imprécise pour le temps très court d'existence de ces particules.



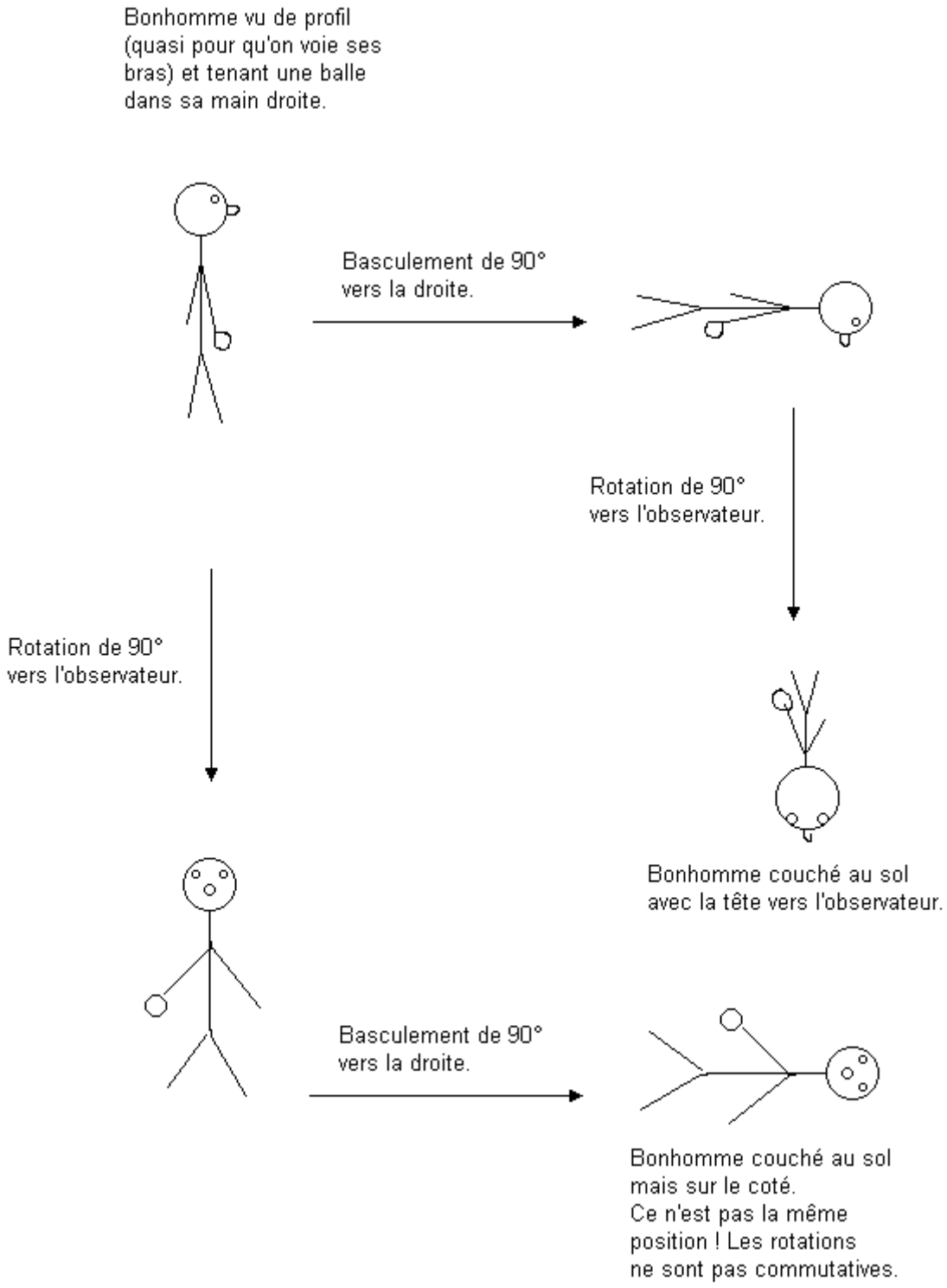
Symétries abéliennes et non abéliennes

Revenons aux translations spatiales. Supposons que j'effectue un déplacement vers la gauche suivi d'un déplacement en avant. Il est évident que si je fais d'abord un déplacement vers l'avant puis vers la gauche, cela revient au même.



On dit que cette symétrie est commutative ou abélienne.

Par contre pour les rotations ce n'est pas vrai. Si j'effectue une rotation dans un sens suivie d'une rotation dans l'autre sens, je n'obtiens pas la même chose en inversant les deux opérations.



On dit que les symétries par rotation sont non abéliennes.

Ce type de symétrie est souvent plus compliqué à étudier.

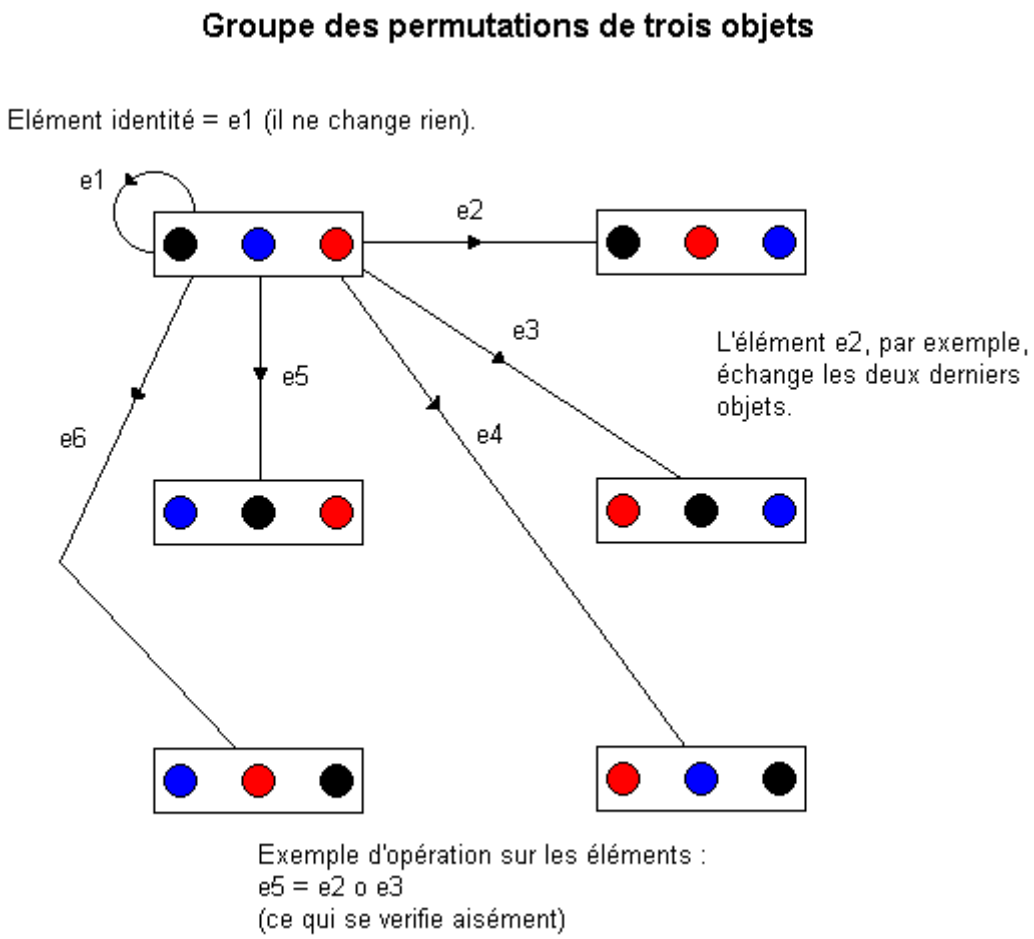
La symétrie par changement de phase, dont nous avons parlé, est une symétrie abélienne. Si j'effectue deux changements de phase, j'obtiens le même résultat en inversant les deux changements.

La théorie des groupes

La théorie des groupes est une branche des mathématiques. Je ne vais évidemment pas donner un cours sur les groupes ici (c'est d'ailleurs une branche assez vaste des mathématiques) ! Mais le mot reviendra souvent et il faut savoir de quoi on parle.

Un groupe est un ensemble d'opérations sur des objets. A chaque opération on fait correspondre chaque objet du groupe à un autre objet. Par la suite j'appellerai éléments du groupe ces opérations, c'est le terme communément utilisé.

Voici par exemple le groupe des permutations de trois objets.



Voici quelques propriétés des groupes :

- il existe un élément unité. Il doit y avoir un élément qui fait correspondre à chaque objet le même objet, donc ne change rien.
- Si j'applique deux éléments successivement (loi de composition du groupe), alors le résultat est un autre élément du groupe. Cette propriété et la précédente caractérisent les groupes.
- Les groupes avec un nombre fini d'éléments sont des groupes discrets. Les groupes avec un nombre infini d'éléments sont des groupes continus. On peut leur associer un certain nombre de "générateurs" qui permettent de parcourir tous les éléments du groupe.

- Il existe des groupes abéliens et non abéliens. Dans un groupe non abélien, l'ordre dans lequel on applique deux éléments est important. Le groupe des permutations ci-dessus est abélien, ce qui se vérifie aisément.

La théorie des groupes est la théorie parfaite pour étudier les symétries. En effet les symétries, comme nous les avons définis, consistent à effectuer une opération qui fait correspondre à chaque objet (par exemple une position) un autre objet (une autre position). A chaque opération de symétrie correspond un élément du groupe.

Les mathématiciens ont donné des noms aux groupes, ils les ont classés.

Ainsi les symétries discrètes obéissent au groupe Z_2. C'est le plus simple de tous les groupes. Il ne contient que deux éléments et s'applique à deux objets. Le premier élément laisse les objets inchangés (c'est l'identité) et l'autre consiste à échanger les deux objets. Il est identique au groupe des permutations de deux objets.

Les translations spatiales (dans une direction), les translations temporelles et les changements de phase correspondent au groupe continu $U(1)$ appelé groupe unitaire à une dimension. Un élément du groupe consiste à faire "glisser" les objets. Si je peux numéroter mes objets, un élément x devient l'élément $x+\delta$. δ étant une valeur caractérisant l'élément. Lorsque δ vaut zéro, c'est l'élément identité. Il est appelé générateur du groupe.

Enfin, les rotations obéissent au groupe (appelé justement groupe des rotations) continu, non abélien $O(3)$. Le trois rappelle qu'il y a trois générateurs (car il y a trois directions pour faire les rotations).

Il est possible de faire des combinaisons de groupe. Ainsi, le groupe correspondant à des translations spatiales dans les trois directions de l'espace correspond au groupe $U(1)\times U(1)\times U(1)$. On dit aussi que le groupe $U(1)$ est un sous-groupe du précédent.

En effet, si dans un groupe (par exemple celui des translations spatiales) je peux prendre une partie des éléments (par exemple les translations dans une seule direction) et qu'ils forment encore un groupe, alors ont dit qu'il contient un sous-groupe.

Voici quelques autres groupes et leur nom :

$O(4)$: groupe des rotations à quatre dimensions.

$U(2)$: groupe unitaire à deux dimensions.

$SU(3)$: groupe spécial unitaire à trois dimensions

$G(2)$: groupe exceptionnel deux.

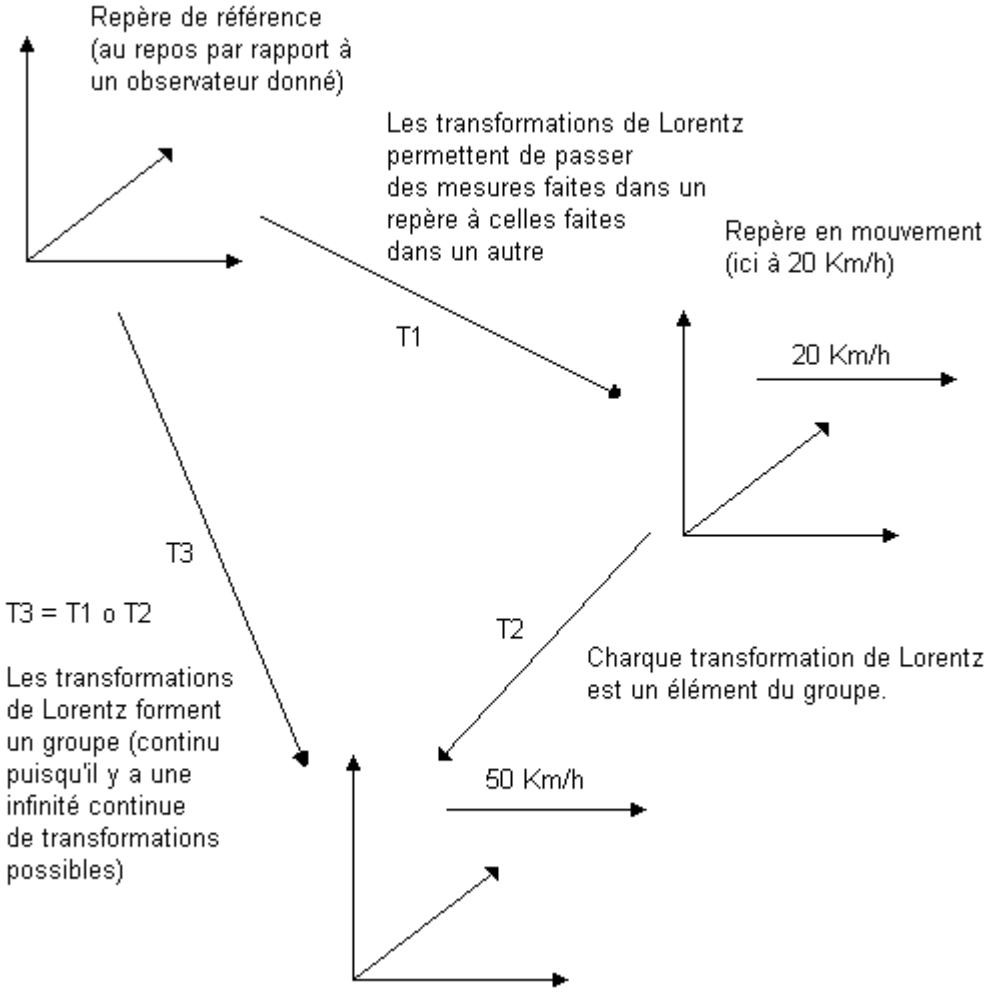
$SU(2)$: groupe spécial unitaire à trois dimensions. Il est mathématiquement équivalent (on dit isomorphe) au groupe des rotations $O(3)$.

Le groupe $SU(2)$ est un sous-groupe de $SU(3)$.

Un groupe très important dont j'aurai l'occasion de reparler est le groupe de Poincaré.

Nous avons vu qu'avec les translations spatiales, par exemple, on pouvait effectuer une transformation géométrique (un décalage) sur tous les points.

Les transformations de Lorentz, d'un autre coté, permettent d'effectuer aussi une transformation sur les points (sur les coordonnées, comme nous l'avons vu). Il se fait que ces transformations forment un groupe. En effet, il y a un élément unité (deux repères immobiles l'un par rapport à l'autre) et deux transformations successives forment encore une transformation de Lorentz.



Le groupe de Poincaré est constitué du groupe de Lorentz, des rotations, des translations et de la parité.

Il est amusant de constater que le groupe de Lorentz est isomorphe au groupe des translations à quatre dimensions. Donc mathématiquement équivalent. Mais il est physiquement différent (les objets auquel s'applique le groupe sont différents) car pour passer des transformations de Lorentz aux simples translations, il faut transformer l'espace-temps et utiliser un temps imaginaire ! Les nombres imaginaires sont des nombres définis par les mathématiciens et qui ont un lien avec les nombres complexes dont nous avons déjà parlé. Bien sûr, un tel temps imaginaire n'est qu'une opération mathématique, il n'a pas de signification physique réelle.

Les équations de la relativité (restreinte ou générale) sont invariantes sous des transformations par le groupe de Poincaré. La quantité conservée qui en résulte n'est rien d'autre que l'énergie et l'impulsion mais définies au sens relativiste, on l'appelle le tenseur énergie - impulsion.

3.5. Les théories de jauge

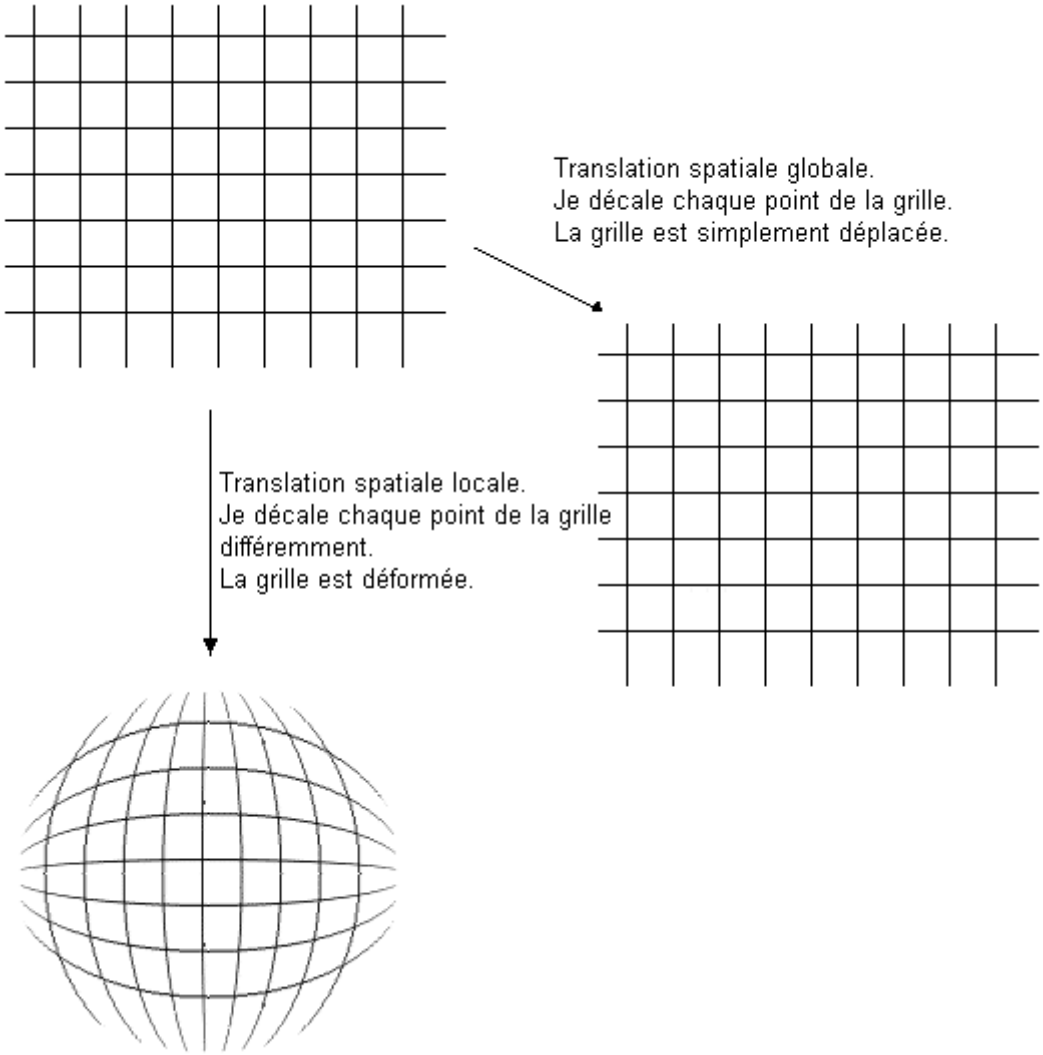
Les théories de jauge furent proposées par Yang et Mils en 1954. Elles ont été essentiellement élaborées dans les années cinquante et soixante. Mais elles n'ont jamais cessé de connaître des développements très importants depuis.

Principe des théories des champs de jauge

Les groupes de symétries internes dont nous avons parlés s'appliquent globalement. C'est à dire que la transformation appliquée est la même en tout point de l'espace.

Lorsque l'on fait varier la transformation de point en point, la symétrie est dite locale. Par exemple au lieu d'appliquer un changement de phase unique, en chaque point de l'espace on applique un changement de phase différent.

On peut facilement visualiser ce type de transformation avec les translations spatiales.



Bien entendu, on travaille ici avec des symétries internes, pas avec symétries géométriques. Mais le principe est le même.

La relativité a pour conséquence que le monde est local. C'est à dire que les interactions à distance peuvent être représentées par des champs où les équations décrivent uniquement des relations locales (pour des points voisins).

Ceci est du à l'existence d'une vitesse limite pour les signaux. Comme ceux-ci ne peuvent se propager instantanément, les forces ne peuvent se propager que de proche en proche. Comme on l'a vu avec les vecteurs de forces et leur portée.

Alors, il semble naturel de n'utiliser que des descriptions locales. Ainsi pour les symétries, on aimerait bien avoir une symétrie locale plutôt que globale comme précédemment.

Malheureusement, on constate rapidement que les équations des champs ne sont pas invariantes sous une symétrie locale.

Pour retrouver l'invariance, on ajoute alors de nouveaux champs appelés champs de jauge (pour des raisons historiques liées au champ électromagnétique qui possède une propriété d'invariance appelée invariance de jauge). On munit ces champs des propriétés de transformation adéquate sous le groupe de symétrie, de manière à rendre les équations invariantes sous le groupe local.

Le cas des électrons

Prenons par exemple le champ électronique.

On a vu qu'il est invariant sous un changement (global) de la phase des électrons. Cette transformation obéit au groupe $U(1)$.

Les équations qui décrivent le champ électronique ne sont pas invariantes sous une transformation locale de cette symétrie.

On ajoute alors un champ de jauge doté des propriétés de transformation requise de manière à ce que les équations totales soient invariante sous des transformations locales du groupe de symétrie.

Que constate-t-on ? Que les équations obtenues ne sont rien d'autres que les équations de l'électrodynamique. Le champ de jauge qui vient d'être ajouté n'est rien d'autre que le champ électromagnétique !

Cela conduit à une profonde signification physique. Le champ électromagnétique, et donc les photons ainsi que la force électromagnétique, est simplement la conséquence de l'invariance locale de la symétrie existant dans le champ électronique !

Les champs de jauges non abéliens

Si les champs obéissent à une symétrie interne non abélienne, alors les champs de jauge ajoutés sont appelés champs de jauge non abéliens.

Les possibilités sont beaucoup plus riches qu'avec les champs abéliens mais aussi… beaucoup plus complexes ! Et même très complexe.

La quantification d'une théorie avec champs de jauge non abéliens est très difficile. Elle est même quasiment inextricable. La méthode de quantification dont nous avons parlé (les variables conjuguées et leur non-commutation) est totalement inadaptée. Les théoriciens ont mis au point des outils de quantification beaucoup plus sophistiqués. La méthode utilisée s'appelle la quantification par les "intégrales de chemin".

Il faut bien comprendre que quantifier à l'aide de ces nouveaux outils ne conduit pas à une théorie quantique différente. La théorie obtenue est parfaitement identique. Simplement ces outils sont les seuls adaptés aux champs de jauge non abéliens mais ils sont aussi plus compliqués à employer.

Une fois fait, on peut alors procéder comme avant. On applique la théorie des perturbations, puis la diagrammatique avec les règles appropriées, etc.

Pour quantifier un système quelconque, on procède comme suit :

- On choisit un (ou plusieurs) groupes de symétries. Par exemple le groupe $U(1)$.
- On classe les particules grâce au groupe en utilisant les propriétés de symétrie et les propriétés des particules (leur charge, leur spin, etc.).
- On écrit les équations (les physiciens parlent de lagrangien) pour les champs correspondant à ces particules en tenant compte de leurs regroupements. Les physiciens nomment ces regroupements des multiplets (ils sont un peu comme des vecteurs). Ils se comportent comme des objets mathématiques uniques (comme les vecteurs), ce qui permet l'écriture d'un lagrangien simple.
- On passe ensuite aux symétries locales, aux champs de jauge, etc.

La richesse des groupes non abéliens permet des classifications des particules beaucoup plus sophistiquées.

Si l'on a une théorie des champs renormalisable pour les particules initiales, alors l'introduction des champs de jauge peut conduire à une théorie renormalisable (pas toujours, nous le verrons avec la gravité).

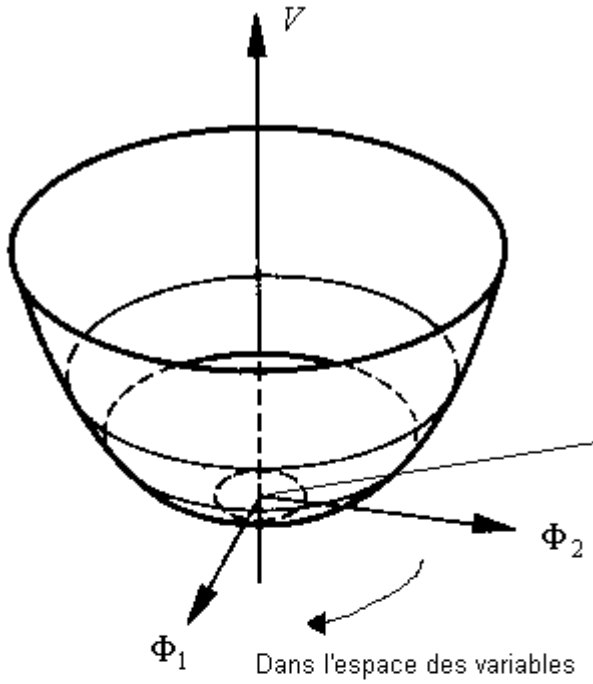
Les champs de jauge sont alors les vecteurs des forces, comme pour le champ électromagnétique. Mais nous avons vu que certains vecteurs étaient massifs. Malheureusement, si l'on introduit brutalement des champs de jauge massifs dans les équations en procédant comme nous venons de le voir, alors les théories quantiques obtenues ne sont pas renormalisables !

Il faut donc procéder autrement pour les champs de jauge massifs. Il faut trouver un "truc". Encore une fois, les théoriciens ont découvert un mécanisme subtil et extraordinaire.

Le mécanisme de brisure de symétrie

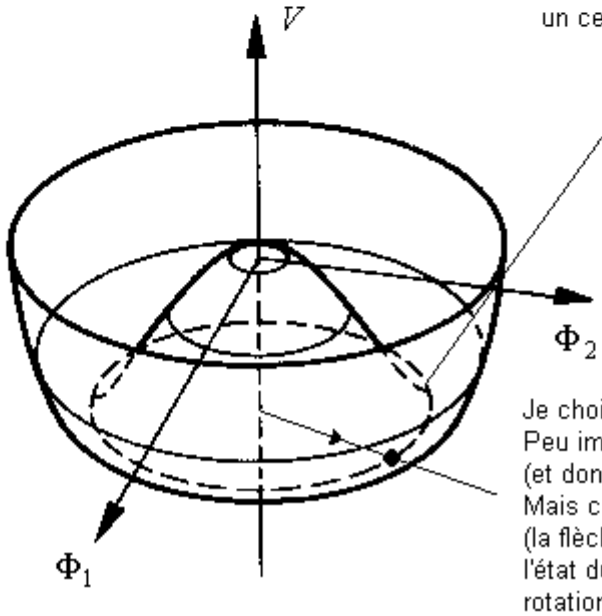
Il peut arriver, dans certains modèles, que l'état du vide ne soit pas défini de manière unique. On dit que le vide est dégénéré.

Voici par exemple deux potentiels tirés de deux modèles physiques. Les deux variables Φ_1 et Φ_2 sont des variables internes du modèle.



L'énergie (potentiel) est minimale ici. Le vide est donc représenté par les valeurs 0 des deux variables internes.

Dans l'espace des variables internes (puisqu'il s'agit de symétries internes et non pas géométriques), le système est invariant par rotation.



Le système a toujours une symétrie par rotation. Mais le vide n'est plus obtenu pour les valeurs (0,0) des variables.
L'état minimal d'énergie forme un cercle.

Je choisis un état particulier comme vide. Peu importe lequel puisque j'ai une symétrie (et donc une invariance) par rotation. Mais ce choix donne une direction privilégiée (la flèche joignant (0,0) au point représentant l'état du vide) et donc brise l'invariance par rotation. La symétrie est brisée.

Dans le cas de gauche, l'état d'énergie minimal est unique. Par contre, dans le cas de droite, l'état d'énergie minimal est un cercle. Tout état sur ce cercle peu correspondre au vide.

