QUANTUM GRAVITY:

A FUNDAMENTAL PROBLEM AND SOME RADICAL IDEAS

G. 't Hooft

Institute for Theoretical Physics
University of Utrecht
The Netherlands

Introduction

A complete theory of the gravitational forces cannot possibly be constructed without taking into account the basic principles of quantum mechanics. Indeed, many attempts are made by general relativists to "quantize" their theory. Being a particle physicist with some experience in second quantization problems, I feel that I should give my point of view, which differs somewhat from most conventional approaches.

In the first part of my lectures I will show how one derives the Feynman rules of quantum gravity as a gauge theory. The method is the conventional one: that is through a perturbation expansion.

Even though, obviously, perturbation expansion cannot tell us everything about the system, it is crucial to understand its structure. The difficulties associated with this expansion will have to be taken into account in the real theory also. If you really want to go beyond perturbation expansion you are often tempted to make models and approximations that obscure these difficulties. However, on the basis of these I claim that if you really want to do quantum theory (quantum gravity) in a non-perturbative fashion then you might have to change radically the whole concept of the ideas you have about gravity itself.

For a start I will just show you very briefly what the perturbation theory looks like and also that the first phenomenon to arise then is that of divergencies. Well, particle physicists are quite used to divergencies of this nature and we have a good answer to these, which is renormalization of the theory. So we are going to investigate renormalization properties in quantum gravity and we find which counter terms would be necessary to obtain a finer theory. That is where the first real problems will arise, which will force me later on to abandon this scheme. But also there are some important results from this simple continuum theory with its ordinary perturbation expansion. There are some problems, which are occasionally discussed in the literature, but which I will show not be be really physical. Then I will discuss why pure gravity is called renormalizable up to the one loop level. In the last lecture, I will discuss something radically different.

A. CONVENTIONAL THEORY

1. Gauge Theory - Feynman Rules

The first procedure to be employed is the so-called Wick rotation into Euclidean space [1]. That implies that we obtain a flat background metric with signature (+,+,+,+). Physically this procedure corresponds to analytic continuation to imaginary time. The nice thing about it is that all field equations become elliptical in nature rather that hyperbolic which makes computations easier and more convergent. Oscillations about this flat metric are considered as a small perturbation :

$$g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu} . \tag{1}$$

As our Lagrangian we take the familiar one :

$$\mathcal{L} = -\sqrt{g}\, R . \tag{2}$$

(Note that now $g > o$.) This is to be expanded with respect to $h_{\mu\nu}$. Let us write the terms quadratic in h explicitly :

$$\begin{aligned} \mathcal{L} = & -\tfrac{1}{4}(h_{\alpha\beta,\nu})^2 + \tfrac{1}{4}(h_{\alpha\alpha,\mu})^2 \\ & -\tfrac{1}{2} h_{\alpha\alpha,\beta} h_{\beta\mu,\mu} + \tfrac{1}{2} h_{\nu\beta,\alpha} h_{\alpha\beta,\nu} \\ & + \mathcal{O}(h^3) + \text{total derivative} . \end{aligned} \tag{3}$$

The commas denote ordinary derivatives, and the total derivative is irrelevant here.

The presence of higher order terms implies that the equations for gravitational waves are non-linear so that you can get scattering. The linear parts of the wave equations are trivial to quantize : the waves with given frequency must simply be quantized in energy units $h\nu$, called gravitons. The higher order terms in the Lagrangian now cause gravitons to interact. The scattering problem is really the problem we are going to address ourselves to, in particular the quantum corrections to the scattering amplitudes. The computational rules to obtain these amplitudes, and in principle also their quantum corrections have been obtained in particle physics.

The history of the derivation of the various steps in this procedure is rather remarkable [2]. The so-called non-Abelian gauge theories for elementary particles were investigated at first mainly because of their close resemblence to quantum gravity [3]. They served as a simplified model of gravity, and seemed to be of little relevance for the observed world of elementary particles. Then a breakthrough took place and they became extremely successful for the weak and electromagnetic interactions, a few years later the strong interactions. They were intensively studied. Now the details are well-understood and can be applied to quantum gravity.

The point is that there is an invariance in the system : invariance under infinitesimal coordinate transformations, corresponding to

$$h_{\mu\nu} \to h_{\mu\nu} + \xi_{\mu;\nu} + \xi_{\nu;\mu} + \mathcal{O}(\xi_\mu^2) , \tag{4}$$

where the semicolons are covariant derivatives. Any choice of the function $\xi_\mu(x)$ yields an equivalent set for $h_{\mu\nu}(x)$. This redundancy in the description of the solutions to the wave equations must first be removed. The prescription is as follows [4] : choose any four component function $C_\mu(x)$, composed of combinations of $h_{\alpha\beta}$, not invariant under eq.(4). It is preferred but not crucial to have C linear in h. We now require that the equation $C_\mu = 0$ fixes the values of $\xi_\mu(x)$ (up to possible boundary effects). Or, if under (4)

$$C_\mu \to C_\mu + \hat{M}_{\mu\nu} \xi_\nu , \tag{5}$$

then the object $\hat{M}_{\mu\nu}$ which may be an operator in ξ-space containing derivatives, must have an inverse, $M^{-1}_{\mu\nu}(x,x')$.

Instead of Lagrangian (2) we now take

$$\mathcal{L} = -\sqrt{g}\,R - \tfrac{1}{2} C_\mu^2 \,, \tag{6}$$

which transforms under (4) as

$$\mathcal{L} \rightarrow \mathcal{L} - C_\mu \hat{M}_{\mu\nu} \xi_\nu \text{ + total derivative.} \tag{7}$$

Let us consider the Euler-Lagrange equations generated by Lagrangian (6). A function $h_{\mu\nu}(x)$ is a solution to those equations if any infinitesimal variation about $h_{\mu\nu}(x)$ does not change the total action $\int \mathcal{L} d^4x$ linearly. Let us in particular consider the infinitesimal variation (4) which happens to be also a coordinate transformation. According to eq.(7) $\int \mathcal{L} d^4x$ changes by an amount $\int C_\mu \hat{M}_{\mu\nu} \xi_\nu$. This must vanish for any choice of $\xi_\nu(x)$. Since $\hat{M}$ has an inverse it follows that $C_\mu = o$, if $h_{\mu\nu}$ satisfies these field equations. Thus the replacement (6) corresponds to imposing the "gauge condition" $C_\mu = o$ on our fields. A convenient choice for C_μ is

$$C_\mu = h_{\mu\alpha,\alpha} - \tfrac{1}{2} h_{\alpha\alpha,\mu} \,. \tag{8}$$

We then obtain

$$\mathcal{L} = -\tfrac{1}{4}(h_{\alpha\beta,\nu})^2 + \tfrac{1}{8}(h_{\alpha\alpha,\nu})^2 + \mathcal{O}(h^3) + \text{ total derivative.} \tag{9}$$

This Lagrangian is especially convenient because in the quadratic terms the differentiation index ν does not mix with the field indices α, β . Therefore the linear parts of the field equations become just Laplace's equation which is trivial to solve.

To solve the complete classical Euler Lagrange equations for a scattering process perturbatively is now straightforward. What is needed is the inverse of the Laplace operator, called propagator for the graviton, and the explicit form of the interaction terms $\mathcal{O}(h^3)$, called vertices. The combinatorial rules for these can also be interpreted as the history of a particular scattering process between gravitons. An example is given in Fig.1.