A priori, tout choix est équivalent. Le potentiel ci-dessus ayant une symétrie par rotation (symétrie $O(1)$ isomorphe à $U(1)$), n'importe quel choix de vide conduit à une physique équivalente. Il faut noter que les potentiels dessinés ne sont qu'une idéalisation. La situation peut être beaucoup plus

complexe et le modèle être invariant sous un groupe de symétrie plus complexe, par exemple $O(n)$. Mais le raisonnement reste valable, le vide étant dégénéré pour une variation continue de certaines variables, conduisant à une symétrie du vide dégénéré sous un sous-groupe du groupe complet. Dans l'exemple ci-dessus, le modèle pourrait être invariant sous $O(n)$ et l'image montre que le vide est invariant sous une symétrie $O(1)$.

On fait donc un choix de vide et on adapte les équations pour en tenir compte.

En regardant les équations, on constate alors qu'elles décrivent un nouveau modèle où sont présentes de nouvelles particules ! Elles sont appelées des bosons de Goldstone. Ce sont des particules sans masses.

Dans le cas des champs de jauge non abéliens, un mécanisme supplémentaire apparaît. Il est possible de réécrire les équations de façon à "absorber" les Goldstone dans les champs de jauge. C'est à dire que l'on effectue un regroupement des termes de façon à ce que les champs de jauge et les champs de Goldstone n'en forment plus qu'un seul.

Cette possibilité de réécrire les équations en faisant disparaître les Goldstone, montre que, dans ce cas, les bosons de Goldstone ne sont pas des particules physiques réellement observables.

Après cette réécriture, on constate que les champs de jauge (modifiés) qui sont décrits ont une masse. Une partie des champs initiaux est également modifiée ce qui peut changer ou donner une masse aux particules initiales du modèle.

Une autre conséquence est la diminution de la symétrie. Le choix d'un vide a brisé la symétrie du modèle. En choisissant, dans le potentiel ci-dessus, une direction privilégiée qui brise la symétrie de rotation. Par exemple, avec les groupes de symétries ci-dessus, le modèle n'est plus qu'invariant sous la symétrie $O(n-1)$.

Il est possible de démontrer que le mécanisme de brisure de symétrie préserve la renormalisabilité de la théorie. Il nous fournit donc un moyen de donner une masse aux champs de jauge sans tomber dans une théorie non renormalisable.

L'interaction faible

En plus de la force de gravitation et de la force électromagnétique, l'univers est gouverné par une autre force appelée interaction faible.

Cette interaction se manifeste dans certains processus de radioactivité et est responsable de la désintégration des particules fondamentales. Par exemple, le neutron est instable et se désintègre en quelques minutes en un proton, un électron et un anti-neutrino. Les noyaux contiennent des neutrons, ceux-ci ne se désintègrent pas (dans les atomes non radioactifs) car ils sont "stabilisés" par la force nucléaire.

Nous avons vu aussi que l'interaction faible brisait l'invariance par parité.

En 1967, Weinberg et Salam proposèrent un modèle qui unifie le champ électromagnétique et l'interaction faible. Ce modèle intègre tout ce que nous venons de voir. Il implique que le champ électromagnétique et l'interaction faible dérivent d'une force plus fondamentale appelée champ électrofaible.

Les particules du modèle regroupent les particules en deux doublets. On a d'une part l'électron et le neutrino, et d'autre part le muon (une espèce d'électron lourd et instable) et le neutrino muonique (une autre espèce de neutrino associé au muon).

Les symétries de ce modèle sont $SU(2) \times SU(1)$.

Il est possible d'écrire un modèle plus complet intégrant d'autres particules (comme le tau et le neutrino muonique ainsi que les quarks que nous verrons plus loin).

Le modèle autorise une brisure de symétrie. Les résultats sont les suivants :

- Un sous-groupe $SU(2) \times SU(1)$ est brisé.
- Il reste une symétrie associée au champ électromagnétique.
- Les électrons et les muons acquièrent une masse.
- Les champs de jauge se divisent en un photon de masse nulle (car la symétrie restante est non brisée). Et trois bosons W+, W- et Z. Les deux premiers portant une charge électrique. Ces trois bosons ont une masse et sont responsable de la portée finie de l'interaction faible.
- Il reste encore un boson de Higgs, résultat de la brisure de symétrie.

En dehors du boson de Higgs, non encore observé, toutes ces particules ont été observées et les prédictions de ce modèle ont été largement confirmées par l'expérience. Ce fut un grand succès des modèles des champs de jauge non abéliens.

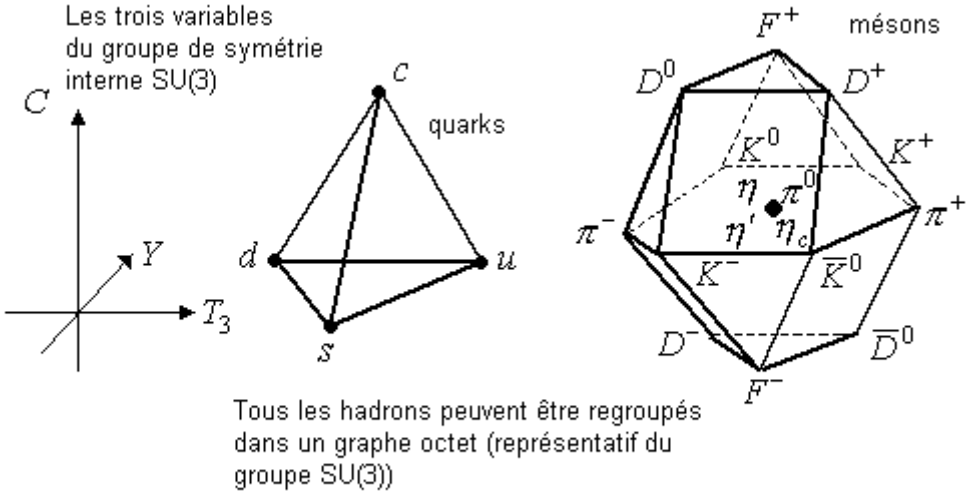
L'interaction forte

L'interaction forte est la quatrième et dernière force fondamentale connue. Elle est responsable de la cohésion des protons et des neutrons. La force nucléaire dérive directement de cette force. Le vecteur de la force nucléaire est le méson pi (appelé aussi méson de Yukawa qui fut le premier à proposer ce modèle pour la force nucléaire). Ce dernier est massif ce qui explique que la force nucléaire a une portée limitée au noyau. Mais la force nucléaire n'est qu'une manifestation de l'interaction forte, plus fondamentale.

La chromodynamique est un modèle qui incorpore toutes les interactions : électromagnétisme, interaction faible, interaction forte.

Si l'on regarde les particules fondamentales, on constate que les fermions peuvent se diviser en deux catégories. Les leptons (électrons, muons, neutrinos) et les hadrons (protons, neutrons, mésons pi, etc.). Les leptons sont insensibles à l'interaction forte contrairement aux hadrons. Nous avions déjà cité l'existence d'une "charge" leptonique. C'est ce qui explique que les électrons orbitent loin du noyau, maintenu par l'interaction électromagnétique. Tandis que les protons et les neutrons sont confinés dans le noyau de très petite dimension (car l'interaction forte est beaucoup plus puissante que l'interaction électromagnétique).

Il est possible de regrouper tous les hadrons, en utilisant leurs propriétés (charges, etc.), sous un groupe de symétrie $SU(3)$



Tous les hadrons peuvent être regroupés dans un graphe octet (représentatif du groupe SU(3))

Tout ce passe comme si les hadrons étaient composés de particules plus fondamentales appelés les quarks. Par exemple le méson pi serait constitué de deux quarks (u et d ou u et u), le proton de trois quarks (uud) et le neutron de trois quarks (udd).

Le quark u a une charge électrique 2/3 et le quark d une charge électrique -1/3. Cela correspond bien aux charges électriques des particules données en exemple :

Méson pi neutre : quark u (2/3) + anti quark u (-2/3) = 0

Méson pi chargé : quark u (2/3) + anti quark d (1/3) = +1

Proton : u (2/3) + u (2/3) + d (-1/3) = +1

Neutron : u (2/3) + d (-1/3) + d (-1/3) = 0

Le groupe de symétrie complet (hadrons + leptons et les trois interactions) est alors $SU(3) \times SU(2) \times SU(1)$

On procède comme avant. Seule l'interaction faible a une symétrie brisée. Les bosons de jauge de l'interaction forte restent donc non massif. Ces bosons sont appelés gluons, ils sont au nombre de huit et portent une charge appelée charge de couleur (rien à voir avec les couleurs habituelles) d'où le nom de chromodynamique. Le groupe $SU(3)$ s'appelle parfois groupe de couleur. Pour expliquer la masse des quarks, il faut faire dériver le groupe ci-dessus d'un groupe plus général qui est brisé.

Pour des raisons que nous verrons bientôt, ce modèle n'est pas encore totalement confirmé par l'expérience (mais les résultats sont tout de même déjà excellent).

Un phénomène appelé confinement (il oblige à avoir une charge de couleur nulle) empêche l'isolement des quarks ou des gluons. Mais avec des processus de collision à haute énergie, ceux-ci ont toutefois pu être observés et étudiés.

Il peut sembler étonnant que la portée des forces nucléaires soit si courte alors que les bosons de jauge correspondant (les gluons) sont sans masse. Cela s'explique par le confinement. Les gluons ne pouvant pas être isolés, ils ne peuvent pas transmettre la force entre deux particules fort éloignées. Cela ne peut se faire que via un intermédiaire "composite". Le plus simple (et le plus léger) étant le méson pi, composé de deux quarks. La masse du méson pi explique la portée finie de la force.

Nous avons parlé de beaucoup de particules. Il est temps d'en dresser un tableau complet avant d'aller plus loin.

3.6. Les particules

Le modèle standard représente la situation à ce stade. C'est un modèle de théorie de jauge avec des groupes de symétries correspondant à l'ensemble des particules connues et expliquant l'électromagnétisme, la force faible et, à ce qu'il semble, la force forte.

Voici l'ensemble des particules connues. Nous y avons également indiqué le graviton, vecteur hypothétique de la gravitation, et quelques particules composées.

Classe	Type	Génération (3)	Symbole	Nom	Forces (1)	Stabilité (5)	Antimatière
Fermions (matière)	Leptons	1	E	Electron	EM, Ff, FG	Oui	Positron
		2	μ	Muon	EM, Ff, FG	Non	Anti-muon
		3	τ	Tau	EM, Ff, FG	Non	Anti-tau
		1	υ_e	Neutrino électronique	Ff, FG	Oui	Antineutrino
		2	υ_μ	Neutrino muonique	Ff, FG	Oui (2)	Antineutrino
		3	υ_τ	Neutrino tauique	Ff, FG	Oui (2)	Antineutrino
	Hadrons (quarks)	1	τ	Down / Bas	Toutes	Oui (4)	Anti-d
		2	s	Strange / Etrange	Toutes	Non	Anti-s
		3	b	Bottom / Fond	Toutes	Non	Anti-b
		1	u	Up / Haut	Toutes	Oui (4)	Anti-u
		2	c	Charm / Charme	Toutes	Non	Anti-c
		3	c	Top / Sommet	Toutes	Non	Anti-t
Bosons (vecteurs des forces)			γ	Photon	EM		
			g	Gluons (8 différents)	FF		
			W^+	Boson intermédiaire	Ff		W^-

			Z_o	Boson intermédiaire	Ff		
				Graviton	FG		
			H	Boson de Higgs	Responsable de la masse des autres particules		
Particule composée			n	Neutron	Composé de u d d		
			p	Proton	u u d		
			π_0	Méson pi neutre	u + anti-d		

1. EM = électromagnétique

   Ff = faible

   FF = forte

   FG = gravitation

   A chaque force est associée une charge :

   EM - charge électrique

   Ff - charge faible

   FF - couleur (3 différentes)

   FG - masse

   La couleur est une charge "confinée" comme les quarks, la somme des couleurs donne toujours la charge neutre (le blanc). Tout comme dans la composition additive des couleurs (les vraies), d'où le nom. Les effets de cette charge ne peuvent donc être observés qu'à très haute énergie (dans des collisions par exemple).

   Il existe aussi quelques autres "charges" résultant de symétries. Tel que la charge leptonique, hadronique, muonique,…
2. Si les neutrinos ont une masse même faible, ils peuvent se transformer l'un en l'autre ("oscillation" des neutrinos).
3. Des charges sont également associées aux différentes générations. On parle ainsi de particules charmées ou étranges. Toute la matière "ordinaire" (celle qui compose les atomes) est composée de la génération 1.
4. Certaines théories super symétriques (voir plus loin) prévoient une instabilité à très long terme.
5. Une particule instable se désintègre en particules plus légères au bout d'un certain temps (généralement très court).

Parmi les particules présentées ci-dessus, il reste à observer le boson de Higgs et le graviton. Pour ce dernier, l'observation sera difficile voir impossible à cause de la faiblesse de la force gravitationnelle. Pour le boson de Higgs, cela sera probablement possible dans un avenir proche (s'il existe !). Le boson de Higgs est une particule qui apparaît dans la théorie lors de l'utilisation de la brisure de symétrie, il est responsable de la masse des particules. Si l'on n'observe pas de bosons de Higgs à l'état libre, on pense que cela est du au fait qu'il ne peut peut-être se manifester que sous forme de particules virtuelles. A suivre.

4. Les problèmes

Si le modèle standard a donné des résultats extraordinaires, riches, précis, largement confirmés par l'expérience, il faut bien admettre qu'il existe encore quelques problèmes. Nous allons les passer en revue.

4.1. Les paramètres libres

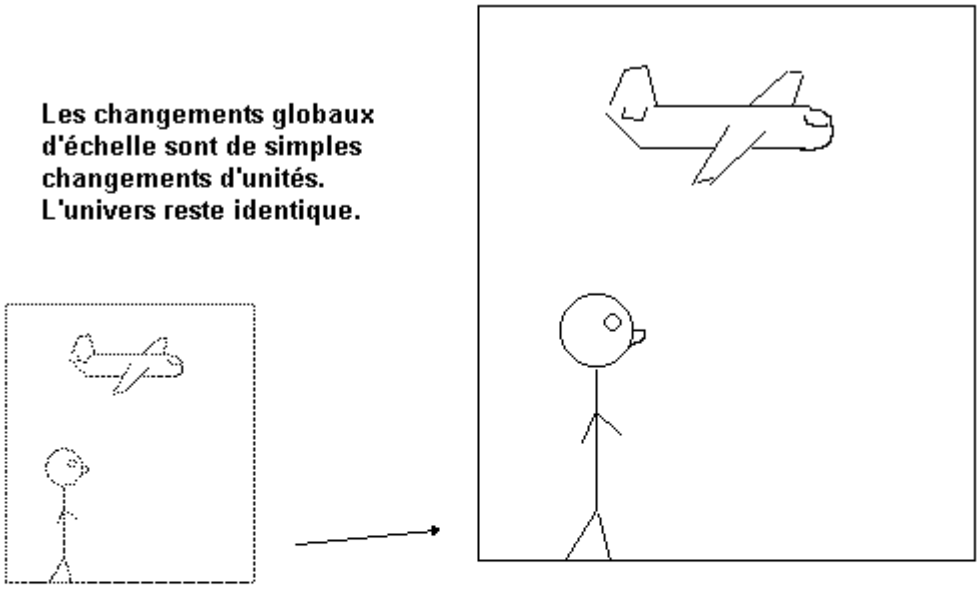
Nous avons vu que pour résoudre le problème des valeurs infinies, on introduisait le mécanisme de renormalisation. Celui-ci consiste à introduire un certain nombre de paramètres, appelés les paramètres libres, dont la valeur est déduite de l'expérience. Puis à réexprimer la théorie en fonction de ces paramètres plutôt qu'en fonction des paramètres nus.

Cela signifie que ces paramètres ne sont pas déduits de la théorie. Il en est ainsi de la masse de l'électron, de la charge électrique, de la masse des quarks et du boson de Higgs, etc.

D'autres paramètres sont déduits de l'observation. Tel que la valeur de la vitesse de la lumière dans le vide, la constante de Planck,…

Pourquoi ces valeurs et pas d'autres ? Une théorie correcte devrait pouvoir prédire ces valeurs.

Bien sûr, une partie de ces constantes peut-être éliminée par le choix des unités. Par exemple, la vitesse de la lumière en mètres par seconde vaut 300000000, tandis qu'en kilomètres par seconde elle vaut 300000. Donc sa valeur exacte est fonction des unités de mesure. Seul importe le rapport entre les différentes constantes.



Mais le choix des unités est limité : longueur, temps, masse, charge électrique,… On ne peut éliminer que quelques constantes fondamentales. Par exemple, les physiciens travaillent souvent avec un choix d'unité tel que la vitesse de la lumière et la constante de Planck valent un. Ce qui simplifie les formules. Avec ce choix d'unité une masse est égale à une énergie, une longueur à un temps, etc. Ce ne sont donc pas de réelles constantes fondamentales. Le fait que la vitesse de la lumière aie une valeur précise n'est dû qu'à note système de mesure et le rapport entre nos propres vitesses (quand je marche, j'avance à quelques kilomètres par heure) et celle de la lumière. Et les vitesses à notre échelle quotidienne dépendent elles même des mécanismes physiques qui nous font fonctionner. Le seul aspect physique pertinent n'est pas la valeur de la vitesse de la lumière mais le fait qu'elle existe, c'est à dire qu'elle ne soit pas infinie. On a vu aussi que cela revenait à voir l'univers comme "local".

Cela ne permet donc d'éliminer que quelques constantes. Les modèles, que nous avons vu (la chromodynamique, par exemple), introduisent aussi quelques contraintes et relations entre constantes. Entre les masses des quarks par exemple. Mais il reste tout de même pas mal de paramètres libres comme la masse de l'électron et du boson de Higgs, par exemple.

Mais les paramètres libres ne sont pas les seules choses déduites de l'observation. Par exemple, pourquoi le groupe de symétrie de la chromodynamique doit-il être $SU(3) \times SU(2) \times SU(1)$? Pourquoi ces groupes là et pas d'autres ?

De même, pourquoi une telle pluralité de particules (on a vu que le problème est lié aux symétries) ? On appelle d'ailleurs parfois le tableau des particules précédent le "zoo" des particules !

4.2. La gravité quantique

Dans le modèle standard, nous avons montré que nous pouvions unifier en une seule théorie quantique les interactions électromagnétiques, faibles et fortes. Mais nous n'avons pas parlé de la gravité. Comment inclure celle-ci ?

En réalité, quantifier la relativité générale n'est apparemment pas un gros problème. On peut procéder en quantifiant directement le champ de gravité de la relativité générale. Ou, ce qui est encore mieux, utiliser les théories de jauge.

On a vu que la relativité restreinte était invariante sous le groupe de symétrie (global) de Poincaré. On procède alors comme on a déjà vu, on introduit des champs de jauge de manière à obtenir une théorie localement invariante. La théorie (classique) ainsi obtenue n'est rien d'autre que la relativité générale. Puis on passe à la quantification du champ de jauge.

On retrouve la signification physique profonde des champs de jauge : la gravité n'est qu'une conséquence de la symétrie locale au groupe de Poincaré de la relativité restreinte, elle-même une conséquence de l'absence de référentiel absolu et de l'existence d'une vitesse limite.

Cette quantification n'est même pas trop complexe en soit car le groupe de Poincaré n'est pas des plus difficiles.

Malheureusement, la théorie quantique de jauge ainsi obtenue est non renormalisable !

Il apparaît donc des valeurs infinies que l'on ne peut éliminer qu'au prix d'une introduction d'un nombre infini de paramètres rendant la théorie totalement non prédictive et … inutilisable. Comme nous l'avons déjà vu.

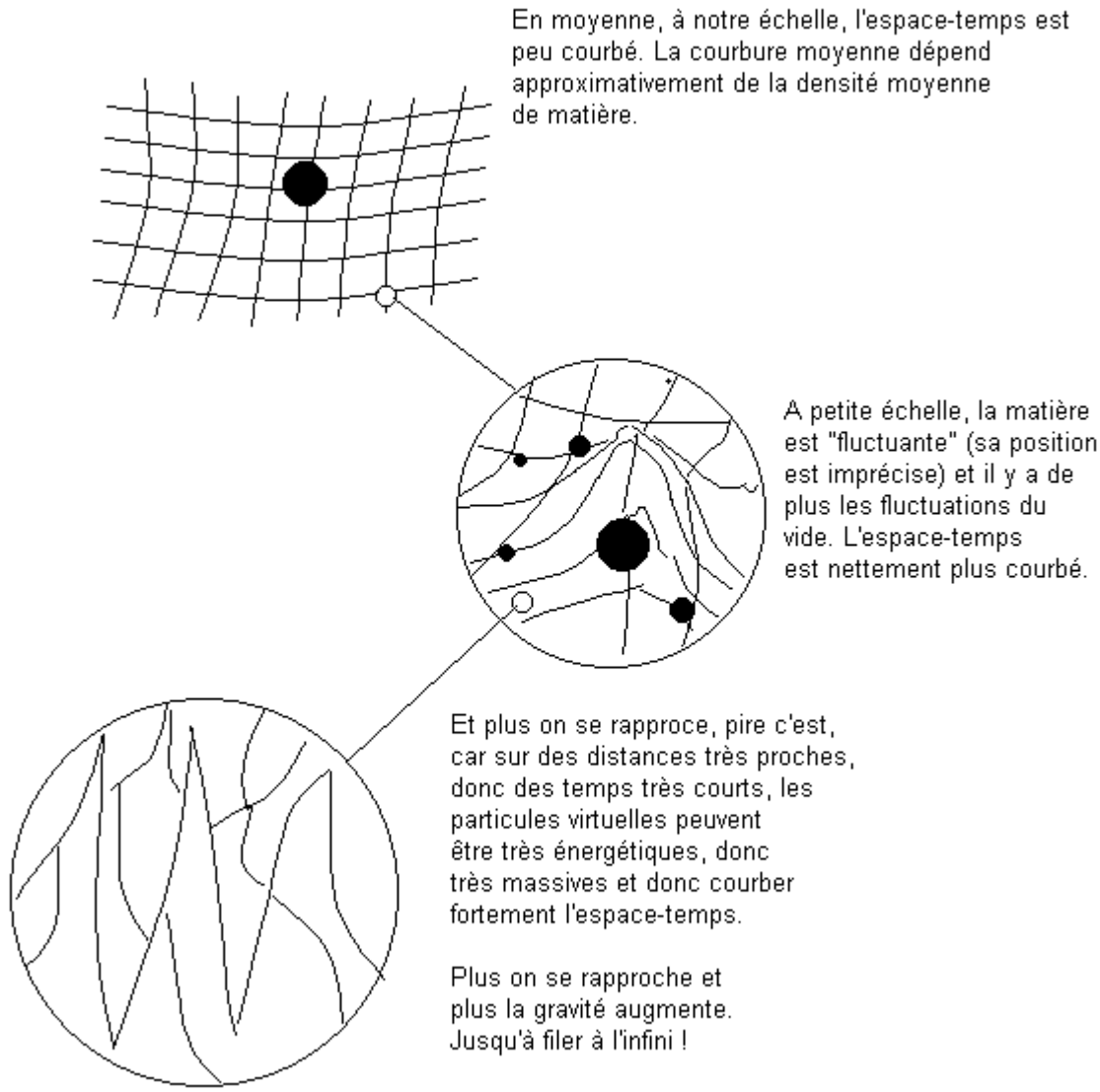
On peut comprendre intuitivement l'origine de ce phénomène. Nous avons vu que le vide était plein de fluctuations conduisant à une énergie du vide infinie. Nous avons dit que cela n'avait pas d'importance car on mesurait l'énergie des états par rapport au vide et non pas l'énergie du vide lui-même. On peut en effet fixer le zéro de notre échelle d'énergie où l'on veut.

Mais la gravité est sensible à toute forme d'énergie puisqu'elle est sensible à la masse et que masse et énergie sont équivalents en relativité. Comme ces fluctuations existent bel et bien , leur énergie infinie doit agir sur la gravité, du moins aux très petites échelles. Donc, à petite échelle la gravité doit devenir infinie, ce qui est absurde !

Lorsque l'on regarde les diagrammes de Feynman pour la gravité, on voit que le graviton (la particule quantique associée au champ de gravitation) peut interagir avec lui-même. Cela provoque la naissance d'autres gravitons qui interagissent à nouveau… etc. Cela conduit à une espèce de réaction en chaîne qui est à l'origine des infinis.

A faible énergie, c'est à dire pour une attraction gravitationnelle à grande distance, on peut conserver uniquement les premiers termes dans les développements perturbatifs. C'est à dire les diagrammes en arbre (sans boucle). Là, il n'y a aucun problème. En ignorant les effets quantiques (c'est à dire en prenant les moyennes en utilisant les probabilités), on retrouve parfaitement le comportement de la relativité générale.

Mais ce qui nous intéresse ce sont les comportements à petite échelle, c'est à dire à haute énergie. Là, il faut ajouter de plus en plus de termes au développement perturbatif. Les divergences se démultiplient et toute prédiction théorique devient impossible.



Lorsque l'on étudie un peu plus attentivement les aspects mathématiques de ce phénomène, on constate que le problème est lié au spin du graviton. Nous avons dit que le spin associé au champ de la gravitation était égal à deux. Et effectivement, on découvre que toute théorie des champs incluant un boson de spin 2 est non renormalisable.

Nous avions eu une difficulté analogue avec les champs de jauge massifs. Les théoriciens avaient découvert le "truc" de la brisure de symétrie. Existe-t-il un truc analogue pour le spin deux ? Malheureusement non. Tel quel, le problème est réellement incontournable. Par exemple, les théoriciens avaient remarqué que la gravité quantique possède un très grand nombre de symétries. Ils avaient espéré que le nombre élevé de symétries permettrait d'obtenir des annulations des divergences. Un phénomène analogue se passe dans une certaine mesure pour l'électromagnétisme (identités de Ward). Mais les espoirs furent vites déçus.

Les théoriciens ont bien réussi à tirer quelques informations de ces théories non renormalisables, mais cela reste très limité. Si on veut progresser, si on veut trouver une théorie renormalisable, il va falloir trouver tout autre chose. Le comportement à très petite échelle de la gravité doit être très différent de ce qui est décrit par la simple quantification du champ de la gravitation de la relativité générale. Mais quel comportement ? Là est tout le problème !

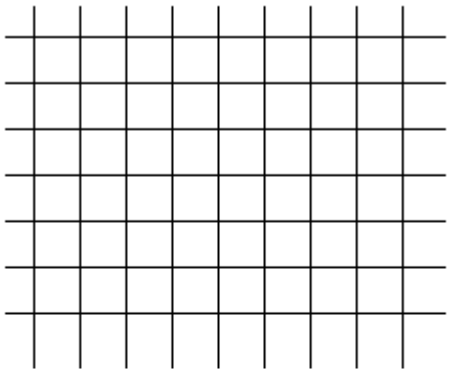
La théorie de Kaluza - Klein est-elle une solution ? Non. Cette théorie est, avant tout, une théorie analogue à la relativité générale. Elle décrit donc aussi un champ de spin deux. Et la théorie reste non renormalisable.

Nous avons vu qu'en augmentant le nombre de dimensions, les divergences disparaissaient en électrodynamique. N'est-ce pas le cas de Kaluza - Klein ? Ne pourrait-on pas passer à des dimensions encore plus élevées ? Ce n'est malheureusement pas la solution. Non seulement le problème est lié au spin deux, et augmenter les dimensions ne change rien à ce problème. Mais pire encore, dans le cas de la relativité générale, augmenter les dimensions empire encore le problème des singularités ultraviolettes ! Contrairement au cas de l'électrodynamique comme nous l'avons vu. Intuitivement cela est du au fait qu'avec une dimension supplémentaire les gravitons ont plus de possibilité encore de réagir entre eux empirant la réaction en chaîne dont nous avons parlé.

La super symétrie et la théorie des cordes, que nous allons voir plus loin, sont dans la droite ligne des théories de jauge mais il y a une autre approche possible.

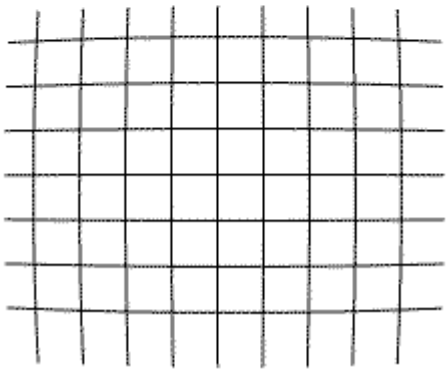
Une analyse précise de la relativité générale et des postulats de la physique quantique montre qu'il y a une incompatibilité. La relativité générale manipule des événements qui sont parfaitement localisés dans l'espace et le temps. Cette possibilité entre en conflit direct avec les relations d'incertitudes de la physique quantique. Cela nous oblige à réviser complètement les notions géométriques habituelles. A très petite échelle, les détails ne peuvent pas être connus et n'ont même pas de signification physique. Une géométrie qui décrit un tel espace-temps est assurément spéciale.

Les théoriciens ont développé une branche très active des mathématiques, appelée géométries non commutatives, pour décrire ces concepts. C'est un domaine vaste, très complexe et très abstrait. Difficile à décrire. On arrive ainsi à décrire la géométrie en termes de boucles, de réseaux de spins, de "mousses" de spin,… Les boucles forment des tresses, des nœuds,…

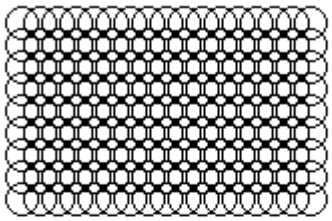


géométrie euclidienne. Les points sont repérés par des coordonnées cartésiennes que l'on peut représenter par ce grillage.

On peut bien sur resserrer le grillage pour repérer plus de points. Il n'y a pas réellement de trous dans l'espace-temps !



géométrie riemannienne (relativité générale), l'espace-temps est courbe et les coordonnées sonr repérées sur un grillage courbé.



Géométrie décrite par un réseau de boucles. Les points sont repérés sur les boucles. Attention les boucles ne se recoupent pas (sinon il y aurait ambiguité pour localiser le point d'intersection : sur quelle boucle ?)

Les boucles ont un diamètre approximativement égal à la longueur de Planck (voir plus loin). Cet espace-temps est plein de trous ! Il est idéal pour décrire des géométries où les petits détails n'ont pas de signification physique.

On peut imaginer des géométries encore plus tordues où le grillage forme des noeuds, des tresses, etc...

A l'aide de ce type de géométrie on peut espérer arriver à une description satisfaisante de la relativité générale à des échelles où les phénomènes quantiques ne sont plus négligeables. Ce type d'approche est appelé quantification canonique de la gravitation. Des progrès considérables ont pu

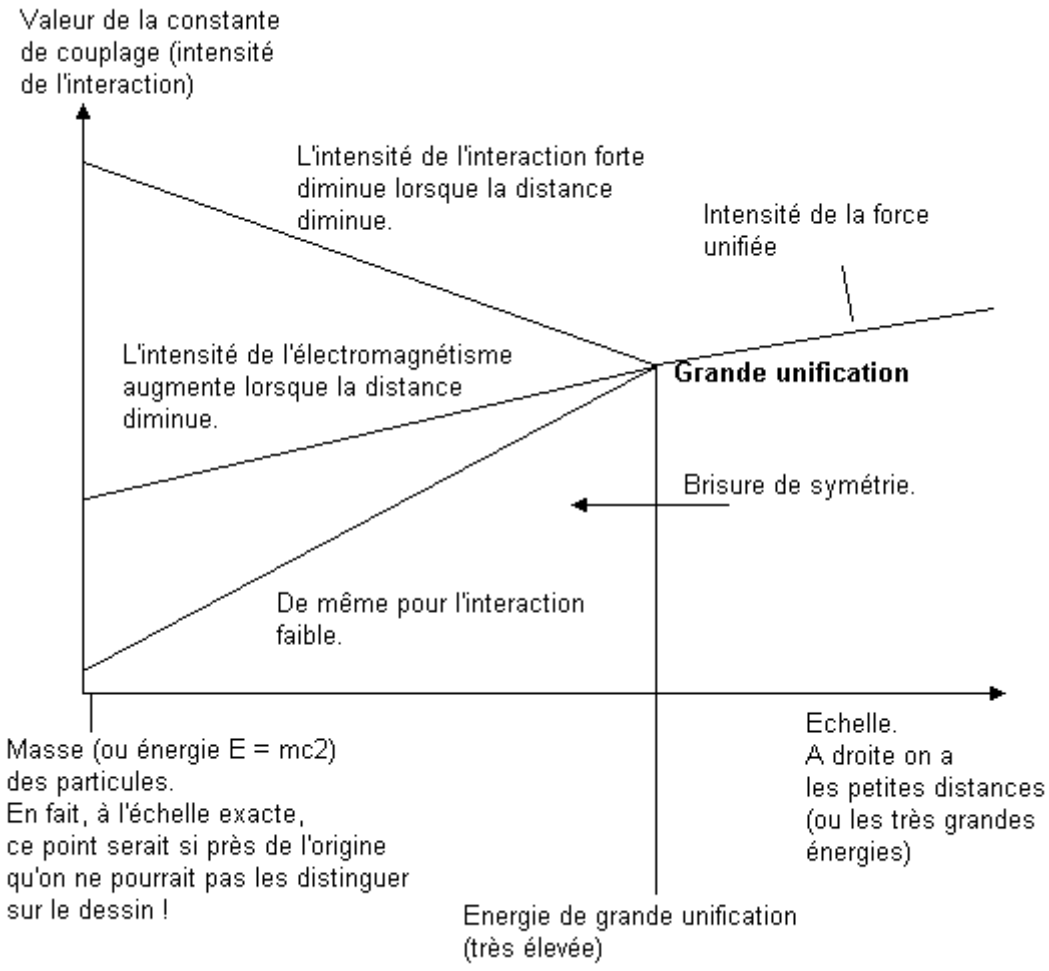
être obtenus. Il semble actuellement que ce type d'approche puisse aboutir et qu'une certaine convergence avec les théories des cordes se dessine.

Le but de cette étude n'est pas d'aborder ces aspects hautement abstraits que sont ceux des géométries non commutatives. Nous allons donc continuer notre petit bonhomme de chemin vers les théories des cordes.

4.3. Le problème de la hiérarchie

L'étude des interactions montre que leur intensité varie avec l'énergie mise en jeu (voir aussi plus loin avec l'interaction forte). Cette intensité s'appelle constante de couplage. Notons en passant que le terme de "constante" est un peu mal choisi puisque ces valeurs varient !

On peut tracer un graphique représentant la valeur des constantes de couplage.



A haute énergie ces interactions ont une intensité identique. C'est normal puisque ces interactions dérivent toutes d'une interaction plus fondamentale comme nous l'avons vu.

Puis, lorsque l'énergie diminue, le mécanisme de brisure de symétrie provoque une différentiation de ces forces.

L'énergie où l'unification de ces interactions a lieu est appelée énergie de grande unification.

Premier problème

Nous avons vu que la masse des particules était une conséquence du mécanisme de brisure de symétrie. Mais nous avons vu aussi que la masse réelle exacte de ces particules n'étaient pas déduite de la théorie mais de l'observation.

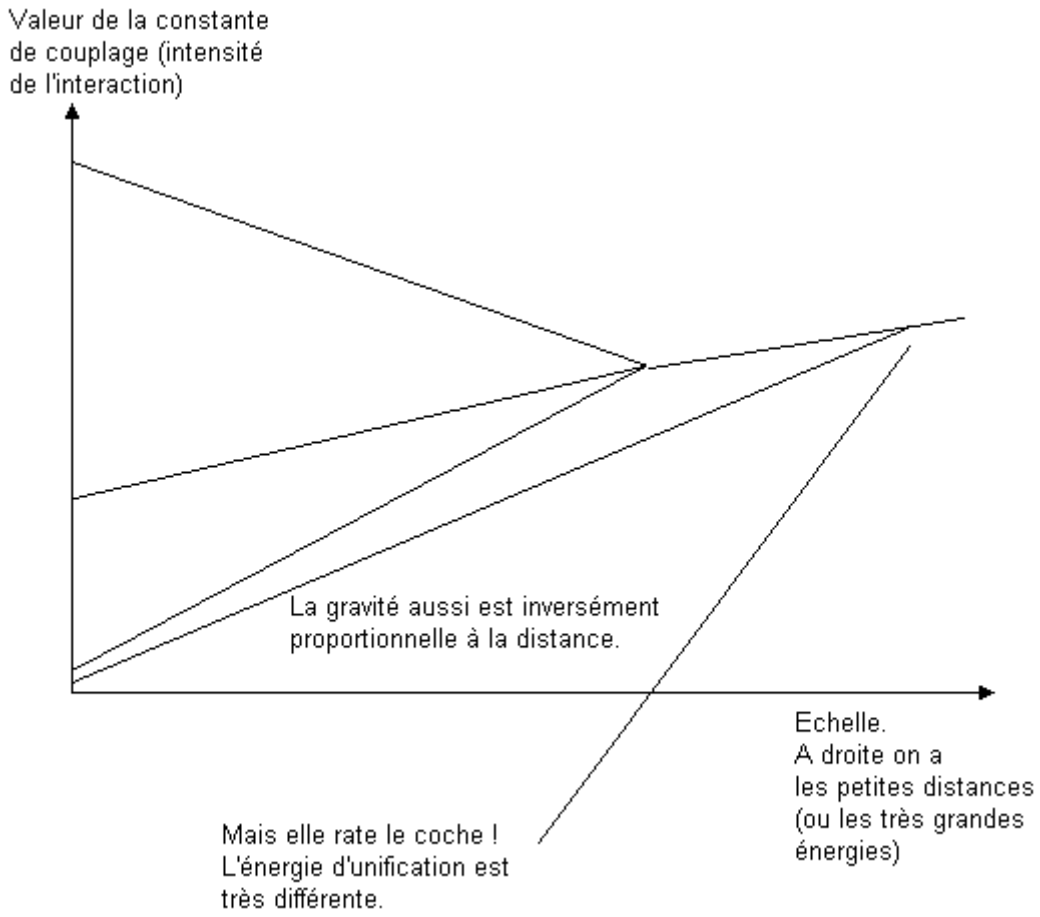
L'écart entre les masses et l'énergie de grande unification est considérable. Cette dernière est vraiment énorme (plusieurs milliards de milliards de fois la masse des plus lourdes particules connues, encore largement inaccessible à notre technologie). Si la brisure de symétrie à cette énergie provoque l'apparition des masses, comment ce fait-il qu'il y ait un écart aussi important entre masses des particules et énergie de grande unification ?

L'écart est vraiment énorme (voir le dessin ci-dessus), et son origine est un mystère. Ce problème est appelé le problème de la hiérarchie.

En réalité, le problème de cet écart entre échelle d'énergie et masse des particules n'est pas seulement "esthétique" ou qualitatif. Il y a de bonnes raisons techniques dans les théories des champs pour affirmer que les masses de certaines particules (comme le boson de Higgs) a tendance à être proche de l'énergie de grande unification.

Deuxième problème

La gravité a aussi une intensité variable suivant la distance et donc l'énergie.



Zut, elle passe à coté des autres !

Si les quatre forces dérivent d'une même force fondamentale. Comment ce fait-il que la brisure de symétrie, la séparation en quatre forces différentes, ne se fait pas en même temps, à la même énergie ?

Notons que la gravité a une force comparable aux autres interactions à très courte distance appelée distance de Planck. C'est aussi la distance où les effets quantiques de la gravitation devraient se manifester.

D'une manière générale, on peut construire, avec les constantes fondamentales, plusieurs combinaison conduisant à :

- la longueur de Planck : environ $1.61 \cdot 10^{-33}\ cm$ (32 zéros après la virgule !).
- l'énergie de Planck : environ $1.22 \cdot 10^{28}\ eV$ (eV est l'électron-volt : l'énergie acquise par un électron dans un potentiel de un volt), soit un milliard et demi de Joule. Ce qui est colossal pour une particule qui aurait cette énergie, elle atteindrait la masse d'un petit insecte (un acarien) !
- temps de Planck : environ $5.39 \cdot 10^{-44}\ cm$ (43 zéros après la virgule), temps mis par la lumière pour parcourir la longueur de Planck.

Ces échelles fixent le domaine où le modèle standard n'est plus d'application. D'après la théorie du Big Bang, lorsque l'univers avait l'age du temps de Planck (c'est vraiment peu), l'univers était tellement dense que la gravité avait une intensité comparable aux autres forces et le modèle standard ne peu s'y appliquer. On l'appelle parfois le mur de Planck.

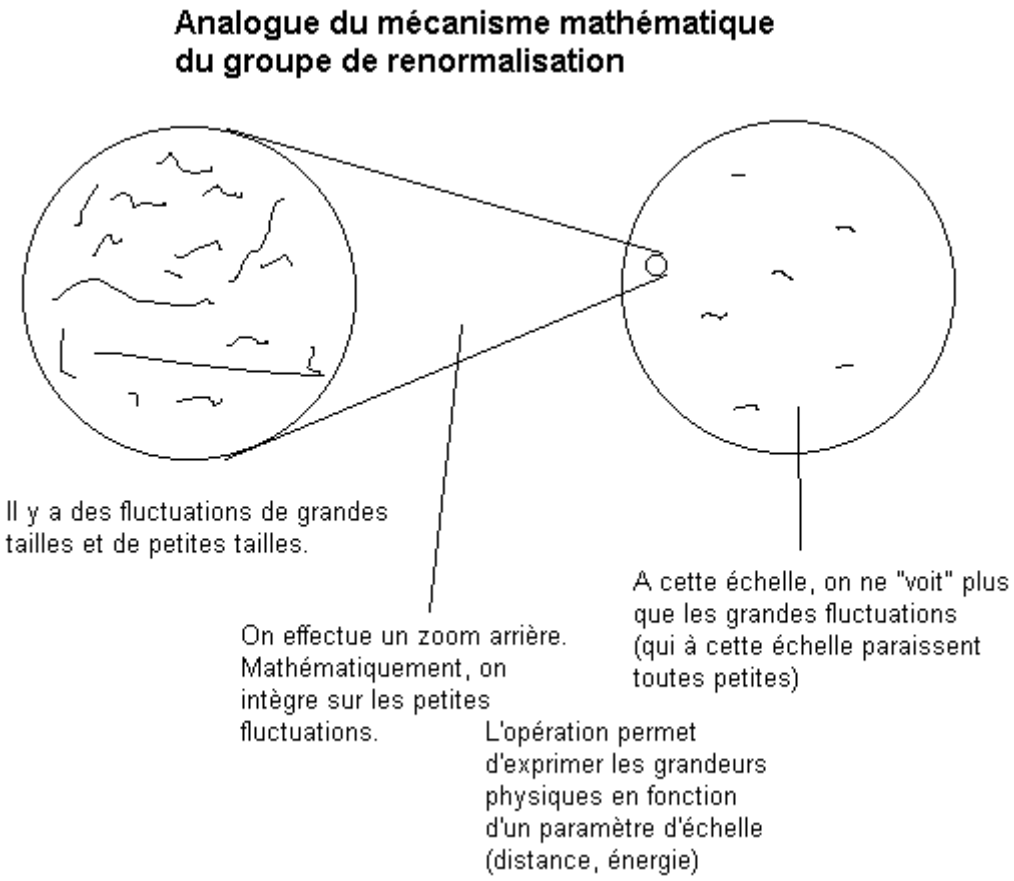
4.4. L'interaction forte

Sur les graphiques précédents on voit que la constante de couplage de l'interaction forte varie en sens inverse des autres.

Effectivement, lorsque les particules sont très proches (à haute énergie) la constante de couplage est petite. Cela signifie que l'interaction forte augmente avec la distance. Une situation assez extraordinaire.

Une étude mathématique approfondie de la renormalisation montre que celle-ci forme un groupe appelé bien évidemment groupe de renormalisation.

Les aspects mathématiques du groupe de renormalisation sont assez complexe mais peuvent être vus de manière intuitive suivante.



Pour résoudre un système, on peut "intégrer" les fluctuations à petite échelle et il ne reste alors plus que des fluctuations à grande échelle. Un peu comme si on regardait le système de loin. Ce processus de "rétrécissement" est analogue à la renormalisation où l'on décrit les particules réelles comme étant globalement les particules nues plus les fluctuations.

Le grossissement (ou le rétrécissement) peut se faire d'un facteur quelconque. Le paramètre de grossissement (paramètre d'échelle) est analogue à un groupe qui autorise un traitement mathématique sophistiqué.

Des généralisations de cette théorie ont permis de montrer que certains modèles pouvaient être doués du phénomène de diminution des interactions lorsque les distances diminuent. On les appelle des théories "asymptotiquement libres". Ce type de comportement s'étudie sur les équations du groupe de renormalisation appelées équations de Callan - Symanzik. Encore un beau succès théorique.

Quelques résultats remarquables ont pu être montré. Par exemple que seule une théorie avec des champs de jauge non abéliens peut présenter ce type de comportement et que la chromodynamique est une théorie asymptotiquement libre.

On a vu que la théorie des perturbations n'était pas applicable lorsque les constantes de couplages sont grandes. Il est alors nécessaire de trouver des méthodes exactes ou d'autres méthodes d'approximation. On les appelle des méthodes non perturbatives.

A haute énergie, par exemple dans des collisions très violentes, la constante de couplage de l'interaction forte est faible. La théorie des perturbations est applicable et de beaux résultats ont pu être trouvé par les théoriciens (et confirmés par l'expérience).

Mais les problèmes se posent à basse énergie. Par exemple, dans les hadrons les quarks sont liés avec une énergie relativement faible. Donc, la théorie des perturbations n'est pas applicable à l'étude des hadrons (par exemple pour calculer leur masse à partir de celle des quarks).

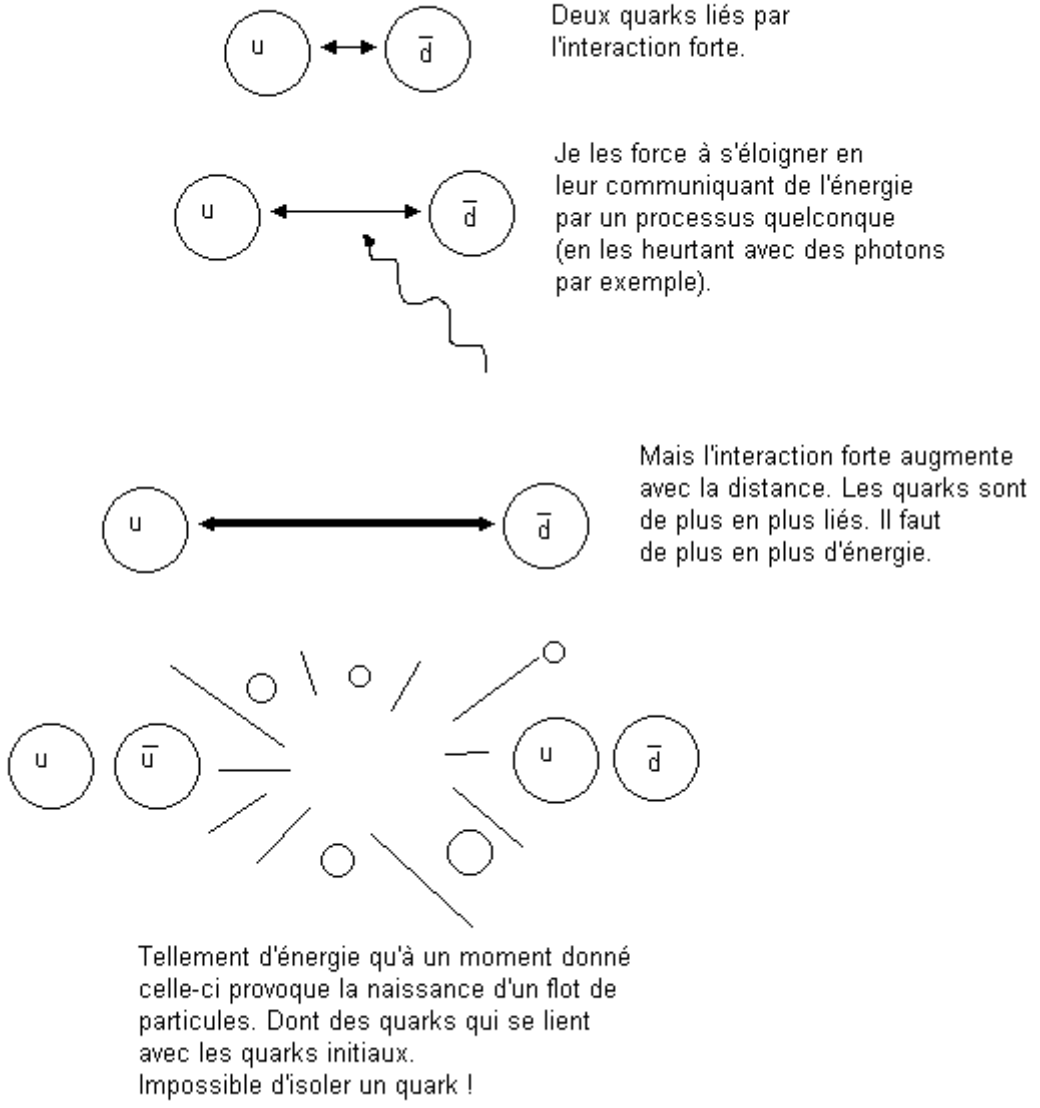
Diverses méthodes ont été développées tel que des développements sur "réseaux" (on place une grille sur l'espace et on ne calcule les valeurs qu'aux intersections) qui autorisent des traitements numériques sur ordinateur. Mais les résultats, bien qu'encourageants, sont encore mitigés. La précision tournant seulement autour de quelques dizaines de pour cents.

Bien entendu ces difficultés sont essentiellement d'ordre mathématique. Ce ne sont pas de vraies difficultés théoriques fondamentales. Elles ne remettent pas en cause le modèle standard mais nécessitent un grand effort (encore en cours) pour améliorer les outils mathématiques.

Une des conséquences les plus extraordinaires des théories asymptotiquement libres est le confinement. C'est à dire que les quarks ne peuvent être isolés et restent confinés dans les hadrons.

On peut comprendre intuitivement le processus comme suit.

Le confinement des quarks



Supposons deux quarks liés dans un hadron. J'essaie alors de les séparer par un processus quelconque en leur communiquant de l'énergie. Tout comme on ionise un atome d'hydrogène en fournissant de l'énergie à son électron pour l'éjecter de l'atome (par exemple avec des rayons ultraviolets) en photochimie ou en photoélectricité.

Plus les quarks s'écartent et plus l'intensité de l'interaction forte augmente. Il devient de plus en plus difficile de les écarter et il faut de plus en plus d'énergie.

Mais n'oublions pas que l'énergie c'est aussi de la masse. A un moment donné, il faut communiquer tellement d'énergie que celle-ci se convertit en particules ! Des nouveaux quarks sont créés et au lieu de se retrouver avec deux quarks isolés je me retrouve avec seulement de nouveaux hadrons.

Malgré les importants développement du groupe de renormalisation et des méthodes non perturbatives et malgré le raisonnement intuitif précédent, le confinement dans les hadrons reste encore assez mal compris (quantitativement) puisqu'il s'agit là d'un problème de basse énergie. Lorsque l'on a communiqué suffisamment d'énergie pour entrer dans le domaine de validité des calculs perturbatifs, cela fait longtemps que des particules sont créées.

5. La théorie des cordes classiques

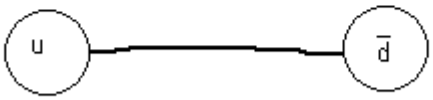
5.1. Les cordes et l'interaction forte

En 1968 Veneziano présenta une idée nouvelle.

Pourquoi ne pas modéliser l'interaction forte par une corde ? A cette époque on avait déjà constaté le phénomène de confinement et d'absence apparente de force dès que les constituants étaient très proches.

Une manière simple d'essayer de modéliser ce comportement était d'attacher les constituants par une corde !

Deux quarks liés par une corde. Lorsqu'ils sont proches la corde est peu tendue. Ils se comportent comme des particules libres. Bien entendu, les cordes de Veneziano sont des objets mathématiques un peu plus compliqué qu'une simple corde de nylon !



Lorsqu'ils s'éloignent la corde se tend.

Puis se brise. L'énergie accumulée est libérée.

L'énergie libérée crée des quarks aux bouts de chaque morceau de corde (plus d'autres particules).
Les cordes reprennent leur taille normale.

Bien entendu, cette idée intuitive reçut un support mathématique concret. Et on modélisa mathématiquement cette notion de corde quantique.

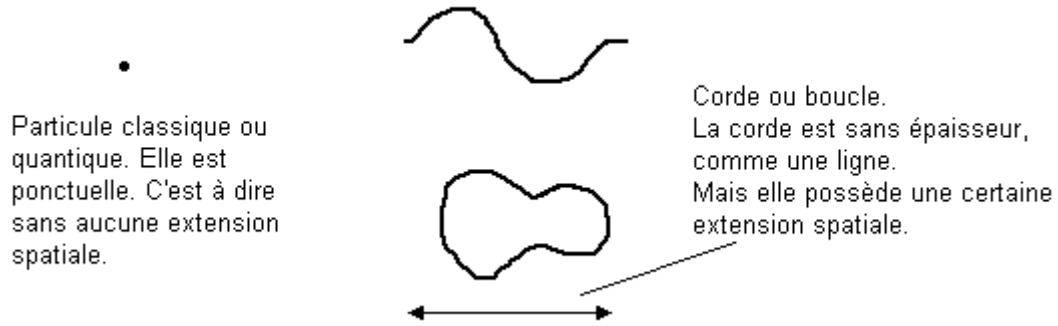
Le modèle ne donna pas de résultat très probant et tomba rapidement en désuétude suite à l'avènement de la chromodynamique quantique en 1973.

5.2. Les cordes classiques

La théorie des cordes classiques fut imaginée dans les années septante. Ce nom de "classique" n'est pas utilisé dans le sens "non quantique" mais par opposition avec les supers cordes que nous verrons plus tard.

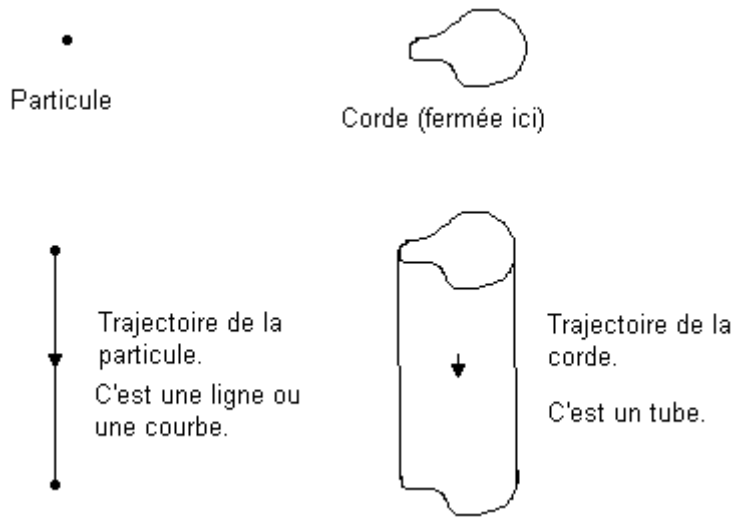
L'idée de Veneziano n'était pas tombée dans l'oreille d'un sourd. La modélisation de l'interaction forte par des cordes est équivalente à une tentative de modéliser les gluons par des cordes. Et si toutes les particules pouvaient être modélisées de cette manière ?

Une corde est comme une particule mais au lieu d'être concentrée en un point elle possède une certaine extension spatiale.



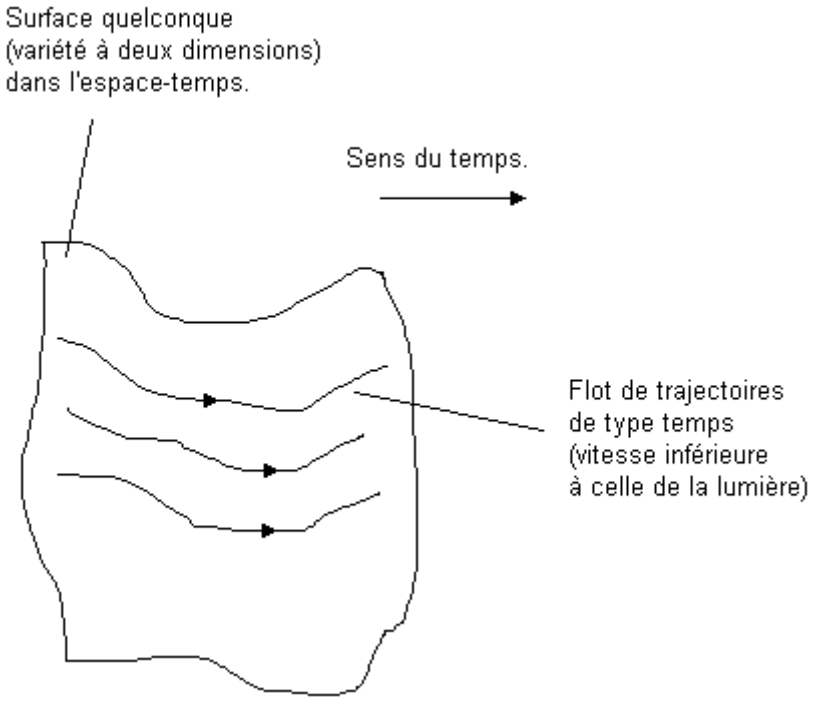
A ce moment, les théoriciens ne pensaient pas encore à l'unification. Ils se demandaient seulement si cette idée pouvait être intéressante et si elle pouvait trouver des applications dans certains contextes physiques.

Au cours du temps, une particule parcourt une trajectoire courbe appelée, en relativité, une ligne d'univers. Tandis qu'une corde trace une surface appelée surface d'univers.



Tout cela peut se passer dans un espace courbe. Donc la surface est une variété à deux dimensions dans un espace-temps à quatre dimensions. Bien sur, cette variété représente une trajectoire

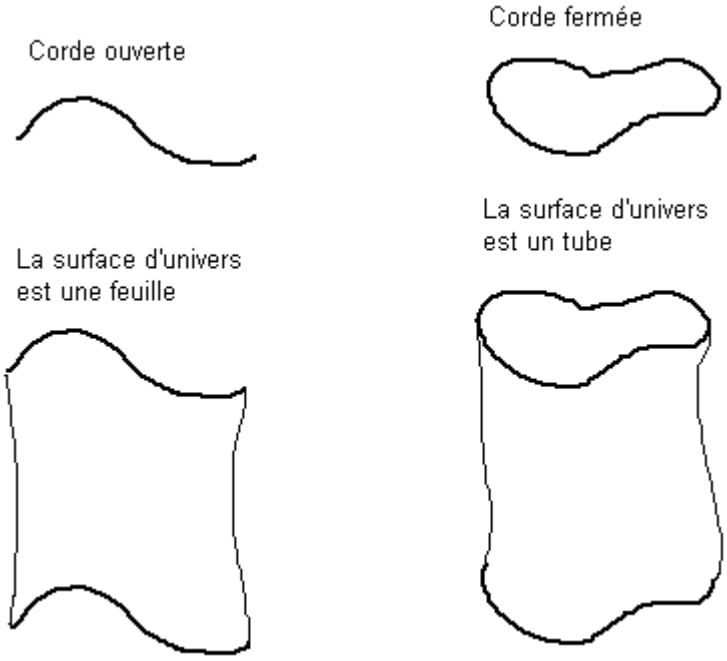
physique pour une corde. Donc, il doit être possible de tracer des trajectoires (des lignes) de type temps sur la surface. C'est à dire des trajectoires correspondant à une vitesse inférieure (ou égale) à la vitesse de la lumière.



Il est bien sûr possible d'écrire une expression mathématique précise de ce type de surface.

L'écriture d'un modèle se fait alors comme en théorie des champs traditionnelle. On choisit un type de corde (voir plus loin), on écrit un lagrangien (comme nous avons vu) pour l'objet mathématique représentant les cordes, puis on passe à la quantification et la suite.

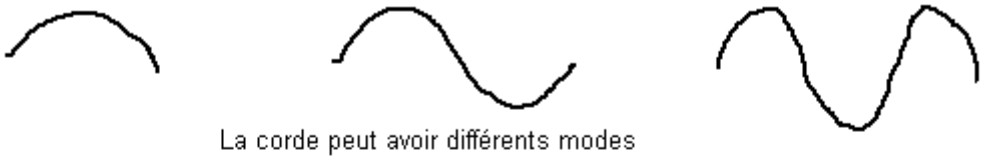
La théorie permet d'étudier deux types de cordes. Des cordes ouvertes et des cordes fermées.



Une corde fermée, comme une boucle, a comme surface d'univers un tube.

Les cordes peuvent avoir une propriété appelée tension et qui est l'analogue de la masse pour les particules.

Les cordes peuvent avoir différents modes de vibration. Tout à fait comme une corde de violon.



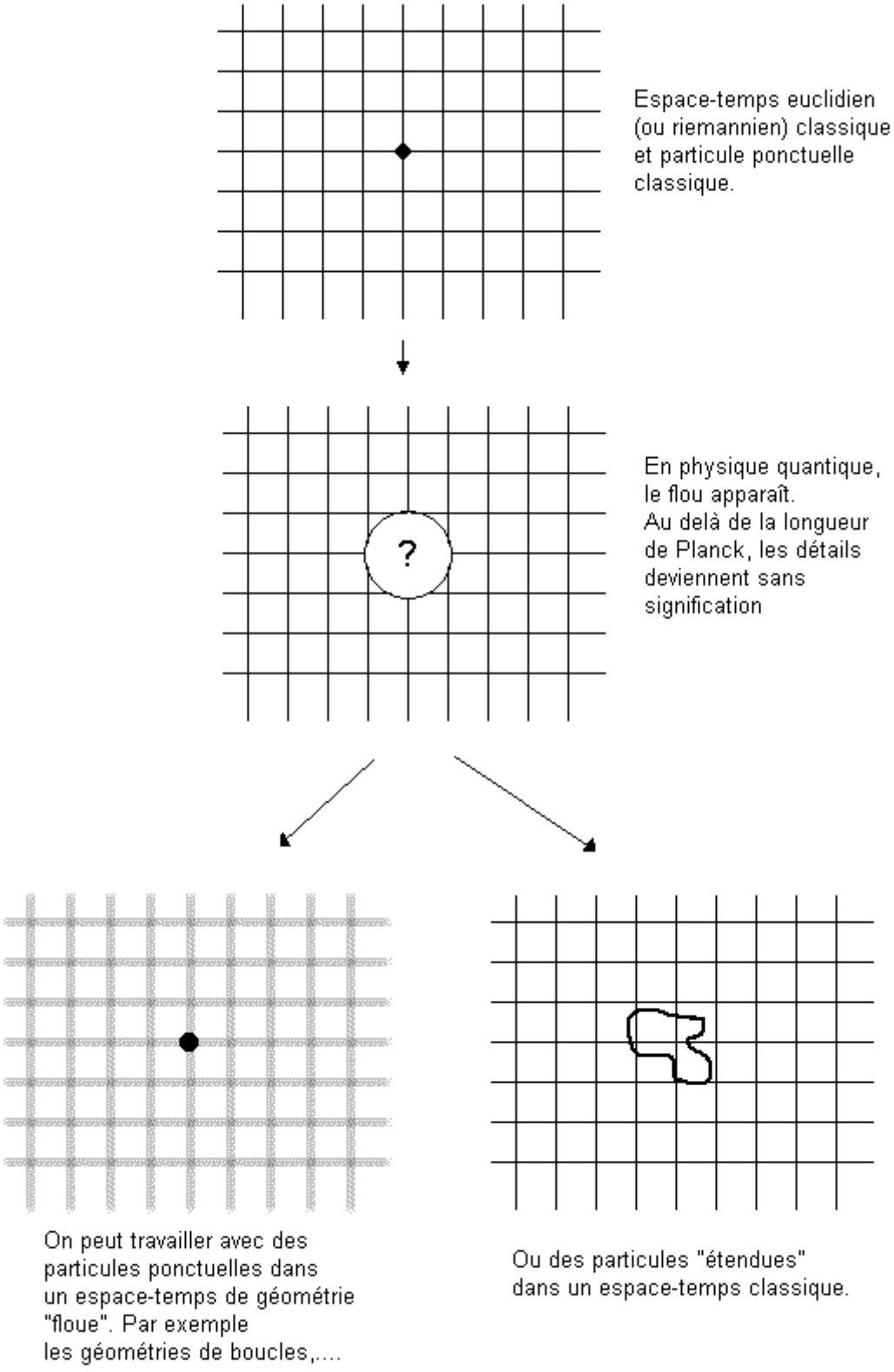
La corde peut avoir différents modes de vibrations avec une, deux ou plusieurs "bosses". Un peu comme une onde.

Ce qui se traduit par des ondulations de la surface d'univers.

L'idée majeure est alors d'associer à chaque mode de vibration un type de particule. Cela peut sembler à priori révolutionnaire mais après tout nous avons déjà fait ce genre d'analogie avec les champs (les modes de vibration des champs étant les particules après quantification) ! La seule différence est qu'ici on a un champ donnant les cordes (au lieu des particules) et les vibrations des cordes conduisent aux différents types de particules (au lieu d'avoir plusieurs champs).

Bien entendu, on imagine bien que la description mathématique des cordes est plus complexe que celle d'un simple point ! La théorie devient vite complexe au fur et à mesure des étapes (quantification, théorie des perturbations, etc.).

Il faut noter ici un lien avec l'approche par les géométries non commutatives. Celles-ci considèrent qu'à petite échelle la géométrie doit être différente pour tenir compte du fait qu'en dessous de l'échelle de Planck les détails ne sont pas "utiles". Mais une autre façon de tenir compte de cet aspect est de travailler avec une géométrie normale et de "disperser" les objets eux-mêmes !



Il semble même que la seule manière consistante de développer une gravitation quantique avec des objets étendus est celle des cordes.

5.3. Avantages et problèmes

Après avoir développé cette théorie, on peut en tirer une série d'avantages et d'inconvénients.

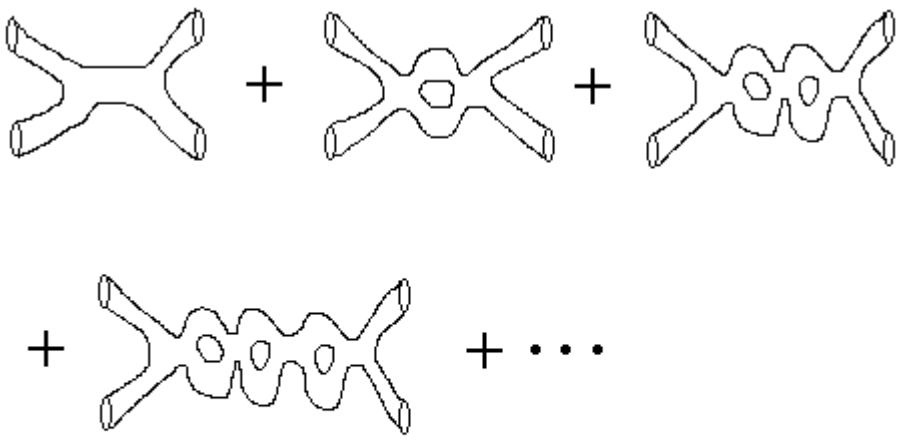
Avantages

Le premier avantage de cette théorie est évident. C'est la fin du zoo de particules. Au lieu d'avoir une pluralité de particules, on a plus qu'un seul type d'objet : une corde.

Un autre avantage est l'absence de renormalisation. En effet, une étude attentive des singularités ultraviolettes montre que leur apparition est due à la nature ponctuelle des particules. Cela n'est pas surprenant, une particule ponctuelle est une singularité dans l'espace-temps et toute singularité a tendance à provoquer ce genre de désagrément.

Les cordes ayant une extension spatiale, elles ne provoquent pas l'apparition de singularités ultraviolettes. C'est un avantage considérable. En effet, non seulement on peut se passer de renormalisation, mais en plus on ne risque pas de tomber sur une théorie non renormalisable !

Une autre facilité se révèle lors de l'application de la théorie des perturbations. Les règles de Feynman pour les cordes conduisent à une diagrammatique particulièrement simple.



Dans la théorie des cordes, les calculs montrent que les graphes à prendre en compte sont très simples. Puisqu'il s'agit de la répétition d'un motif simple (une boucle).

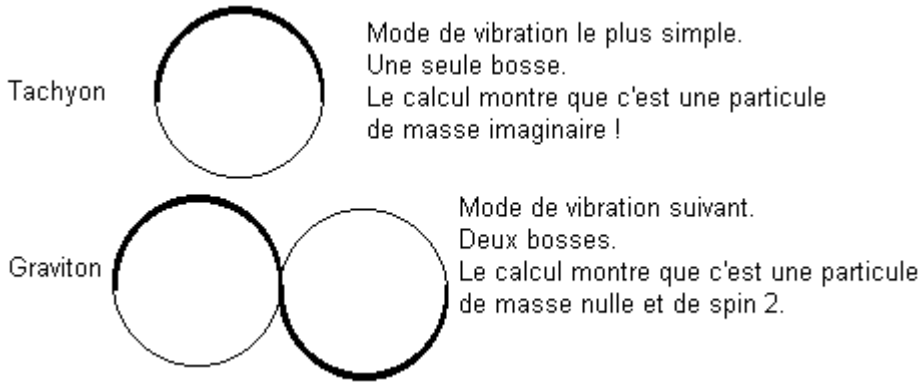
Malgré la complexité de la théorie, on a au moins une simplification , et pas des moindres. Au lieu d'avoir un nombre considérable de diagrammes lors de l'augmentation du nombre de boucles, avec les cordes le nombre de diagrammes reste faible.

La mise en équation de la théorie montre que des inconsistances peuvent exister sauf si l'on travaille dans un espace-temps avec 26 dimensions ! Donc 25 dimensions spatiales et une de temps. Selon la manière d'aborder la théorie ces inconsistances prennent différentes formes : violation de l'invariance de Lorentz (violation de la relativité) ou normes négatives des états. La norme des états est liée aux amplitudes que nous avons vues, qu'elles soient négatives implique des probabilités négatives, ce qui est gênant ! On montre que lorsque l'on sélectionne les états physiques parmi tous les états théoriques, les états à norme négative ne disparaissent tous que si l'on travaille à 26 dimensions.

Pour les théoriciens c'était très excitant. C'était la première fois que l'on trouvait une théorie imposant le nombre de dimensions pour des raisons de consistance. En effet, dans toutes les autres théories le nombre de dimensions était imposé. On pouvait faire varier le nombre de dimensions mais sans

rencontrer d'inconsistances. On pouvait parfois tomber sur des théories non renormalisables, mais celles-ci ne sont pas inconsistantes, c'est à dire ne possèdent pas de contradictions internes, elles sont seulement inutilisables en pratique à cause du nombre infini de paramètre libres. Ici, le nombre de dimensions est une conséquence de la théorie. Extraordinaire.

Il est également possible d'étudier les propriétés des particules. C'est à dire des différents modes de vibration des cordes. Le mode le plus simple s'appelle tachyon, nous allons y revenir. Le mode suivant est une particule de masse nulle et spin 2.



Mais une particule de masse nulle et de spin deux, c'est le graviton ! Fantastique ! Et si cela nous donnait une possibilité d'unification de la gravité avec les autres théories quantiques ? Là aussi pour la première fois, le graviton apparaissait comme une conséquence de la théorie sans que l'on doive l'introduire de manière "artificielle" (en choisissant les groupes de symétries par exemple).

Problèmes

Mais la théorie ne possède pas que des avantages. Ce serait trop beau.

Le premier ennui, grave, est le tachyon. Celui-ci correspond à une particule de masse imaginaire (ce n'est pas un nombre classique) ! De plus, il se déplace à une vitesse supérieure à celle de la lumière.

Le tachyon, avec une masse qui n'est pas un nombre classique, est un objet non physique (qui avait d'ailleurs déjà été imaginé). Que la théorie prédise son existence est évidemment assez grave.

Un autre ennui est le nombre de dimensions nécessaires à la théorie : 26. Ce nombre semble absurde. Malgré l'excitation d'avoir trouvé une condition imposant le nombre de dimensions, on aurait préféré trouver 4. Pourquoi ce nombre aussi élevé ?

Enfin, malgré sa diagrammatique simple et l'absence de renormalisation, la théorie reste complexe à manipuler mathématiquement. De plus, la théorie ne semblait pas pouvoir reproduire le monde réel. On ne retrouvait pas les particules existantes (le zoo). Par exemple, la théorie ne contient pas de fermions, donc pas de matière !

Notons enfin, que même si les divergences traditionnelles ne se produisent pas, d'autres divergences peuvent apparaître. En effet, les cordes sont des objets plus complexes que des particules ponctuelles. Une corde est constituée d'une infinité de points. Elle peut se tordre ou se courber d'une infinité de manière différente (les physiciens parlent de degrés de libertés et une corde à une infinité de degrés de liberté). Ces infinités de contorsions peuvent en retour provoquer l'apparition d'infinis dans les calculs. Mais ils peuvent être contournés. L'apparition de ces divergences est évidemment aggravée si l'on emploie des objets encore plus complexes comme des membranes, d'où la préférence pour des cordes. Mais comme on le verra, les membranes n'ont pas dit leur dernier mot !

A ce stade on est donc complètement bloqué. La théorie semble seulement être une curiosité mathématique intéressante. Dommage, mais l'histoire n'a peut-être pas dit son dernier mot.

En effet, l'expérience a montré que les divergences en théorie des champs n'étaient pas toujours faciles à surmonter. Par exemple, avec l'interaction faible, des divergences apparemment insurmontables se produisaient aussi (théorie de Fermi) jusqu'à l'arrivée de la théorie des champs de jauge de Yang et Mills. Il a été montré que c'était la seule manière consistante de surmonter ces divergences tout en conservant les propriétés de l'interaction faible. Par conséquent, le fait d'avoir trouvé une méthode consistante pour surmonter les divergences de la gravité doit être pris très au sérieux, d'autant que c'est la seule connue à ce jour (bien que la quantification canonique semble approcher aussi la solution, mais beaucoup d'aspects de ces théories se recoupent avec la théorie des cordes) !

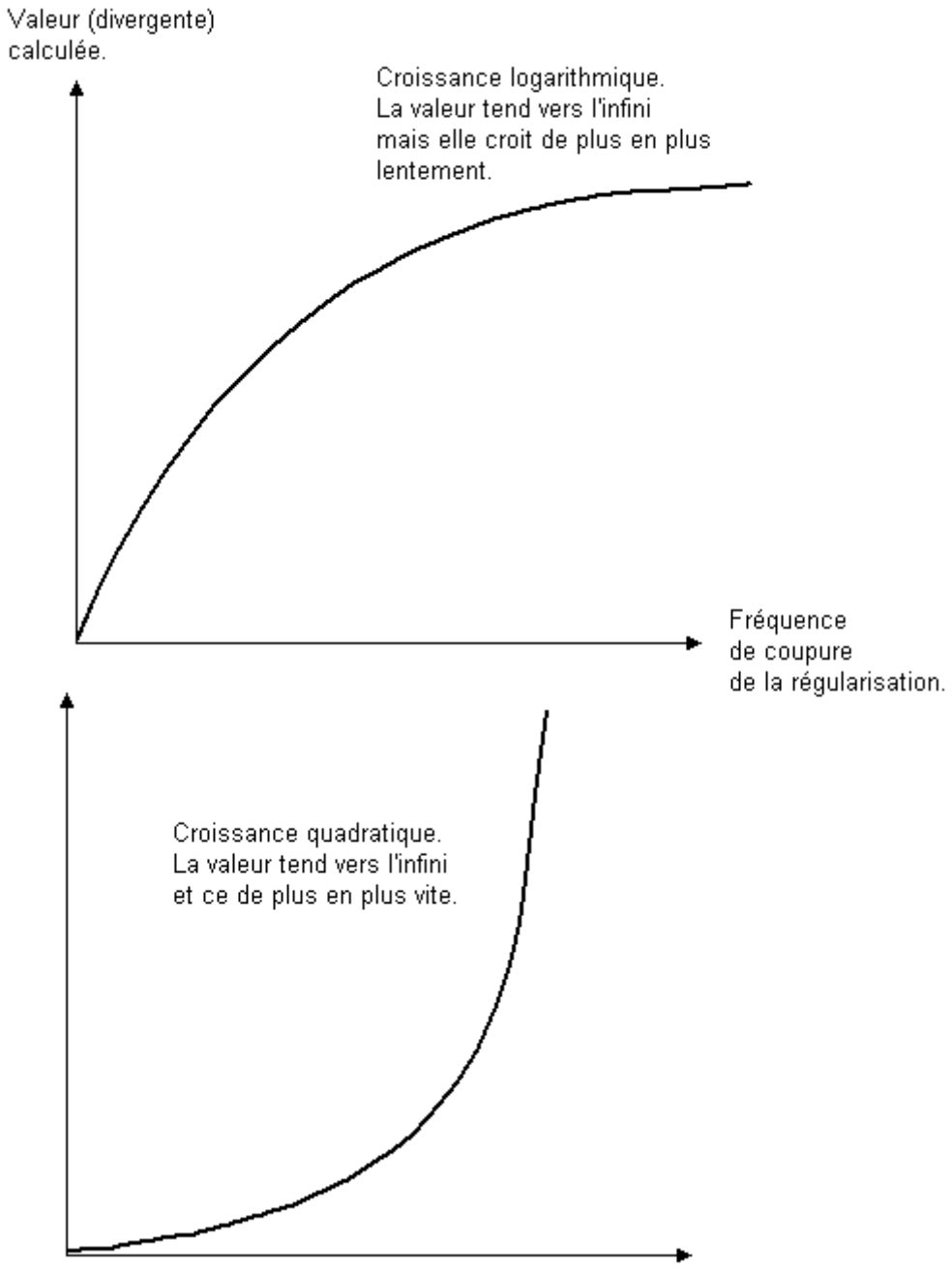
6. La première révolution

6.1. La super symétrie

L'idée de la super symétrie apparut au début des années septante mais prit véritablement son essor avec les travaux de Ferrara, van Nieuwenhuizen et Freedman, Deser et Zumino.

Le problème initial était l'existence dans la théorie des champs de divergences dites quadratiques.

Lorsque l'on regarde les divergences ultraviolettes, on constate qu'il en existe de plusieurs types. Les divergences logarithmiques sont les moins graves.

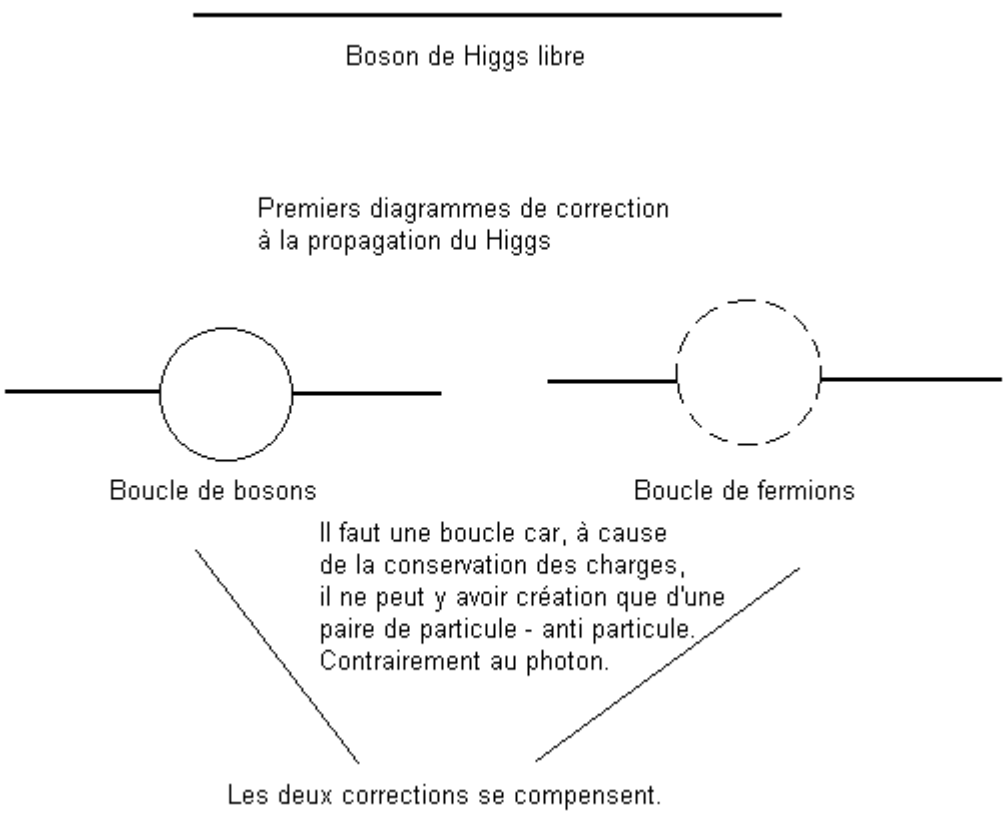


Une divergence logarithmique fait tendre la masse nue vers l'infini, comme toute divergence, mais si l'on considère que l'échelle maximale est celle de Planck, c'est à dire que l'on fait une régularisation en limitant l'énergie à celle de Planck, les corrections introduites aux paramètres sont seulement de l'ordre de quelques dizaines de pour cents. Par exemple la masse nue de l'électron est corrigée de 24 %.

Les divergences quadratiques, par contre, entraînent des corrections énormes. Ainsi, à l'échelle de Planck elles entraînent pour le boson de Higgs une correction d'un facteur de 30 ordres de grandeurs (un suivi de 30 zéros) ! Ce problème est d'ailleurs lié au problème de la hiérarchie.

On peut comprendre que les particules virtuelles entourant un électron introduisent une correction modeste (à condition d'effectuer la coupure à l'échelle de Planck). Par contre une correction de 30 ordres de grandeur semble inexplicable.

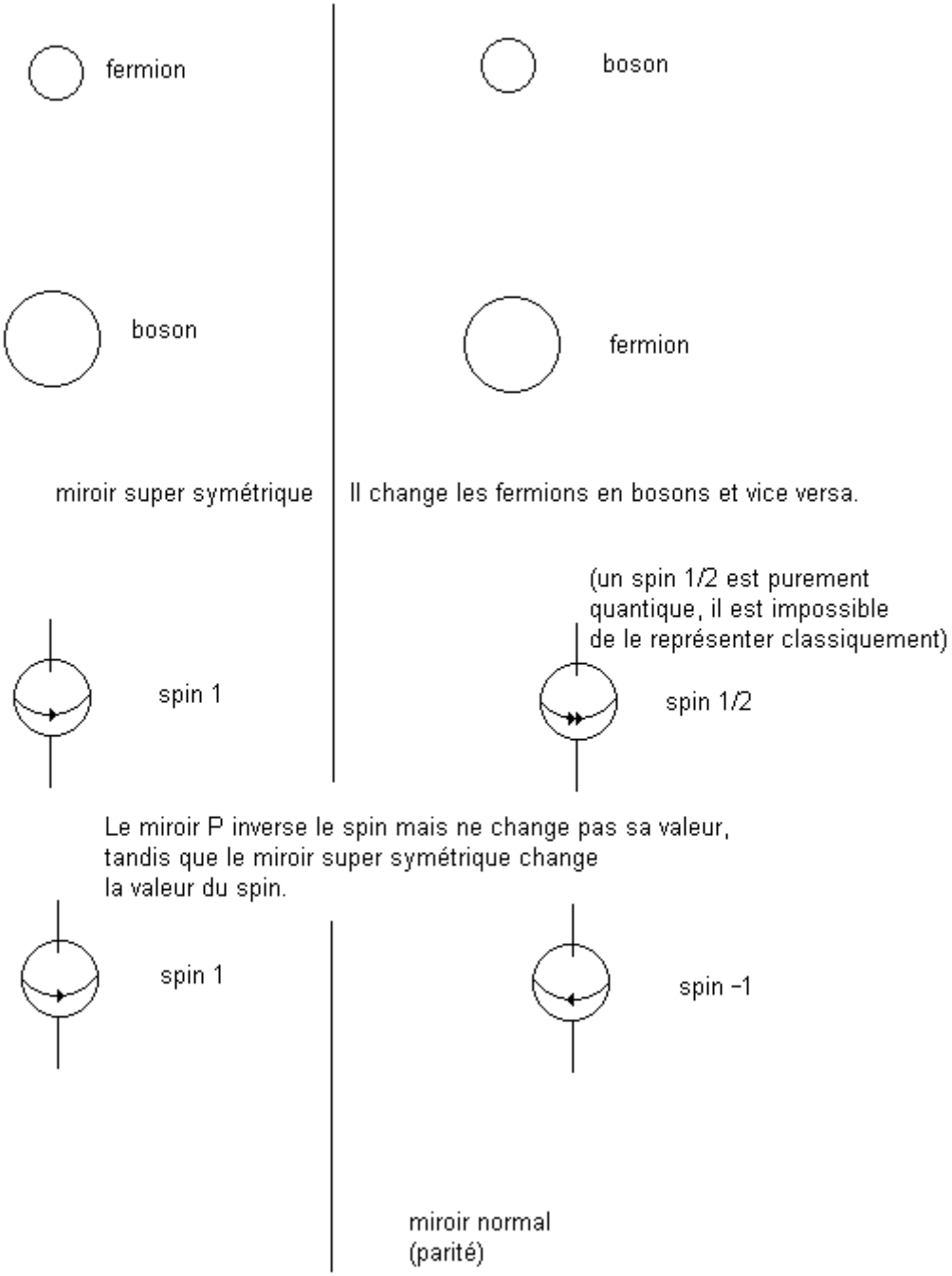
Il existe pourtant un moyen simple d'éliminer ces divergences quadratiques. Il suffit de s'arranger pour que dans les diagrammes de Feynman les boucles de fermions et de bosons se compensent.



Comme dans les diagrammes du boson de Higgs ci-dessus (qui présentent des divergences quadratiques).

Mais pour y arriver, il faut qu'il y ait une parfaite symétrie entre les fermions et les bosons. Sinon la contribution des diagrammes serait différente.

C'est le principe de la super symétrie. On introduit une symétrie qui fait correspondre les fermions et les bosons.



Tout comme une opération symétrique de parité échange un gant gauche et un gant droit, une opération avec la super symétrie échange les fermions et les bosons.

Le monde des particules est divisé en fermions et bosons. On peut se demander pourquoi et il est naturel d'imaginer une symétrie responsable de cette dualité.

Il existe d'autres bonnes raisons pour introduire la super symétrie. Mathématiquement, cette symétrie est la seule extension non triviale du groupe de Poincaré. Ceci est une conséquence d'une relation entre spin (lié aux fermions et aux bosons) et l'espace-temps. Relation qui avait déjà été constatée dès le début de la théorie des champs et qui est à la base du théorème spin - statistique. C'est un théorème de la physique quantique qui montre le lien entre le comportement statistique des particules et leur spin et qui est valable pour toute théorie quantique des champs. Ce théorème explique que les fermions ne peuvent tous se retrouver exactement dans le même état contrairement aux bosons. Le fait que les fermions ne puissent être tous dans le même état (principe d'exclusion de Pauli) explique

aussi pourquoi les électrons restent bien sagement sur des orbites séparées autour du noyau et empêchent ainsi toute la matière de s'effondrer !

Comme le groupe de Poincaré est nécessaire dans une quantification de la relativité générale, on voit l'intérêt de disposer d'une extension de ce groupe dans un cadre quantique.

De plus, un théorème du à Haag - Lopuszanski - Sohnius a établit que la plus grande symétrie qu'une théorie du champ en interaction puisse avoir est obtenue par un produit des groupes de jauge (éventuellement très large), l'invariance de Lorentz et la super symétrie. On a déjà les deux premiers, pourquoi pas le troisième ?

S'il le monde est super symétrique, alors chaque fermion doit avoir son boson correspondant avec la même masse, même charge électrique, etc.

Malheureusement, un coup d'œil aux particules du tableau standard, montre vite qu'il est impossible d'établir une telle correspondance entre les particules connues. Rien que le nombre de fermions est déjà beaucoup plus grand que celui des bosons.

Le monde ne serait-il pas super symétrique ? Qu'à cela ne tienne, ajoutons des particules ! A chaque particule connue, on associe donc un "super partenaire" de même masse, même charge, etc.

	PARTICLES	***Spin***	***SPARTICLES***	***spin***
BOSON	Photon	1	Photino	1/2
BOSON	Gluons	1	Gluinos	1/2
BOSON	W^+, W^-, Z^0	1	Winos, zino	1/2
BOSON	Graviton	2	Gravitino	3/2
BOSON	Higgs	1	Higgsino	1/2
FERMION	Quarks	1/2	Squarks	0
FERMION	Lepton : électron	1/2	Slepton : sélectron	0
FERMION	Lepton : muon	1/2	Slepton : smuon	0
FERMION	Lepton : tau	1/2	Slepton : stau	0
FERMION	Lepton : neutrinos	1/2	Slepton : sneutrinos	0

L'opération de super symétrie diminue donc le spin de un demi.

Le tableau contient déjà le gravitino anticipant sur la super gravité.

On montre que les super partenaires ne peuvent être produit qu'en paires. De même, un super partenaire ne peut se désintégrer qu'en un autre super partenaire (plus éventuellement des particules habituelles). Comme une particule ne peut se désintégrer qu'en une particule plus légère, cela implique que le plus léger des super partenaire doit être stable.

Mais si un des super partenaires est stable, cela veut dire que des particules de ce type restant de l'époque du Big Bang, lors de la création de l'univers, devraient "traîner" dans le coin. Or on n'en observe pas. Cela ne peut s'expliquer qu'à plusieurs conditions : ces particules doivent être rares, ce qui impose qu'elles soient assez lourdes (une particule massive est plus difficile à créer, puisqu'il faut beaucoup d'énergie, et donc plus rares), ces particules sont électriquement neutres (sinon leur détection serait facile), elles n'interagissent pas par interaction forte (sinon elles modifieraient les propriétés des noyaux atomiques).

Mais même pour les autres particules la question se pose. Pourquoi ne les observe-t-on pas dans les accélérateurs de particules ? La seule explication, si elles existent, est que leur masse est trop grande.

Mais nous avons dit que les particules et les super partenaires devaient avoir la même masse ! C'est vrai, mais si la super symétrie est une symétrie brisée (comme l'interaction faible), alors la masse de super partenaires peut devenir plus grande.

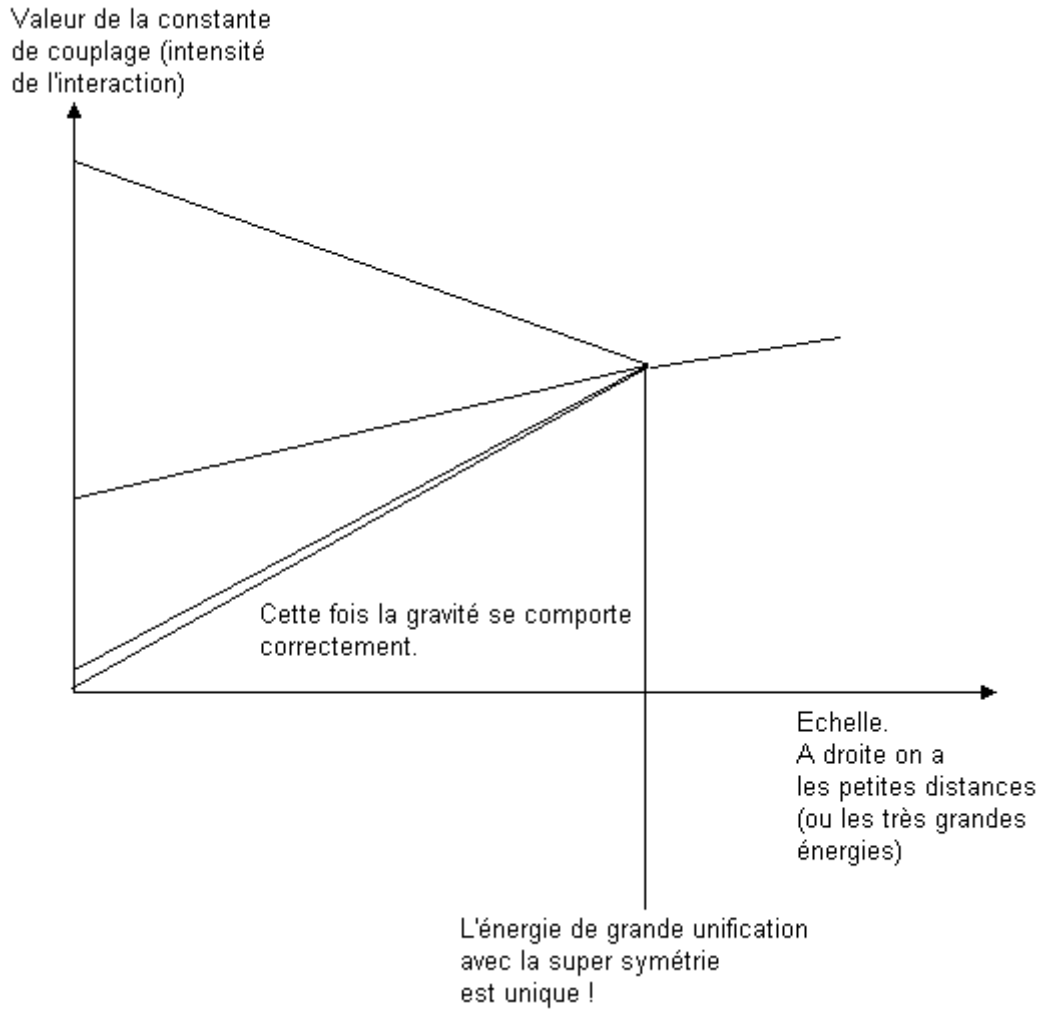
Des considérations théoriques montrent que la masse de ces particules ne peut pas être trop grande. Ce qui donne un espoir de les observer relativement vite avec la prochaine génération d'accélérateurs de particules… ou de jeter la théorie au bac si elles restent absentes ! Ou du moins, cela nous forcerait à retourner à nos planches à dessin pour comprendre l'absence des super partenaires.

La super symétrie améliore la renormalisabilité des théories par la suppression de certaines divergences, comme nous l'avons vu, ce qui est un atout.

Elle permet également de supprimer certains paramètres libres (pour la même raison) mais elle en introduit malheureusement de nouveaux. De plus, le zoo des particules a empiré puisque nous avons carrément doublé le nombre de particules !

Un des problèmes épineux est résolu : c'est la disparition du problème de hiérarchie. Comme on l'a dit, celui-ci est lié aux divergences quadratiques qui ont disparu. De plus, si la super symétrie est brisée à une échelle modeste, et non à l'échelle de Planck, cela explique l'ordre de grandeur des masses des particules.

Mieux encore, le problème de convergence des quatre interactions est résolu !



La gravitation "rentre dans le rang" alors que nous n'avons rien fait pour cela ! C'est une raison supplémentaire pour penser que la super symétrie est vraiment la voie.

A noter qu'il existe plusieurs versions de la super symétrie. On peut l'introduire une fois (comme ci-dessus) ou deux fois ou, enfin, quatre. Les super partenaires sont donc en un, deux ou quatre exemplaires.

Dans certaines versions, le proton peut même être instable. Mais la durée de vie du proton serait énorme (10^{70} ans). Même en observant un très grand nombre de protons l'espoir d'observer la désintégration de l'un d'eux est faible (un litre d'eau en contient environ un milliard de milliards de milliards = 10^{27}). Les tentatives en ce sens ont échoué jusqu'à maintenant.

6.2. Les dimensions supplémentaires

Si l'on essaie de construire une théorie complète avec les symétries de jauge $SU(3) \times SU(2) \times SU(1)$, l'invariance de Lorentz et la super symétrie on découvre une belle surprise. En effet, le commutateur des générateurs de la super symétrie (les opérateurs générant cette symétrie) est un opérateur de déplacements spatio-temporels. Cela conduit donc naturellement à la relativité générale à condition d'utiliser une invariance locale.

Cette fois aussi il est agréable de voir surgir la gravité sans l'avoir explicitement incluse. Pour que la théorie soit cohérente, il faut toutefois ajouter le graviton et le gravitino. Cette théorie a prit le nom de super gravité.

Pour que la théorie soit consistante, il est nécessaire que l'espace-temps aie 11 dimensions. 7 des dimensions étant enroulées sur une distance de l'ordre de la longueur de Planck. Chacune de ces dimensions supplémentaires correspond à une symétrie interne. Revoici Kaluza - Klein ! En fait, toute théorie de super gravité doit se dérouler dans un espace-temps de 11 dimensions.

On montra même que la super gravité, sous certaines conditions, autorisait un enroulement spontané de ces sept dimensions supplémentaires. Ce phénomène est analogue à une brisure de symétrie. La stabilité du vide nécessite cet enroulement. Ainsi, la théorie mène naturellement à l'espace-temps habituel à quatre dimensions et non plus de manière artificielle. C'est à dire que l'on n'impose pas cet enroulement mais que celui-ci découle de la théorie. C'était un grand succès.

Mais tout n'est pas rose. Ainsi la théorie présente toujours des paramètres libres. Pire encore, même si la super symétrie a amélioré la renormalisabilité de la théorie, la super gravité est encore non renormalisable. Ce qui est un sérieux écueil.

Et il faut encore trancher parmi toutes les théories possibles. Au début, on avait espéré aboutir avec une méthode analogue à Kaluza - Klein. Partir d'une pure gravité à 11 dimensions, plus la super symétrie, et en tirer toutes les théories. Mais cet espoir fut déçu, il était nécessaire de rajouter des champs de matière. Ainsi le nombre de théorie possible est énorme et il est difficile de trancher. Quel est la bonne théorie et pourquoi ?

- Différents champs de matières. Différents groupes de symétrie de jauge.
- Une, deux ou quatre super symétries ?
- Comment les 7 dimensions s'enroulent telles ? Sur une sphère, un tore, … ?

Le dernier problème, appelé problème de la compactification, est identique à celui de la théorie des cordes et nous y reviendrons.

6.3. La théorie des cordes super symétriques

Le moment est venu de mixer toutes les idées.

L'introduction de la super symétrie dans la théorie des cordes eut lieu à la charnière 1984 - 1985. Elle fut appelée la première révolution.

La théorie des cordes super symétriques, ou théorie des supers cordes, offre de nombreux avantages. En effet, la plupart des problèmes rencontrés disparaissent automatiquement !

Le premier avantage est la disparition des tachyons. Ceux-ci peuvent être reliés au principe d'incertitude. Nous avons vu que ce principe interdit l'existence d'un vide absolu. Il y a toujours des fluctuations. En fait, le principe d'incertitude interdit, pour un système quelconque, d'avoir une énergie minimale nulle. Il y a toujours une énergie résiduelle appelée énergie de "point zéro". Les tachyons, mode de vibration le plus bas, ne sont rien d'autre que cette énergie de point zéro. Dans une théorie

des cordes super symétriques, les modes de point zéro des fermions et des bosons se compensent et les tachyons disparaissent.

Le point zéro, c'est à dire le mode d'énergie le plus bas, devient le mode suivant qui est le graviton. Celui-ci est donc conservé par la théorie ce qui est tout à fait satisfaisant.

N'oublions pas que la théorie des cordes est renormalisable. Nous voilà donc avec une théorie de super gravité renormalisable. Quel succès.

Le deuxième avantage est une conséquence évidente de la super symétrie. Les fermions font leur apparition. Nous avons vu que la théorie classique ne contenait que des bosons. Mais la super symétrie reliant les bosons et les fermions, ces derniers surgissent donc naturellement. Il est donc peut-être possible de retrouver un spectre de particules tel que celui du modèle standard expliquant le zoo.

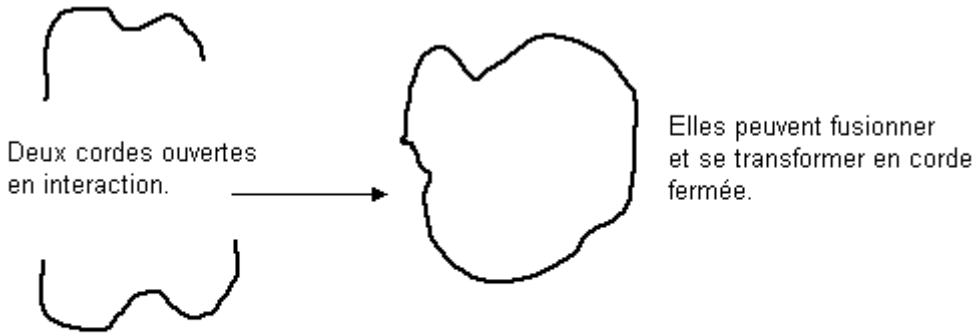
Enfin, les conditions de consistance de la théorie, c'est à dire le respect de l'invariance de Lorentz et la nécessité d'avoir des états de norme positive, conduit à une contrainte différente sur le nombre de dimensions. Celui-ci est maintenant de 10. C'est déjà beaucoup mieux que 26 ! Il n'y a plus que 6 dimensions à enrouler.

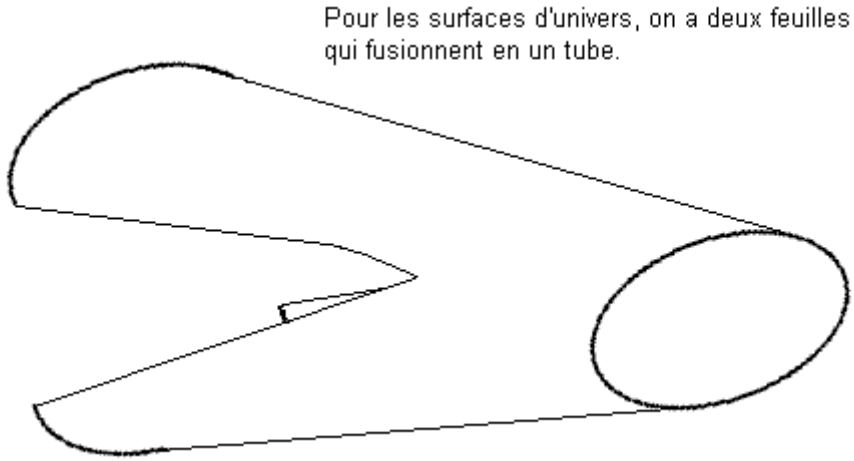
Le seul défaut est justement ce nombre de 10 dimensions. Nous avons vu qu'une super gravité nécessite d'avoir 11 dimensions. Il en manque une ! Mais ne nous formalisons pas pour si peu. On a déjà tellement progressé. Continuons pour voir si ce problème ne va pas se résoudre.

6.4. Plusieurs théories

La théorie des supers cordes permet également de construire plusieurs théories différentes que nous allons décrire. Il faut préciser que cela ne suit pas totalement la ligne historique. Tout s'est déroulé en seulement quelques années dans une explosion d'idées et de résultats mathématiques. En particulier, ce qui suit chevauche partiellement la seconde révolution qui sera présentée plus loin. Mais la courte durée de cette période rend impossible un suivit strictement historique sans nuire à la clarté de l'exposé.

On peut déjà séparer les théories en deux grandes classes selon que l'on a des cordes ouvertes ou fermées. En fait, la présence de cordes ouvertes implique automatiquement celle de cordes fermées car les cordes ouvertes peuvent interagir pour fusionner.





Les théories se divisent donc en :

- théories avec uniquement des cordes fermées
- théories avec des cordes ouvertes et des cordes fermées

La théorie des supers cordes présentait des anomalies dites gravitationnelles et des anomalies dites de Yang et Mills. Mais Green et Schwarz découvrirent que ces anomalies disparaissaient si le groupe de jauge utilisé était $SO(32)$ ou $E(8)\times E(8)$. Ce qui limite singulièrement le nombre de théories possibles.

Les théoriciens découvrirent ainsi une théorie consistante baptisée théorie de type I. Basée sur des cordes ouvertes.

Ils découvrirent également deux théories avec des cordes fermées, appelées théories de types IIA et IIB.

Gross, Harvey, Martinec et Rohm découvrirent les cordes hétérotiques hybrides (fermées) avec le groupe $E(8)\times E(8)$. Et ont découvrit également une théorie hétérotique avec le groupe $SO(32)$.

Les cordes fermées admettent des vibrations se déplaçant à gauche ou à droite.

Corde fermée animée de vibrations.

Les vibrations peuvent avancer dans un sens (un peu comme des vagues qui tourneraient dans un canal circulaire).

Ou dans l'autre sens.

Des vibrations peuvent se propager dans les deux sens à la fois. Un peu comme deux vagues peuvent se croiser sans se perturber (essayez en jetant deux petits cailloux en même temps dans l'eau).

Les vibrations peuvent se déplacer dans les deux sens indépendamment. La super symétrie peut donc être attribuée à seulement un sens. D'où le nom d'hybride puisqu'un sens est super symétrique tandis que l'autre sens est classique.

Il semble en l'état actuel des choses que ce soient les seules théories possibles mais ce n'est pas démontré.

Les théories peuvent donc se regrouper dans un tableau.

Nom	***Type de cordes***	***Graviton***	***Symétrie (de jauge)***	***Super symétries***	***Parité***
Type I	Ouvertes	Oui	$SO(32)$	1	Violée
Type IIA	Fermées	Oui	Non	2	Conservée
Type IIB	Fermées	Oui	Non	2	Violée
Hétérotique SO(32)	Fermées	Oui	$SO(32)$	1	Violée
Hétérotique E(8)xE(8)	Fermées	Oui	$E(8)\times E(8)$	1	Violée

Décrivons maintenant ces théories plus en détail.

Théorie de type I

La théorie de type I contient le graviton ainsi que des bosons de jauge correspondant aux bosons intermédiaires (de l'interaction faible), le photon et les gluons. Ainsi que les super partenaires correspondants.

Cette théorie contient aussi des fermions qui peuvent interagir avec les bosons de jauge. Ils portent ainsi des charges.

De plus, la théorie viole la parité tout comme l'interaction faible. On dit que la théorie est chirale.

Tout cela est encourageant et permet d'espérer retrouver toutes les interactions habituelles. Malheureusement le groupe $SO(32)$ est beaucoup trop large. Il inclut bien les interactions habituelles, dérivant du groupe $SU(3) \times SU(2) \times SU(1)$, mais il implique aussi l'existence d'interactions supplémentaires qui ne sont évidemment pas observées.

Il y a aussi d'autres problèmes. Ainsi la théorie est à 10 dimensions. Et rien ne semble induire automatiquement un enroulement des six dimensions supplémentaires. Pire encore, les particules sont toutes sans masse !

Il semble donc que cette théorie ne puisse pas réellement correspondre au monde réel. Toutefois, à très haute énergie les particules se comportent comme des particules sans masse. Cette théorie pourrait donc, éventuellement, être une candidate pour une théorie limite s'appliquant à haute énergie.

Théories de type II

Ces théories possèdent des gravitons et des fermions. Malheureusement la situation est encore pire qu'avec la théorie de type I. En effet, ces théories ne portent pas de groupe de jauge et ne possèdent donc pas de bosons de jauge ! Donc, elle ne prédit pas le photon et les autres bosons porteur des interactions. C'est la situation inverse de la théorie des cordes classiques, cette fois nous avons de la matière mais pas d'interaction.

Comme les fermions ne peuvent pas réagir entre eux (via les bosons de jauge), ils ne portent pas de charge.

Il y a bien un champ bosonique tensoriel, analogue au photon mais décrit par un tenseur au lieu d'un vecteur, mais il n'interagit pas avec les fermions qui ne peuvent donc pas porter de charge vis à vis de ces bosons. Ce champ ne véhicule donc pas une interaction.

Ces théories semblent donc, à priori, très éloignées du monde réel. Elles furent pendant longtemps considérées comme exotiques.

Théories hétérotiques

Avec les théories hétérotiques un problème d'un tout autre genre se pose. Comme la super symétrie est imposée à un sens des vibrations, alors la consistance impose que les cordes soient plongées dans un espace-temps à 10 dimensions. Mais l'autre sens ne possède pas la super symétrie. Les cordes doivent donc être dans un espace-temps à 26 dimensions comme pour les cordes classiques ! Cela pose évidemment un gros problème.

Heureusement, le mécanisme de Kaluza - Klein vient à notre secours. Il suffit de prendre un espace-temps à 26 dimensions puis d'enrouler complètement 16 des dimensions supplémentaires pour retrouver un espace-temps à 10 dimensions.

Cet enroulement à même d'autres avantages. Il introduit des degrés de libertés supplémentaires qui se traduisent par l'apparition de bosons de jauge qui étaient normalement absents dans les théories avec des cordes fermées (comme les théories II). Ces degrés de libertés suivent des symétries internes selon le groupe $SO(32)$ ou $E(8) \times E(8)$.

Le fait d'avoir la super symétrie pour un seul sens est ennuyant car l'équilibre entre les fermions et les bosons (l'équilibre avec les super partenaires) n'est plus restauré. Il manque des fermions. Or cet

équilibre est indispensable à la renormalisabilité de la théorie. A nouveau, l'enroulement des 16 dimensions nous aide car il restaure l'équilibre entre les fermions et les bosons.

Les théories suivant le groupe $SO(32)$ ont un phénomène déplaisant. Lorsque l'on enroule spontanément les 6 dimensions pour revenir à l'espace-temps habituel à quatre dimensions, l'invariance sous la parité est restaurée ! Elles ne peuvent donc décrire correctement l'interaction faible.

Par contre, les théories avec le groupe $E(8) \times E(8)$ conservent cette propriété cruciale (ainsi que la théorie IIB d'ailleurs).

Les théories avec le groupe $E(8)$ sont, de plus, beaucoup plus intéressantes vis à vis des interactions et des particules connues. Car ce groupe est beaucoup moins large. Les théoriciens travaillant dans le domaine de la théorie des champs de jauge avaient déjà tenté d'unifier les théories en une théorie obéissant à un groupe unique $E(6)$ (nous en avions touché un mot en disant que le groupe $SU(3) \times SU(2) \times SU(1)$ pouvait dériver d'une symétrie brisée plus large). Et la théorie des cordes hétérotiques lui ressemble beaucoup (après être passé à quatre dimensions).

La théorie des cordes hétérotiques $E(8) \times E(8)$ prédit ainsi des familles chirales de quarks et de leptons comme dans la théorie des champs classique.

De plus, Candelas, Horowitz, Strominger et Witten découvrirent qu'il existait un mécanisme spontané d'enroulement des six dimensions (sur une variété de Calabi - Yau, nous y reviendrons) comme dans la super gravité.

Cette théorie est donc la meilleure candidate à une théorie unifiée.

Le dilaton

Nous avons vu que l'intensité des interactions était une fonction de l'énergie. Cette variation de l'intensité est une conséquence de la renormalisation et est décrite par le groupe de renormalisation. Comme les théories des cordes sont libres de divergences, on pourrait s'attendre à ce que la constante de couplage des interactions soit une constante. En réalité c'est un peu plus compliqué que cela.

Nous avons vu que la théorie de Kaluza - Klein prévoyait l'existence d'un champ scalaire sans masse appelé dilaton. Ce champ est une conséquence de l'enroulement des dimensions supplémentaires et toutes les théories des cordes prédisent donc de tels dilatons. Ce champ scalaire n'est pas directement observable mais dicte l'intensité des autres interactions (c'est un boson de Goldstone, donc inobservable).

Il donne donc un espoir de déduire l'intensité des interactions de la théorie plutôt que de devoir introduire les constantes de couplages à partir de l'observation.

Techniquement, la constante de couplage est fonction d'une certaine valeur constante que prend le champ scalaire des dilatons dans le vide quantique. Mais nous ne savons pas comment déterminer cette valeur ! La super symétrie permet à cette constante de prendre n'importe quelle valeur. Elle agit donc comme un paramètre libre.

Mais nous savons aussi que la super symétrie est une symétrie brisée. De cette manière la valeur de cette constante du dilaton dans le vide peut-être déterminée. C'est un des buts principaux que les théoriciens espèrent atteindre dans le futur.

Les champs tensoriels de jauge

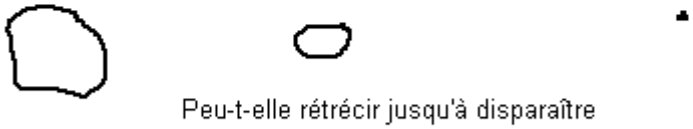
Nous avons vu que les théories de type II prévoyaient l'existence d'un champ de jauge tensoriel. En réalité c'est le cas de toutes les théories des cordes. Quel est la signification physique de ce champ ?

Nous avons dit que les fermions n'interagissaient pas avec ce champ et ne portaient donc pas de charge associée à ce champ. Toutefois on peut se poser la question de savoir s'il existe une certaine sorte d'objet dynamique qui pourrait interagir avec ce champ et donc porter une charge.

La question est importante car l'existence d'une telle charge implique une loi de conservation de telle manière que les objets correspondants peuvent être stables. En particulier, le plus léger de ces objets doit être stable.

Quel est l'objet dynamique associé à ce champ ? La réponse est incroyablement simple. Ce sont les cordes elles-mêmes ! Elles portent une charge unité associée qui est conservée par les interactions. L'existence de ce champ implique donc la stabilité des cordes. C'est du moins le cas pour les cordes fermées. Pour les cordes ouvertes la situation est un peu plus complexe et ne sera pas abordée ici.

Au cour du temps une corde fluctue,
se déforme.



Peu-t-elle rétrécir jusqu'à disparaître complètement ?

NON. Car la corde porte une charge qui assure sa stabilité. Cette charge est conservée et la corde ne peut pas disparaître.

Ce n'est pas la charge électrique, bien sûr, mais une charge associée au champ tensoriel de jauge.

Ce champ n'est donc pas non pas directement observable mais est un mécanisme physique sous jacent (comme les Goldstone) essentiel à la théorie.

Conclusions

Quelles conclusions pouvons-nous tirer de tout cela ?

Nous avons cinq théories au lieu d'une seule. C'est très ennuyant car nous aurions espéré n'en trouver qu'une seule. Quelle est la bonne et pourquoi ?

Même si la théorie des cordes hétérotiques $E(8) \times E(8)$ est la meilleure candidate, il reste à comprendre pourquoi la nature a choisi celle là.

Mais même sans chercher une telle explication, la situation n'est toujours pas parfaite. Car même si cette théorie semble la meilleure, elle ne correspond toujours pas au monde réel ! Nous avons dit qu'elle ressemblait à certaines théories des champs. Mais ressembler n'est pas être identique. Et il existe encore des écarts qui ne s'expliquent pas à ce stade.

7. La seconde révolution

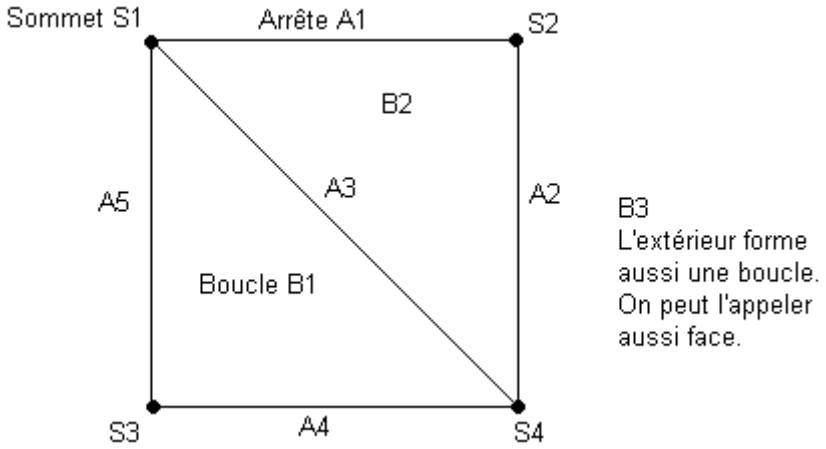
7.1. Les dualités

Introduction

Qu'est que les dualités ?

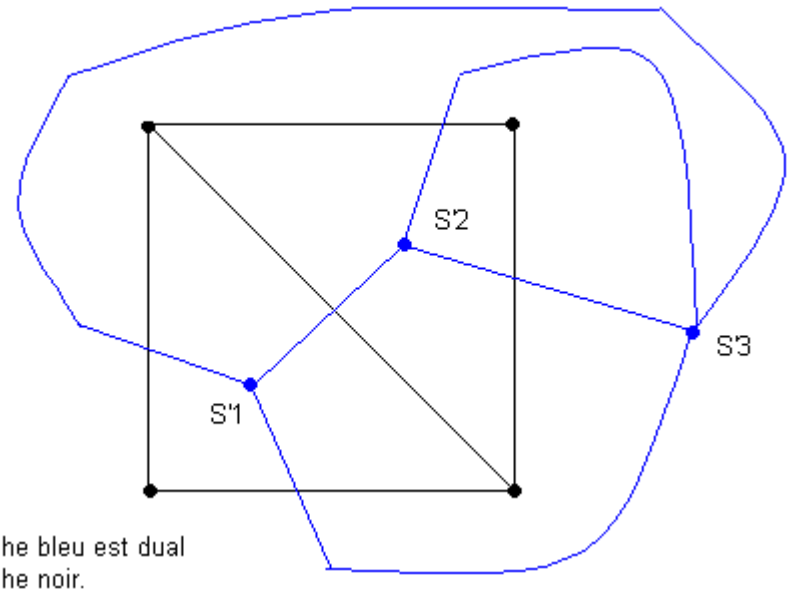
Ont dit que deux théories sont duales lorsque les deux théories décrivent le même système physique. C'est à dire lorsque deux descriptions mathématiques différentes décrivent formellement la même physique. Illustrons cela par quelques exemples.

Prenons un système physique pouvant être décrit par un graphe.

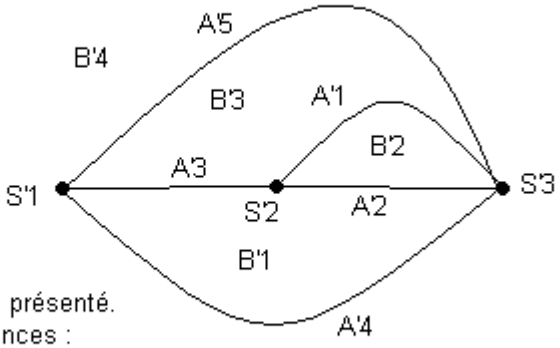


A chaque sommet est associé une valeur, et à chaque boucle est associée une valeur. Les arrêtes du graphe décrivent des relations entre les sommets et les boucles.

A ce graphe il est possible d'associer un autre graphe appelé graphe dual.



Le graphe bleu est dual
du graphe noir.

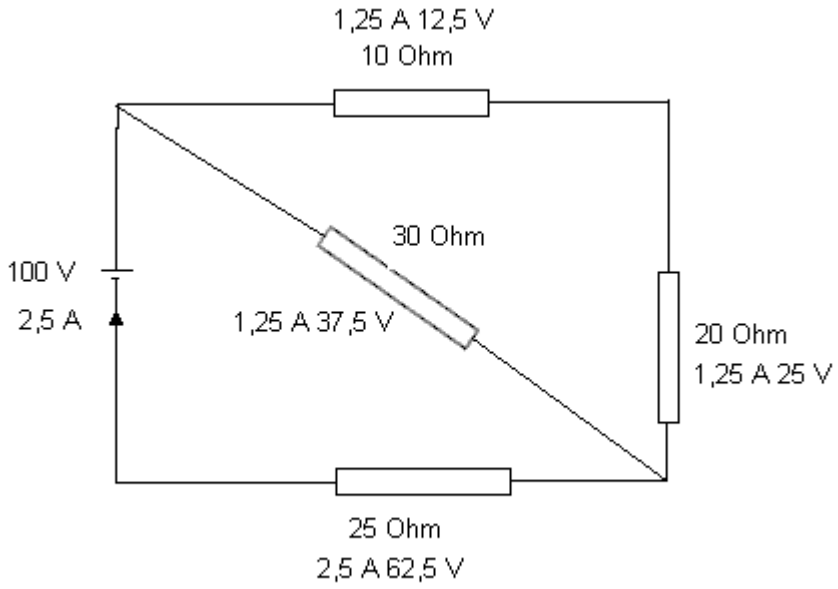


Le graphe dual mieux présenté.
On a les correspondances :
A1 - A'1 A2 - A'2 A3 - A'3 A4 - A'4 A5 - A'5
S1 - B'3 S2 - B'2 S3 - B'4 S4 - B'1
B1 - S'1 B2 - S'2 B3 - S'3

Comme on le voit, à chaque sommet du graphe initial correspond une boucle dans le deuxième graphe et à chaque boucle correspond un sommet. Ainsi le graphe initial à quatre sommets et trois boucles et le deuxième graphe a trois sommets et quatre boucles.

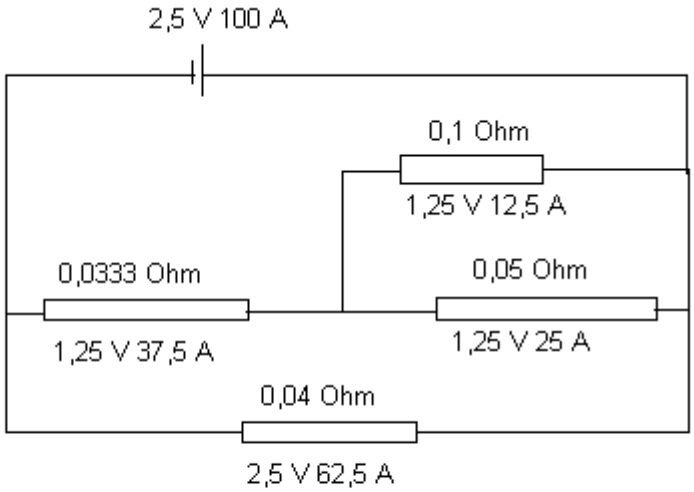
Les relations décrites par les arêtes du deuxième graphe sont les mêmes que pour le premier graphe, sauf que ce qui s'applique au sommet de l'un s'applique aux boucles de l'autre et vice-versa. Bien que différents ces deux graphes correspondent au même système physique.

Une manière simple de voir cela est l'utilisation en électricité.



Les lois sur l'électricité impliquent que la somme des courants qui arrivent à un sommet doit être égale à zéro (le courant qui arrive doit repartir) et la somme des différences de potentiel le long d'une boucle est égal à zéro.

Si je change ce circuit par son graphe dual.



Graphe électrique dual.
On a échangé les potentiels et les courants.
Et on a inversé les résistances.

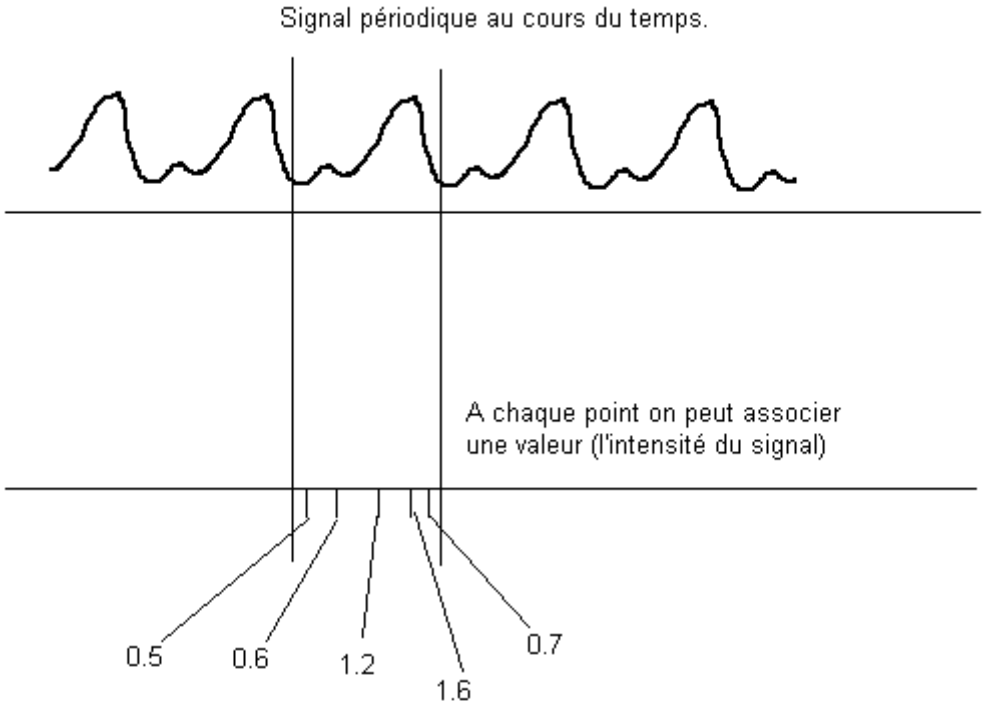
On peut vérifier que les règles sont respectées (somme nulle des courants à un sommet, somme nulle des potentiels pour une boucle, et loi d'Ohm).

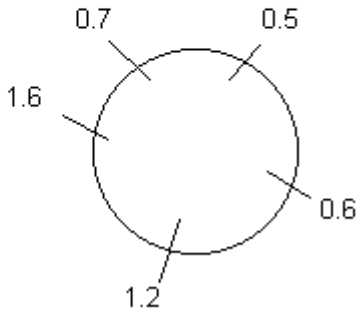
En électricité il s'agit d'une dualité parallèle <-> série.
Ces deux circuits sont analogues.

Alors, les sommets sont échangés avec les boucles et les courants sont échangés avec les potentiels, mais les règles restent les mêmes. Formellement les deux graphes décrivent les mêmes circuits mais où l'on a échangé le rôle du courant et du potentiel et passé d'un montage parallèle / série à un montage série / parallèle.

Donc deux théories duales décrivent la même physique mais vue d'un autre point de vue.

Un exemple de dualité plus "mathématique" est celui du cercle. Supposons un système qui peut être décrit par une valeur sur une droite.





Mais on peut aussi disposer les valeurs sur un cercle puisque le signal est périodique.
Mathématiquement, le cercle est dual d'une droite périodique.

Si la valeur est périodique, c'est à dire si elle se répète à l'identique. Alors, la description sur un cercle est formellement équivalente.

Un autre cas de dualité en physique est l'électricité et le magnétisme. Les charges électriques existent mais par contre il n'existe pas de charge magnétique. Nous avons vu aussi que l'électricité et le magnétisme jouaient un rôle semblable dans les équations de Maxwell de l'électromagnétisme.

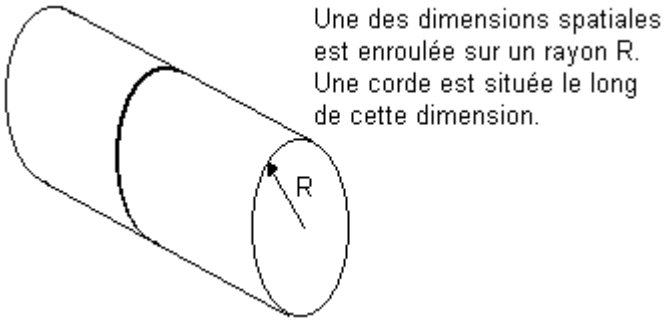
En fait, en l'absence de charge électrique, on peut échanger l'électricité et le magnétisme dans les équations sans changer le système physique qui est décrit !

Quel est l'intérêt de travailler sur ces dualités ? Tout simplement parce que les calculs sont parfois plus faciles à effectuer dans une des formulations ! Les dualités peuvent aussi révéler des aspects physiques cachés qui avaient échappé à une première analyse.

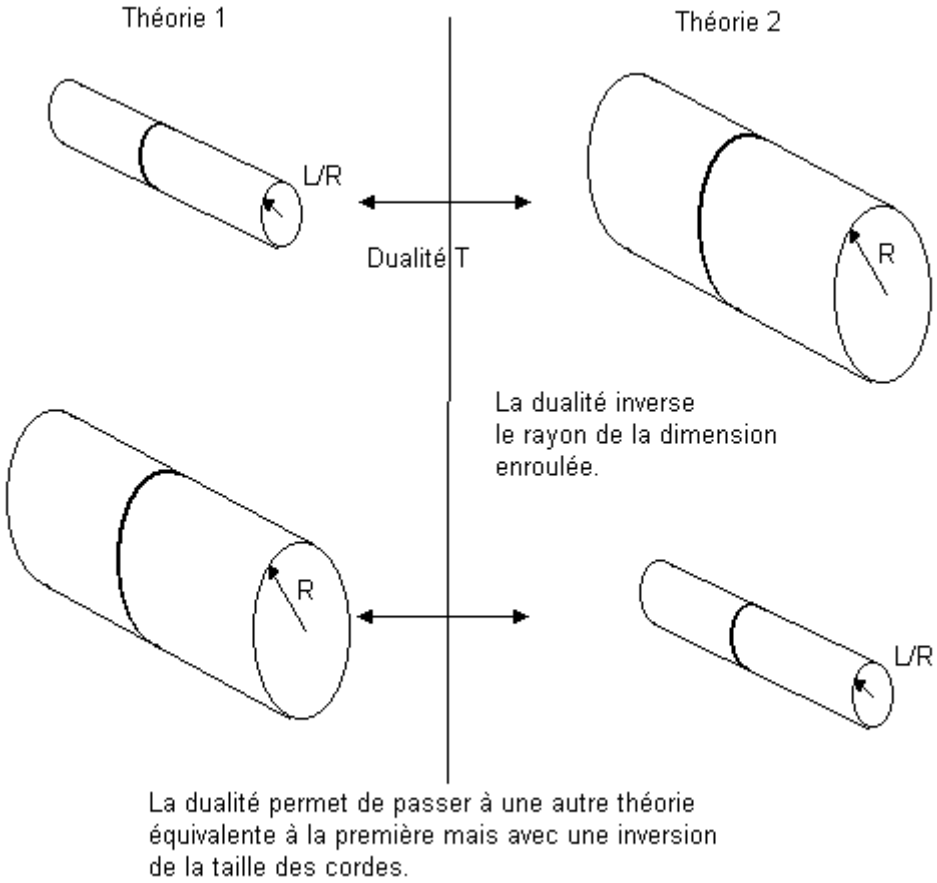
Dans les théories des cordes, des dualités semblables existent. Notamment des dualités de type électricité et magnétisme. Mais des dualités plus profondes encore ont été trouvées.

La dualité T

Cette dualité relie deux théories des cordes qui diffèrent par le compactage des cordes sur une des dimensions.

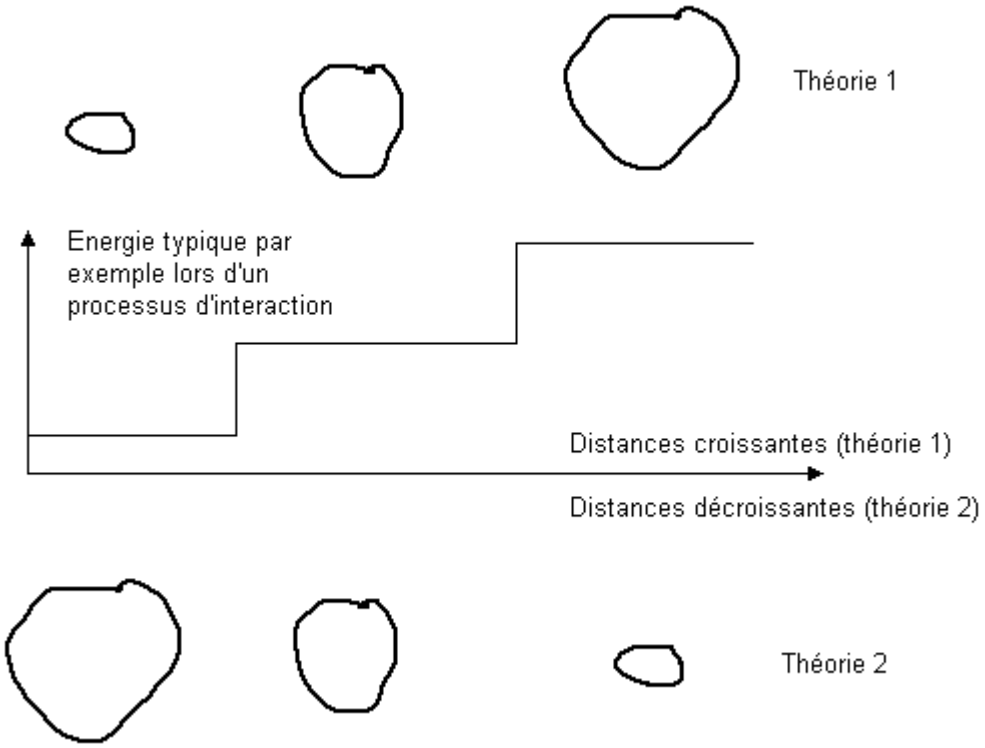


L'une des théories a ses cordes enroulées sur une dimension de rayon R, tandis que l'autre a ses cordes enroulées sur une dimension inverse L/R.



Donc quand une théorie décrit des cordes enroulées sur une petite dimension, l'autre décrit des cordes enroulées sur une grande dimension. La valeur de la constante L est reliée à la tension de la corde.

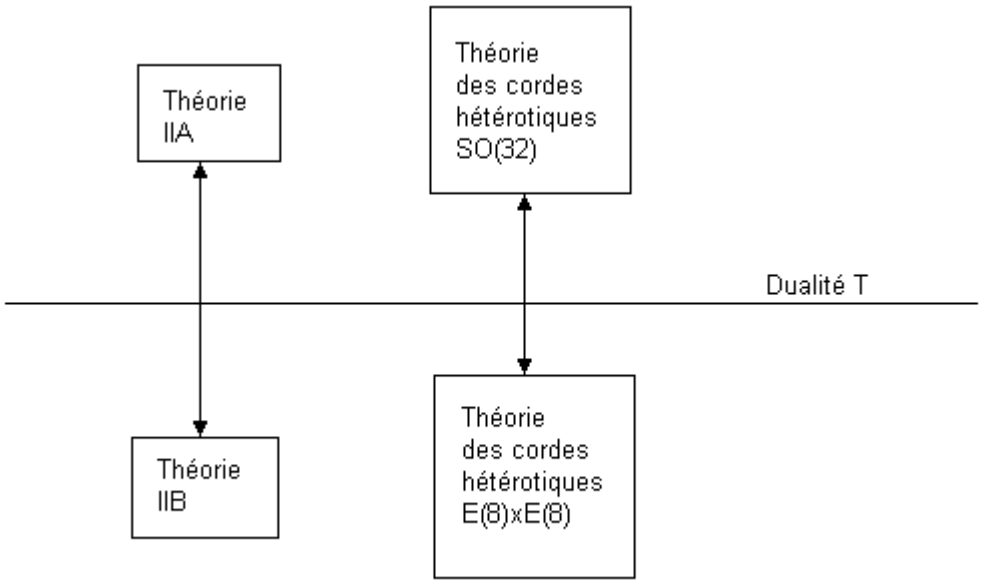
Dans une théorie l'énergie d'une corde est proportionnelle au rayon tandis que dans l'autre l'énergie est inversement proportionnelle au rayon. Ainsi les cordes duales ont même énergie.



Donc ces deux théories décrivent formellement la même physique. Leur description mathématique est différente puisqu'une théorie avec une dimension fortement enroulée correspond à l'autre théorie avec une dimension faiblement enroulée. Mais les conséquences physiques qui peuvent découler des équations sont les mêmes.

On découvrit ainsi que les deux théories des cordes IIA et IIB étaient duales l'une de l'autre via la dualité T.

On découvrit de même que les deux théories des cordes hétérotiques étaient duales l'une de l'autre ! Cela peut sembler étonnant au vu de leurs groupes de jauge différents ($SO(32)$ et $E(8)\times E(8)$) mais la manière d'enrouler une des dimensions peut tout changer et en particulier les symétries.



Ainsi, il s'avère que, en définitive, les deux théories IIA et IIB décrivent la même physique. De même les deux théories des cordes hétérotiques décrivent la même physique.

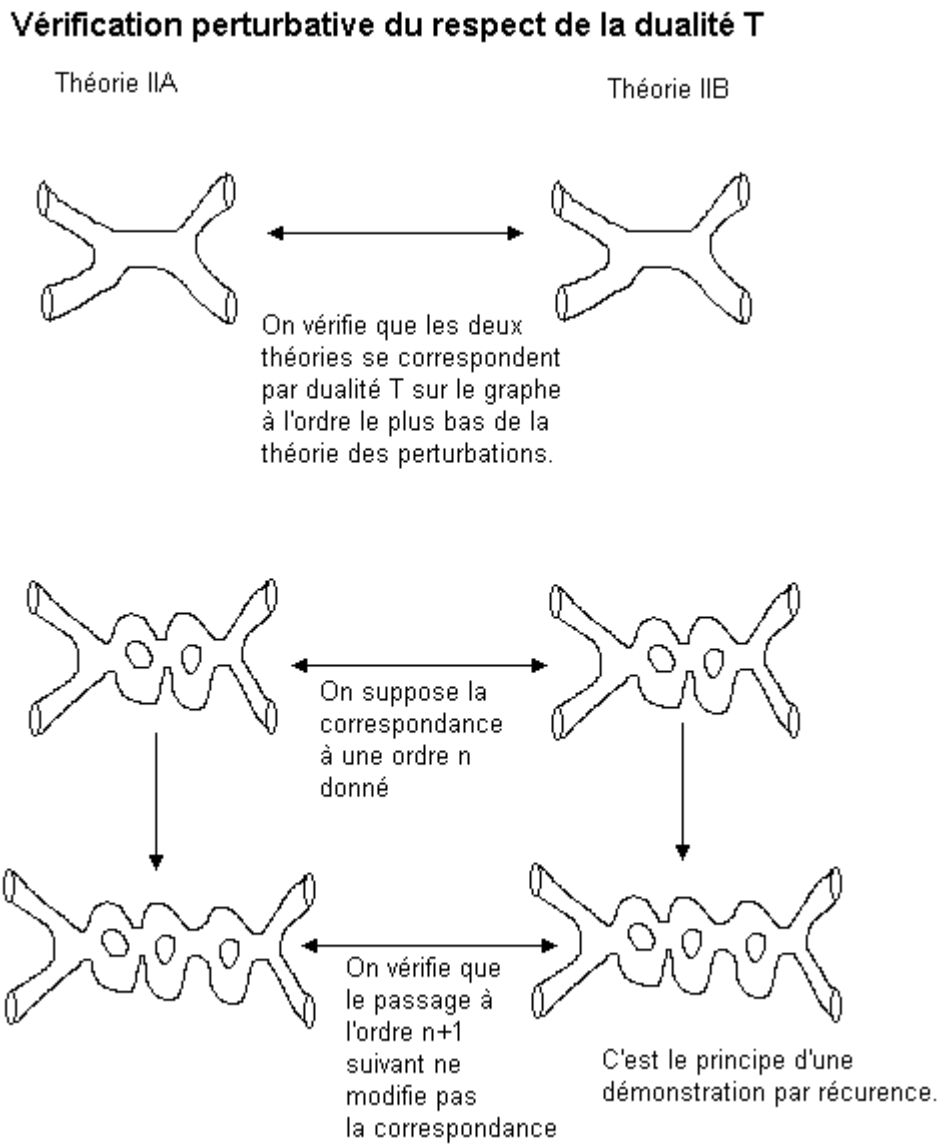
C'est évidemment très satisfaisant de découvrir cela pour deux raisons :

- Les calculs effectués dans une théorie peuvent parfois être plus faciles lorsque l'on effectue ces calculs dans l'autre théorie. Il suffit alors de passer de l'une à l'autre par l'opération de dualité T.
- Au lieu d'avoir quatre théories des cordes, on en a plus que deux. Il est donc plus simple de chercher la "bonne" théorie !

Mais cela montre aussi que ces théories ne sont pas si simple puisque l'on avait pas remarqué qu'elles décrivaient la même physique !

Comment a-t-on démontré la dualité T entre ces théories ? En utilisant la théorie des perturbations. On a vu que la théorie pouvait s'écrire de manière perturbative et sous forme diagrammatique.

Les calculs sur un graphe sont plus faciles que sur la théorie dans son ensemble. Nous avons d'ailleurs vu, avec la théorie des champs, que la théorie n'était en général pas soluble de manière exacte et qu'il fallait utiliser des méthodes approchées. Démontrer une telle dualité sans passer par la théorie des perturbations est parfois très difficile.

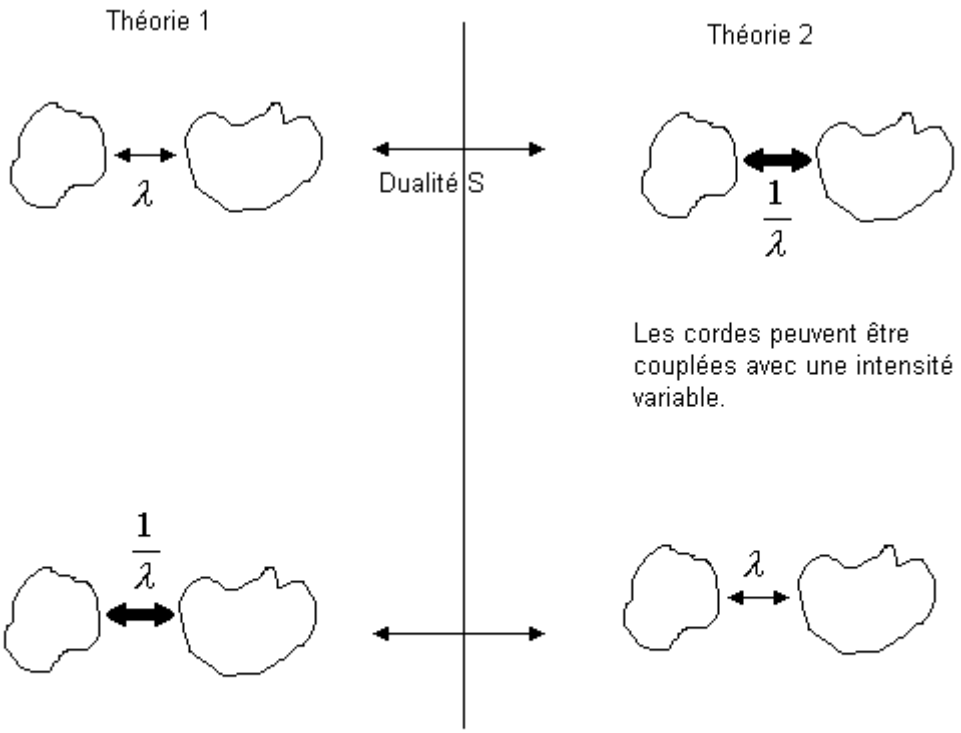


Le fait de pouvoir effectuer la démonstration via la théorie des perturbations est donc satisfaisant. On démontre que la dualité T s'applique ordre par ordre sur chaque graphe de la théorie des perturbations. Bien entendu, on n'effectue pas la démonstration pour tous les graphes (ils sont en nombre infini) ! Mais on le démontre pour le graphe le plus simple (appelé diagramme en arbre car il ne contient pas de boucle) puis on démontre que le résultat est valable pour tous les graphes plus compliqués. En fait, on démontre que si la dualité est valable pour un graphe d'ordre n, alors elle est également valable pour un graphe d'ordre $n+1$.

La dualité S

Les cordes interagissent entre-elles. Nous avons vu qu'il existait une constante de couplage qui dicte l'intensité de ces interactions. Nous avons même vu que cette intensité était reliée au champ de dilaton.

La seconde dualité qui fut découverte est la dualité S. Elle relie deux théories qui ont des constantes de couplages inverses.



La dualité S permet de passer d'une théorie 1 avec des cordes fortement couplée à une théorie 2 avec des cordes faiblement couplées mais décrivant formellement la même physique.

Lorsqu'une théorie décrit des cordes fortement couplées, l'autre théorie décrit des cordes faiblement couplées et vice versa.

A nouveau, certaines caractéristiques doivent varier en sens inverse avec la constante de couplage selon la théorie. Ainsi l'énergie d'interaction (ou les distances entre cordes) doit varier dans le même sens que la constante de couplage dans une des deux théories et dans l'autre sens pour l'autre théorie. Cela rappelle l'interaction forte où la constante de couplage diminue avec l'énergie.

On découvrit que la théorie IIB est duale d'elle-même. C'est à dire que lorsque l'on change la constante de couplage par son inverse, on obtient la même théorie ! Celle-ci possède un comportement identique à courte et grande distance. C'est assez surprenant.

Par contre la théorie des cordes hétérotiques $SO(32)$ est duale de la théorie I par la dualité S. Donc la théorie des cordes hétérotiques $SO(32)$ décrit un comportement à grande distance identique à la théorie I à courte distance et vice versa.

La correspondance duale entre ces deux théories est d'autant plus étonnante que la théorie I possède des cordes ouvertes alors que les théories hétérotiques n'en ont pas ! Ces deux théories décrivent pourtant la même physique mais dans des contextes différents (courtes et grandes distances).

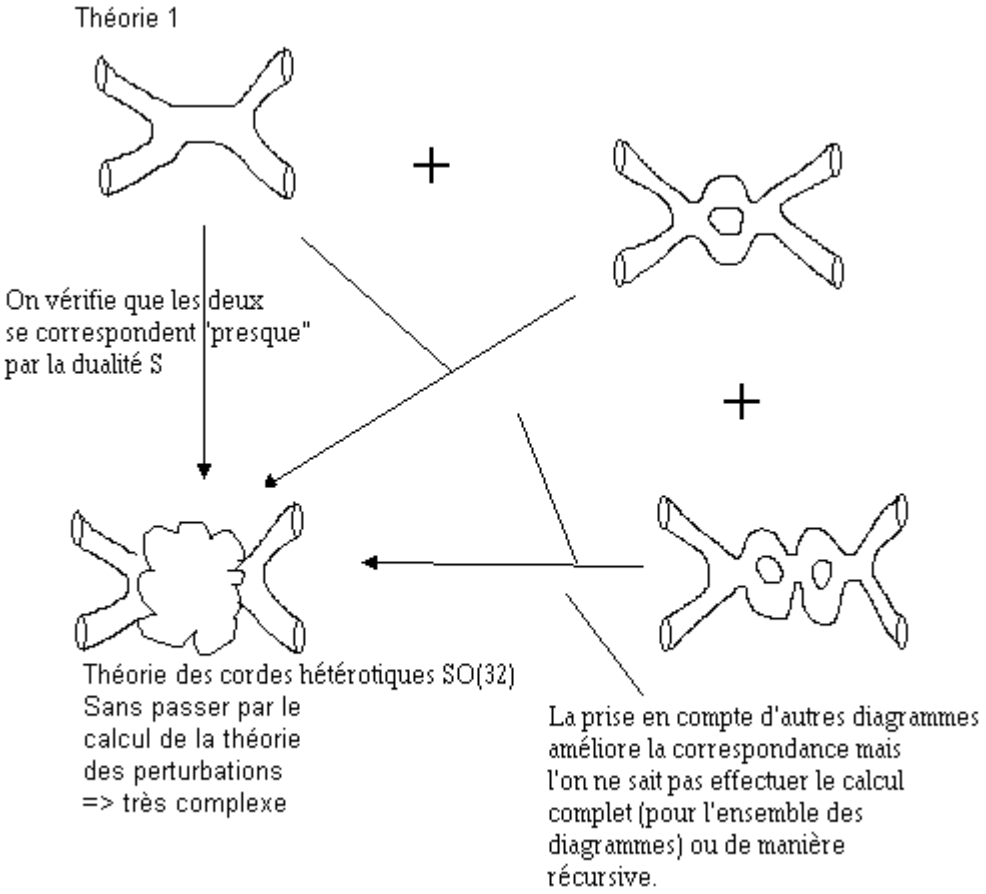
Il est nécessaire de dire que la correspondance de ces théories par la dualité S n'est pas totalement prouvée. Où est la difficulté ?

Nous avons vu que la théorie des perturbations ne pouvait pas s'appliquer pour des grandes constantes de couplage car les termes successifs (qui sont des corrections) sont de plus en plus grand. Les résultats ne convergent pas et la théorie des perturbations est inapplicable. Dans un tel cas on a besoin de méthodes différentes par exemple des méthodes de calculs exacts (non perturbatifs).

Or ici la dualité S fait correspondre une théorie avec un faible couplage avec une autre théorie avec un fort couplage. Quel que soit le couplage, il y a toujours une des deux théories où l'on ne peut utiliser la théorie des perturbations !

Par conséquent, il est impossible de démontrer la dualité S en utilisant la théorie des perturbations. On a besoin d'une démonstration qui se fait sur la totalité de la théorie, de manière exacte, non perturbative. Bien entendu, une telle démonstration est beaucoup plus difficile.

On peut toutefois effectuer une démonstration "approchée". C'est à dire montrer qu'il y a presque correspondance et probablement correspondance exacte.



On effectue la correspondance pour des interactions avec un diagramme en arbre pour l'une des deux théories et on vérifie que par dualité S, on obtient "presque" l'autre théorie. Pourquoi presque ? D'une part on a comparé une théorie "tronquée" (uniquement les diagrammes en arbre) avec une théorie complète, d'autre part la vérification sur l'autre théorie, complète, est très difficile puisque l'on n'y applique pas la théorie des perturbations.

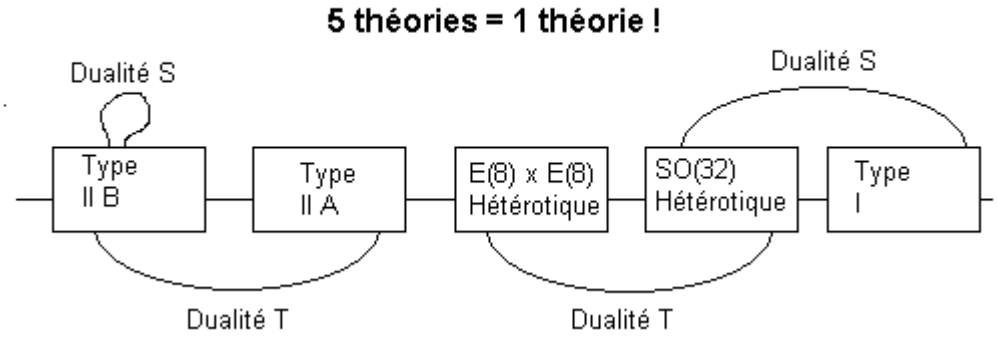
Puis on regarde le diagramme suivant et on vérifie que la dualité est améliorée. Et ainsi de suite. L'idéal serait de montrer qu'en prenant en compte tous les diagrammes la dualité s'améliore sans cesse jusqu'à devenir parfaite mais l'on n'y est pas encore arrivé. C'est un des grands challenge pour le futur.

En plus de ce type de raisonnement, les théoriciens ont découvert un vaste faisceau d'indices qui sont en faveur d'une dualité exacte entre les deux théories. Même en l'absence d'une démonstration rigoureuse (et en attendant celle-ci) les théoriciens conjecturent que la dualité est exacte.

Une telle dualité est très pratique dans l'étude de problèmes concrets. Supposons que je souhaite étudier une réaction particulière entre les cordes où la constante de couplage de l'interaction est grande. Je constate donc que je ne peux pas appliquer la théorie des perturbations et le manque d'outils mathématique m'empêche d'effectuer un calcul exact. Aucun problème ! J'applique la dualité S et je me retrouve avec une situation avec une faible constante de couplage. Je peux donc, cette fois, appliquer la théorie des perturbations. J'effectue les calculs puis j'applique à nouveau la dualité S pour revenir à la description initiale et le tour est joué.

La seconde révolution

La découverte des ces dualités en 1994 constitua ce qui fut appelé la seconde révolution.



Nous verrons qu'il y a également moyen de regrouper également les deux groupes de théories.

Donc toutes ces théories (en anticipant un peu) n'en forment en réalité qu'une seule ! Ce qui est évidemment très satisfaisant puisque l'on a plus à chercher laquelle s'applique à la réalité. Il suffit de le prouver pour l'une et les autres suivront automatiquement.

Quelles sont donc les différences entre-elles ? Formellement, elles décrivent la même physique mais leur domaine d'application est très différent. Les différences sont, par exemple :

- La définition du vide quantique
- L'énergie des interactions
- L'intensité du couplage entre les cordes
- La façon de replier les dimensions (nous y reviendrons)
- Les échelles de distances
- Les conditions aux limites. Par exemple, dans la théorie des collisions on a vu qu'on partait et qu'on arrivait à une situation où les particules sont libres. On appelle ces deux situations des conditions aux limites. Pour une même description de réactions entre cordes, les conditions aux limites de deux théories des cordes peuvent correspondre à des situations physiques très différentes.
- Les D-branes (voir plus loin).
- L'approche du problème : perturbatif ou non perturbatif.

7.2. La théorie M

Nous avons donc besoin d'une description mathématique unique qui unifierait les descriptions de toutes les théories des cordes. Quelle pourrait être la forme d'une telle théorie ?

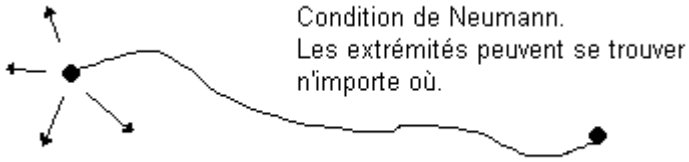
Cette théorie hypothétique est appelée théorie M (M pour matrices, membranes, mystère).

Les D-branes

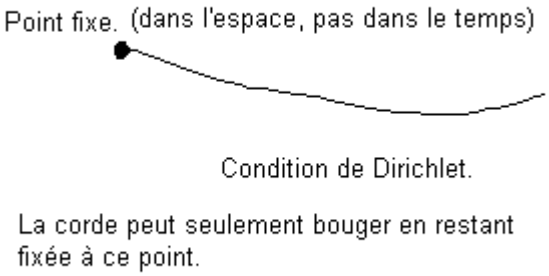
Avant d'approfondir le sujet, il est nécessaire d'étudier un peu mieux les différentes propriétés des cordes. Et en particulier les conditions aux limites des extrémités des cordes ouvertes.

Pourquoi des conditions aux limites pour les extrémités ? Tout simplement parce qu'une corde ouverte à deux bouts et qu'il faut bien décrire comment ils se comportent !

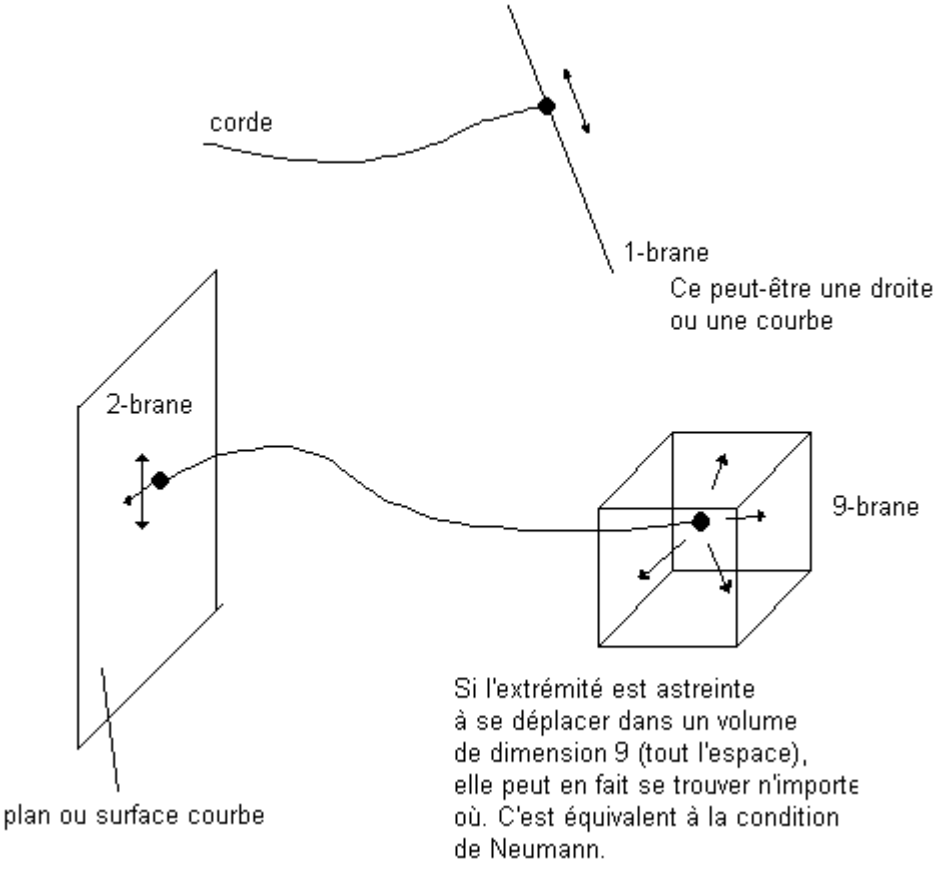
Quelles conditions peut-on fixer aux extrémités ? Les conditions les plus générales et les plus naturelles sont les conditions aux limites de Neumann. Les extrémités sont astreintes à se déplacer… partout ! C'est à dire dans tout l'espace. Elles sont libres.



Une autre idée est de fixer ces cordes en un point particulier de l'espace (donc en fixant 10 des coordonnées la 11ème, le temps, restant évidemment libre).



Enfin, on peut imaginer des conditions mixtes. C'est à dire imposer les coordonnées dans D dimensions et les laisser libre dans 9 - D dimensions spatiales.



Mathématiquement, les conditions de Dirichlet et de Neumann ont une formulation mathématique précise. On impose donc D conditions de Dirichlet et 9 - D conditions de Neumann. Une telle condition aux limites est appelée une D-brane (D pour dimensions ou pour Dirichlet). Une 0-brane est un point, une 1-brane est une courbe, une 2-brane est une surface (une membrane), etc. Une 9-brane est l'espace tout entier et correspond aux conditions de Neumann montrées plus haut.

Attention, une 1-brane n'est pas une corde ! C'est une condition imposée aux extrémités d'une corde mais elle ne se décrit pas, mathématiquement, du tout de la même manière.

Les conditions de Neumann respectent l'invariance de Lorentz mais pas les conditions de Dirichlet puisque ces dernières imposent des coordonnées particulières. Or en relativité, il n'y a pas de repère (donc pas de point ni de direction) privilégié.

Pour ces raisons on a longtemps utilisé uniquement les conditions de Neumann quand, en développant la théorie, on s'est rendu compte que les conditions de Dirichlet apparaissaient naturellement et étaient indispensables !

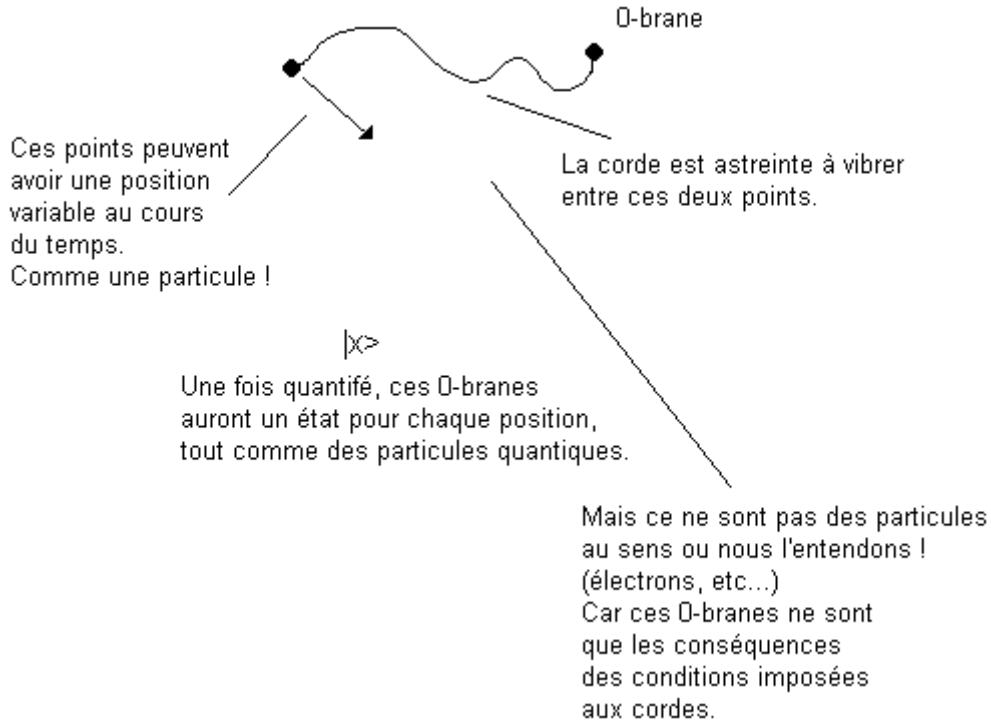
Mais les conditions de Dirichlet violent l'invariance de Lorentz ! Comment s'en sortir ?

En réalité cette violation n'est pas surprenante. Prenons la relativité classique avec des particules. Tant qu'il n'y a aucune particule (le vide) l'invariance de Lorentz est respectée. Maintenant plaçons une particule quelque part. L'invariance n'est plus respectée puisqu'il existe un repère privilégié : celui où la particule est au repos ! C'est normal. La théorie sous-jacente est invariante mais un état particulier ne l'est pas. On doit donc considérer une D-brane comme un état particulier du système.

Donc l'ensemble des D-branes possibles pour une corde forme un ensemble d'états possibles. Etat peut-être pris au sens de la mécanique quantique puisque la théorie des cordes est une théorie quantique. Les D-branes sont donc des objets "dynamiques" décrits par un ensemble d'états. Le "vide

de D-brane" est un état invariant par transformation de Lorentz mais un état particulier contenant des D-branes ne l'est pas.

Par exemple, une 0-brane, étant un point, est décrite par un ensemble d'états, chacun d'eux correspondant à une position donnée. Donc, une 0-brane se comporte comme une particule (dans un espace-temps à 10 dimensions) ! Mais attention, de telles D-branes ne sont pas des objets isolés puisqu'elles sont reliées par les cordes. Les D-branes ne sont que la conséquence des propriétés assignées aux extrémités des cordes.



En particulier une théorie avec des 2-branes est différente d'une "théorie des membranes". C'est à dire une théorie semblable à la théorie des cordes mais qui décrirait des membranes vibrantes au lieu de cordes vibrantes. Ce genre de théorie a déjà été étudié et, bien qu'intéressante, retombe dans les travers de la non renormalisabilité à cause du trop grand nombre de degrés de libertés des membranes.

Les D-branes ne changent rien, formellement, aux cordes (au contraire puisqu'elles fixent des contraintes sur les extrémités des cordes) et la théorie reste renormalisable.

On constate, en étudiant la théorie, que les D-branes peuvent avoir des interactions avec certains champs de jauge et peuvent donc porter une charge dont la conservation assure la stabilité de ces D-branes. Tout comme les cordes ont des interactions avec le champ tensoriel de jauge comme nous avons vu.

Dans ce cas la théorie n'est plus seulement une théorie des cordes mais une théorie des cordes et des D-branes ! Le spectre des particules (bosons, fermions) à un lien étroit non seulement avec les cordes mais aussi avec les D-branes, ce qui n'est pas surprenant puisque la façon dont sont fixées les extrémités joue forcément un rôle pour les vibrations autorisées pour la corde. Tout comme le fait de fixer les cordes d'un violon permet à celles-ci de vibrer selon certains modes précis de vibrations. Ce qui donne des possibilités supplémentaires et… de grandes complications théoriques !

Pour qu'une 2-brane puisse interagir avec un champ de jauge, il faut que celui-ci soit un tenseur du troisième rang. Nous avions présenté les tenseurs mais c'était uniquement des tenseurs du deuxième rang (qui se transforment comme deux vecteurs). Un tenseur du troisième rang est plus complexe. Un

tenseur du troisième rang est au tenseur du second rang ce qu'un tenseur du second rang est à un vecteur (appelé parfois tenseur de premier rang). La raison mathématique à cela est qu'une surface (comme une 2-brane) décrit une trajectoire au cours du temps qui est un volume, c'est à dire une variété à trois dimensions. Donc en général, une D-brane a besoin un champ de jauge tensoriel de rang D+1.

Les D-branes stables dépendent donc des champs tensoriels de jauge prédits par la théorie.

Par exemple, la théorie IIA contient un champ tensoriel de troisième rang qui interagit avec des 2-branes. Elles sont donc stables.

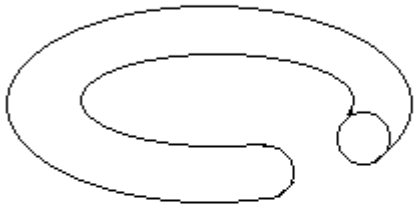
Notons aussi que les D-branes peuvent être définies même en présence de cordes fermées (comme la théorie IIA). On peut en effet décrire les D-branes pour des cordes ouvertes (dans la théorie I) puis appliquer les dualités pour passer à des théories avec des cordes fermées. Cela fournit déjà un cadre plus global pour décrire l'ensemble des théories (mais ce n'est pas encore suffisant, ce n'est toujours pas la théorie M recherchée).

La dualité U

On a également découvert une théorie, utilisant des membranes (en faisant un détour par la théorie des cordes avec des D-branes), qui se déroule dans un espace à 11 dimensions et qui est équivalente aux autres théories !

En enroulant les membranes sur un cercle, on repasse à 10 dimensions et les membranes enroulées forment des cordes donnant la théorie IIA.

On l'enroule sur une dimension.
(appelons là X)

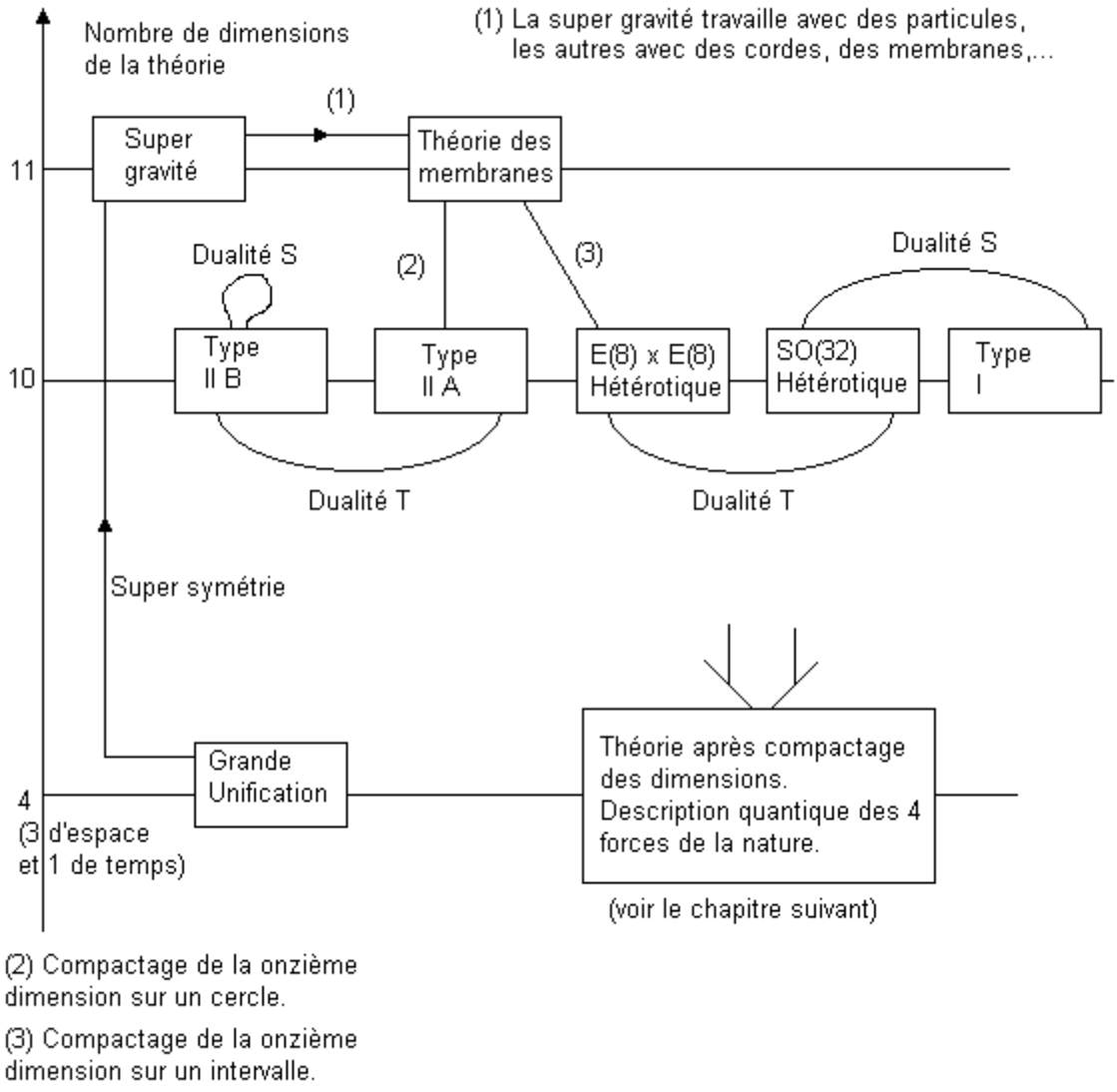


On l'enroule sur une autre dimension
(celle sur laquelle habituellement
les cordes s'enroulent)

On resserre encore la dimension X.
On passe ainsi dans un espace-temps
à 10 dimensions.
HOP -> on a une corde !

En enroulant les membranes sur un segment de droite, on retombe sur la théorie des cordes hétérotiques $E(8) \times E(8)$. La correspondance entre ces théories s'appelle la dualité U.

Nous avons donc, comme annoncé, une relation reliant toutes les théories entre elles.

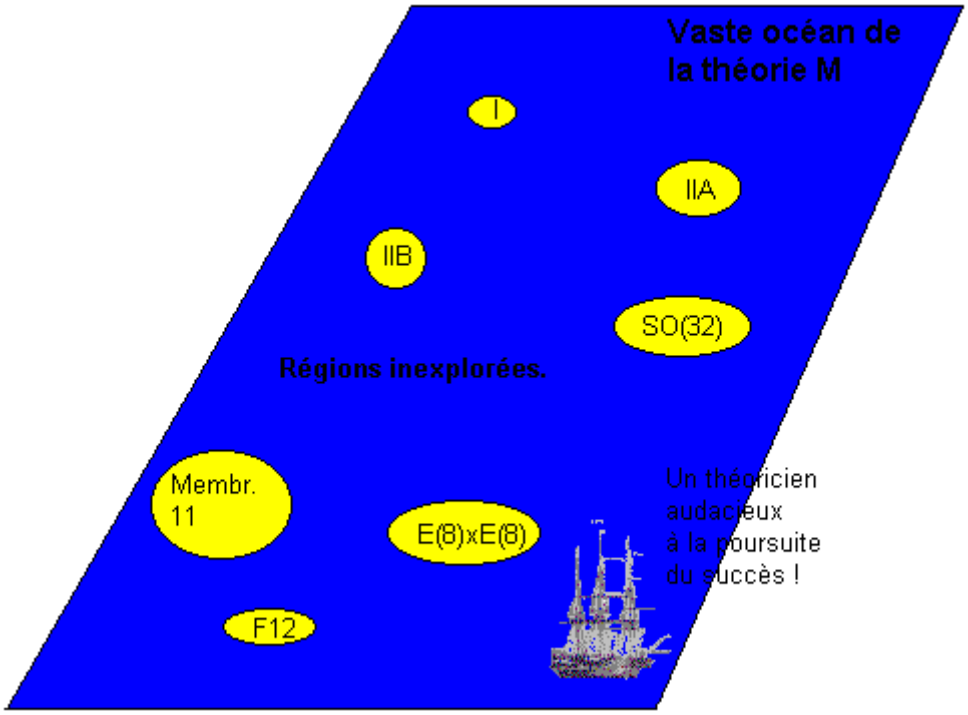


Notons que cette théorie n'est toujours qu'un cas limite avec un domaine d'application particulier. Comme pour chacune des théories des cordes existantes. Ce qui distingue ces théories (fort ou faible couplage, etc.) reste donc aussi d'application pour cette théorie et elle ne constitue donc pas notre fameuse théorie M (bien qu'elle soit parfois abusivement appelée comme cela à cause des membranes).

Il existe même une théorie F à 12 dimensions !

La théorie M

Bien que différente de la théorie présentée ci-dessus, on a pu déterminer que la théorie M serait une théorie à 11 dimensions. Ceci est évidemment très satisfaisant puisque c'était un des problèmes rencontrés. On a vu qu'une théorie de super gravité consistante nécessitait un espace-temps à 11 dimensions.

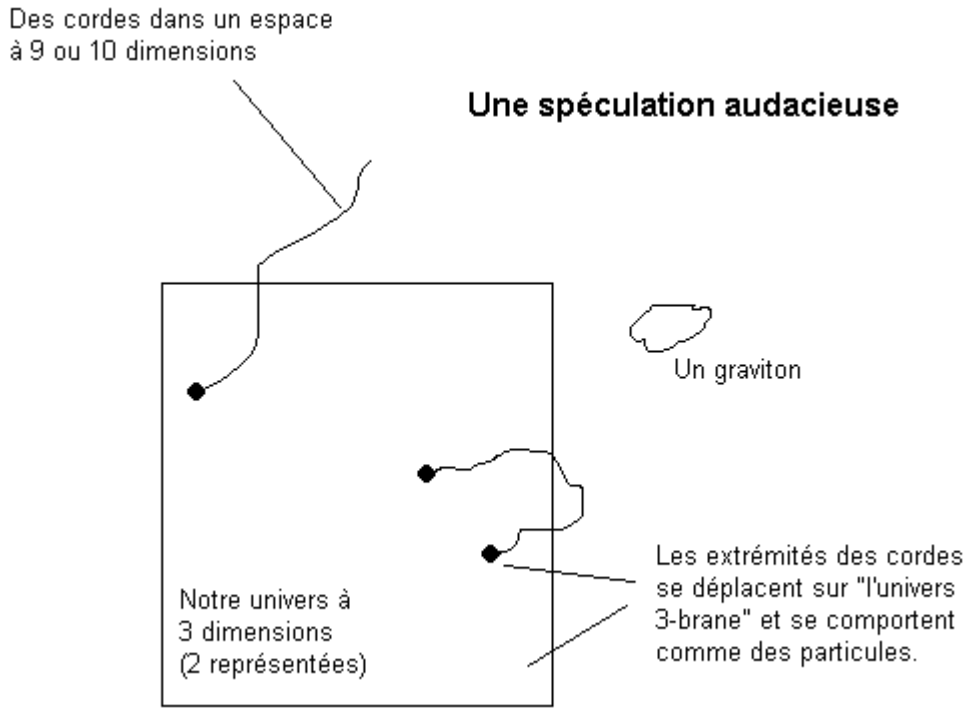


Différentes approches et études des différentes théories ont déjà permis d'audacieuses idées et spéculations.

Ainsi, on a par exemple imaginé que notre univers (à quatre dimensions) ne serait rien d'autre qu'une 3-brane (plus le temps) se déplaçant dans un univers à 10 dimensions (d'espace).

Les particules (fermions et bosons de jauge obéissant aux groupes habituels $SU(3) \times SU(2) \times SU(1)$) ne seraient que les modes de vibrations des cordes dont l'extrémité est attachée à la 3-brane (notre univers). Extrémités qui se comporteraient comme des particules !

Le graviton serait une corde fermée se déplaçant dans la totalité de l'espace à 10 dimensions (il n'est pas attaché à la D-brane). Une conséquence intéressante est que cela explique naturellement la faiblesse de la gravité par rapport aux autres interactions fondamentales.



Ce genre d'univers autoriserait même un repliement de ces dimensions sur des distances beaucoup plus grande que la longueur de Planck. En effet, les photons (la lumière) et les particules (qui nous composent) étant astreint à se déplacer sur la 3-brane, ils ne nous permettraient pas de "voir" les dimensions supplémentaires ! Mais l'enroulement ne peut pas dépasser le millimètre car la gravité a été testée jusqu'à cette distance sans constater d'écart avec la gravité normale. En effet, à grande échelle l'enroulement des dimensions supplémentaires rend celles-ci sans influence et la gravité se comporte comme elle est décrite par la relativité générale (et la théorie de Newton) dans un espace-temps à 4 dimensions. Mais pour des distances de l'ordre des dimensions enroulées, le graviton pouvant se déplacer dans des dimensions supplémentaires, les lois classiques devraient fortement dévier de la gravité observée.

On peut même imaginer qu'il y a plusieurs 3-branes et donc plusieurs univers qui n'interagiraient que par l'intermédiaire de la gravité à travers l'espace entre les 3-branes (appelé bulk). Ce qui donne des idées pour expliquer l'origine de la masse manquante de l'univers (voir les théories cosmologiques) ! Ce genre de théorie a même déjà été envisagé en dehors du contexte de la théorie des cordes (univers gémellaires, etc.).

En tout cas, toutes spéculatives qu'elles soient, ces idées autorisent des enroulements plus variés des dimensions et enrichissent encore la théorie.

Quoi qu'il en soit, la théorie M est encore imparfaitement comprise et sa forme mathématique reste en réalité complètement à imaginer. En particulier on a besoin d'outils non perturbatifs plus puisant pour "débroussailler" la théorie.

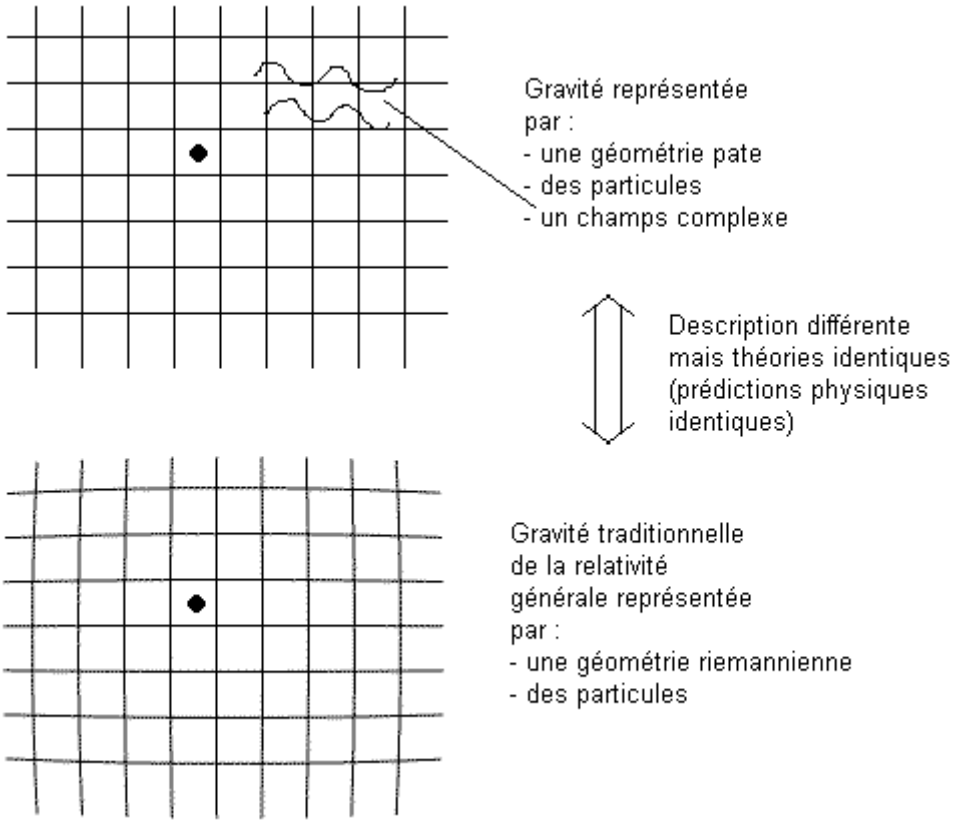
Toutefois de nombreux résultats quant à ses propriétés ont déjà été établis, c'est encourageant.

Une formulation matricielle des théories des cordes semble actuellement donner un cadre qui pourrait s'avérer le bon (cela reste encore à voir).

Un certain rapprochement avec les théories géométriques (géométries non commutatives, théorie des boucles, mousses de spin,…) est également en train de se faire. Mais ce n'est pas sans difficulté puisque l'espace-temps des cordes, quel que soit le nombre de dimensions envisagé, a une géométrie commutative. Pire encore, l'espace-temps de la théorie des cordes est indépendant du

contenu, contrairement à la relativité générale (comme on l'a vu), d'où de grosses difficultés pour relier les deux théories.

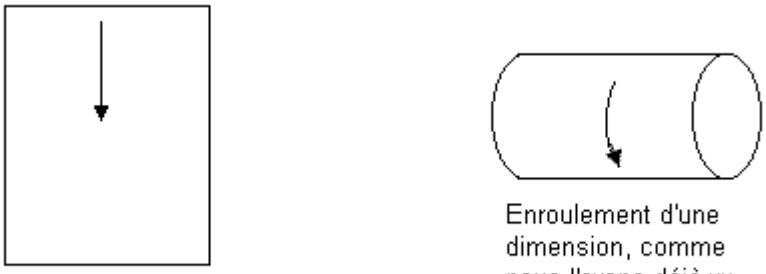
Mais ces rapprochements ne sont pas impossibles. Par exemple, la relativité générale a pour cadre un espace-temps courbe, mais il est possible de la formuler différemment : dans un espace-temps plat où ce qui tient normalement lieu de courbure est décrit par un champ du troisième rang (en relativité générale ce sont les symboles de Christoffel). On s'est ainsi rendu compte que la géométrie classique n'était probablement pas la meilleure pour décrire les cordes et des résultats fort encourageants ont déjà été obtenu.



7.3. Le repliement des dimensions

Nous nous retrouvons donc avec une (ou plusieurs) théorie à 10 ou 11 dimensions. Mais nous savons que l'univers observable n'a que 4 dimensions. Il faut donc que les 6 ou 7 dimensions supplémentaires soient recourbées sur une distance si petite qu'elles sont inobservables.

Nous avons déjà présenté cette idée avec la théorie de Kaluza - Klein. Si l'on a une dimension supplémentaire elle peut être enroulée sur une toute petite distance et elle devient inobservable. Si l'on représente l'espace normal par une feuille, en chaque point on a une petite boucle qui correspond au parcours de cette dimension supplémentaire.

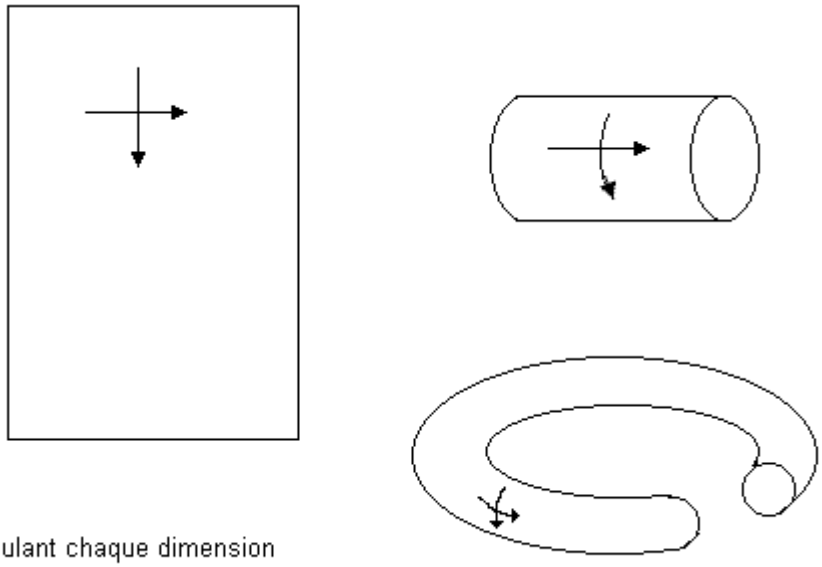


Enroulement d'une dimension, comme nous l'avons déjà vu

Bien entendu cette dimension a des conséquences physiques même si elle ne peut pas être directement observée. Par exemple, dans Kaluza - Klein, cette dimension supplémentaire provoquait, dans une théorie purement gravitationnelle, l'apparition du champ électromagnétique et de la charge électrique. On dit que cette dimension supplémentaire introduit des degrés de libertés supplémentaires autorisant l'apparition de symétries auquel correspond une charge conservée, ici la charge électrique.

Ce repliement est aussi appelé compactification des dimensions.

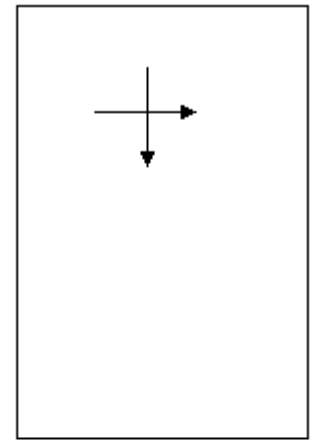
Que se passe-t-il si nous avons deux dimensions à replier ? Le cas le plus simple est l'enroulement séparé de chacune des dimensions sur un cercle.

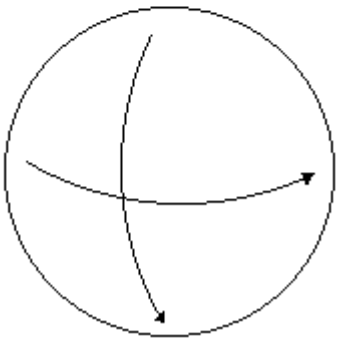


En enroulant chaque dimension séparément, on obtient un tore

Les deux dimensions forment alors une variété analogue à un tore, c'est à dire la forme d'un pneu.

Mais on peut aussi replier les deux dimensions en même temps.



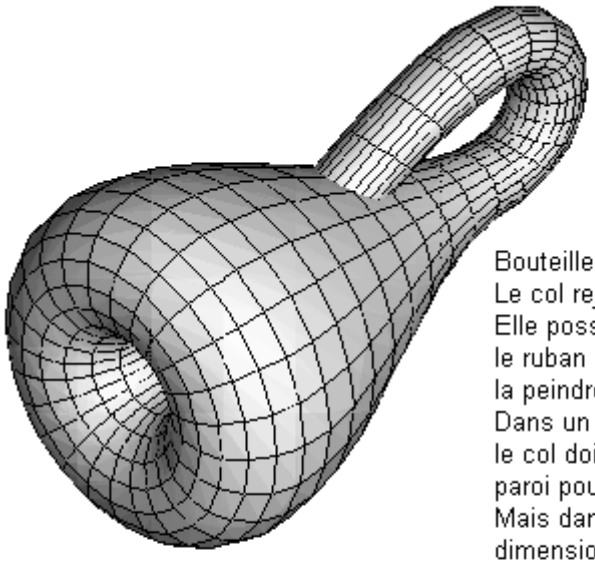


Et en repliant les deux dimensions à la fois ont obtient une sphère

Dans ce cas, les deux dimensions forment une sphère.

D'une manière générale, on peut imaginer que les deux dimensions forment une variété plus ou moins complexe.

On peut imaginer d'autres variétés à deux dimensions pour le repliement.



Bouteille de Klein.
Le col rejoint le fond par l'intérieur.
Elle possède une seule face (comme le ruban de Moëbius) : essayez de la peindre !
Dans un espace à trois dimensions, le col doit forcément perforer la paroi pour rentrer à l'intérieur.
Mais dans un espace avec plus de dimensions, ce n'est pas le cas (difficile à visualiser !)

On peut même replier les dimensions sur un noeud.
Bonjour les complications mathématiques !

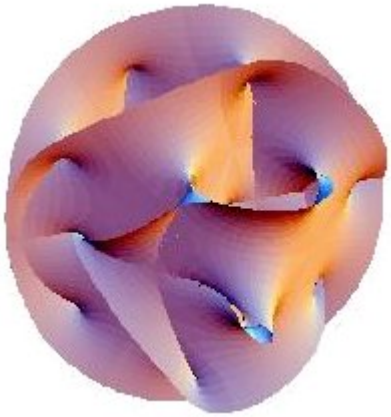
Cela donne déjà un nombre immense de possibilités !

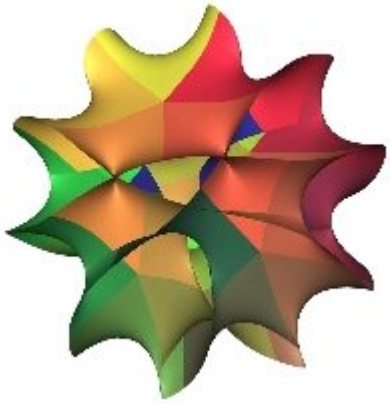
Lorsque le nombre de dimensions croît le nombre de possibilités augmente encore. On peut avoir des tores à plusieurs dimensions (difficiles à dessiner). Par exemple le tore à quatre dimensions est noté T4. Ou des sphères à 5 dimensions (S5), etc.

On peut même avoir des variétés très particulières appelées orbifold. Ce sont des variétés "anguleuses" (dont les dimensions se recoupent en formant des angles aigus au contraire des exemples ci-dessus où les formes sont "douces" et rondes). Les singularités (les points aigus) obéissent à une structure de groupe, comme pour les symétries.

Un type de variété est appelé variété de Kahler (notée K3) et est une variété de type orbifold obéissant à la géométrie dite de Kahler. On trouve aussi des variétés dites de Calabi - Yau qui sont du style de celle de Kahler mais avec d'autres règles (plus complexes) les définissant.

Les variétés de Calabi - Yau forment une classe gigantesque ! On en connaît des milliers d'exemples. Elles sont vraiment très complexes (et difficiles à décrire). Voici par exemple des coupes à deux dimensions (pour pouvoir les dessiner !) effectuées dans une variété particulière de Calabi - Yau à 6 dimensions.

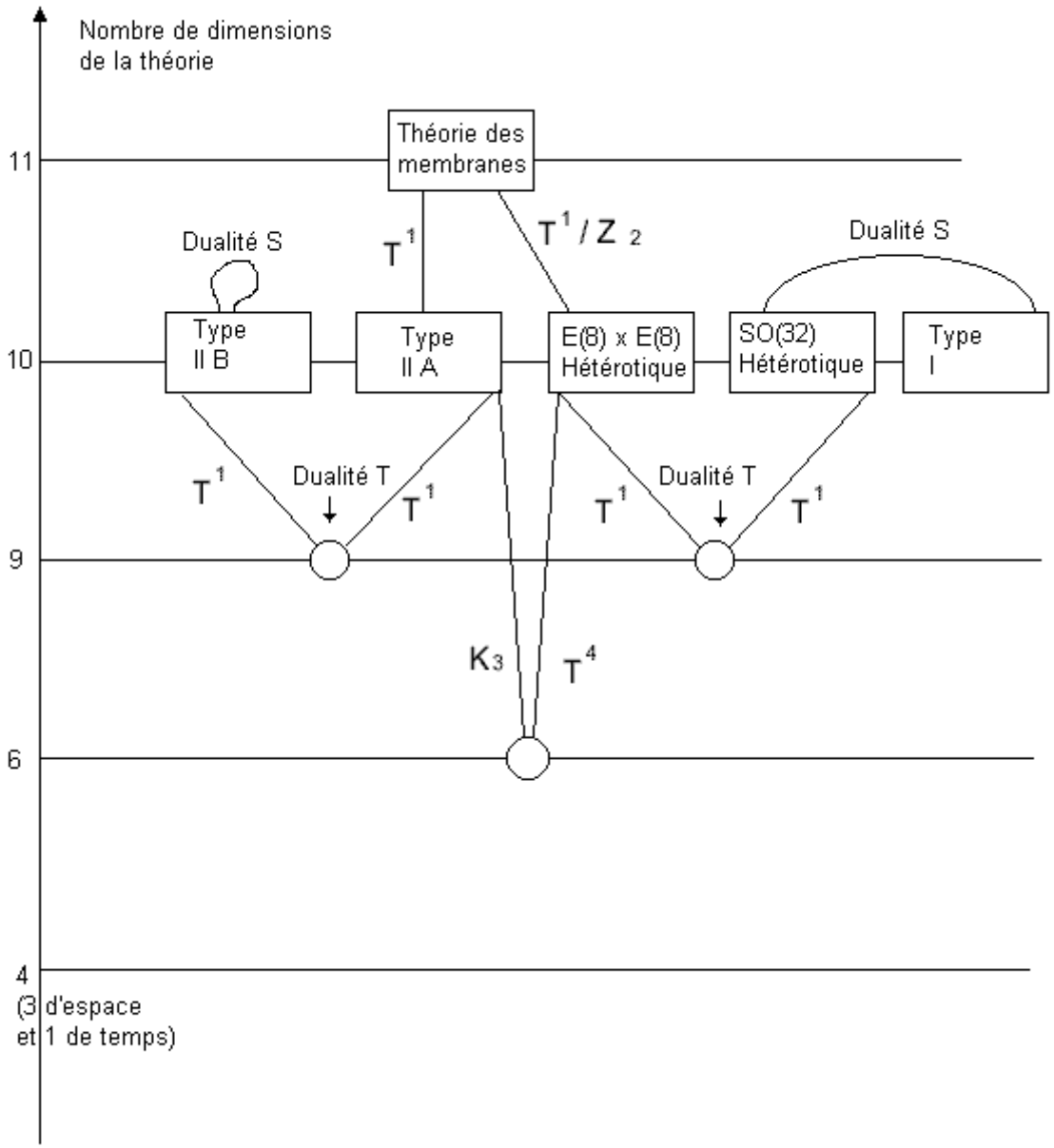




Le domaine des mathématiques qui s'applique à l'étude de ces variétés est la topologie algébrique et c'est une branche très compliquée (et très vaste) des mathématiques.

Lorsque l'on effectue l'analyse de la théorie avec ces dimensions enroulées, on constate que les propriétés de la théorie changent beaucoup. Les résultats peuvent être très variables en fonction de la compactification choisie. Le spectre des particules peut varier, ainsi que les interactions, les symétries de jauge,… Nous avons déjà vu que certaines théories perdaient leur chiralité (la violation de la parité) lorsqu'elles étaient compactifiées.

Mais même dans cet immense océan de possibilités, des dualités ont été trouvées. C'est à dire que l'on a découvert que diverses compactifications de diverses théories pouvaient conduire à la même théorie. On les appelle des dualités miroirs.



Cela donne encore une meilleure vue de l'unité de ces théories.

Nous avons déjà vu aussi que dans certains cas la théorie autorisait un compactage spontané des dimensions. C'est à dire que des conditions sur la stabilité du vide et sur l'énergie conduisent naturellement l'espace à se replier. Ceci est analogue avec la brisure de symétrie et a d'ailleurs un lien étroit avec, car certaines compactifications brisent les symétries de jauge ! Elles peuvent fournir un mécanisme pour ces dernières. En effet, dans la théorie initiale à 10 (ou 11) dimensions, les particules (les cordes dans un état de vibration) sont toutes sans masses. Et après compactification, sans brisure de symétrie, les particules restent toutes sans masse ce qui est passablement irréaliste. C'est un domaine actuellement en pleine effervescence.

En 1981 Witten a démontré que la compactification sur un nombre impair de dimensions empêche l'existence de fermions avec chiralité. C'est un résultat important qui exclut certaines compactifications.

Le théorème de Witten ne s'applique pas pour des compactifications "spéciales". C'est à dire sur des variétés présentant des singularités comme les orbifolds. Cela n'empêche pas d'étudier des compactifications plus simples (sphères, tores). D'une part les calculs sont plus simples ! D'autre part

cela permet de dégager un certain nombre de propriétés permettant d'aborder les cas plus complexes des variétés de, par exemple, Calabi - Yau.

Nous avions déjà dit que la théorie des cordes hétérotiques $E(8) \times E(8)$ autorisait un compactage spontané sur une variété de Calabi - Yau conduisant à une théorie (presque) réaliste : apparition du groupe de jauge $SU(3) \times SU(2) \times SU(1)$, violation de la parité, présence du graviton,…

Le principal problème est la démultiplication des possibilités (malgré les dualités miroirs). On a différentes théories avec différentes descriptions possibles (D-branes, conditions aux limites : par exemple l'univers "vu" comme une 3-brane,…) et des millions de compactages possibles.

On a donc des milliards de "théories" possibles (virtuellement infini) ! En réalité nous n'avons qu'une seule théorie, comme nous l'avons vu, mais les manières de l'appliquer au monde réel (ou de tenter de le faire !) et d'essayer de retrouver les résultats possibles, sont multiples.

Actuellement de nombreux axes de recherches sont en cours, notamment avec l'étude des aspects géométriques, des brisures de symétrie,… Mais pour essayer de retrouver le compactage "réaliste" on utilise une approche phénoménologique. C'est à dire que l'on passe au crible (parfois en s'aidant d'ordinateurs !) ces milliards de possibilités pour essayer de trouver celle qui colle aux observations et aux théories "classiques" (relativité générale, théories des champs et modèle standard).

Le problème avec cette approche est que si l'on trouve une "bonne théorie", comment pourra-t-on réellement affirmer qu'il s'agit d'une prédiction de la théorie des cordes ? Parmi une infinité de possibilités on a simplement choisi celle qui collait aux observations ! Il resterait alors à expliquer : pourquoi celle-là ? Ce qui sera probablement une autre paire de manche. Vu la rapidité de l'évolution des travaux dans ce domaine, on aura peut-être des nouvelles importantes d'ici peu.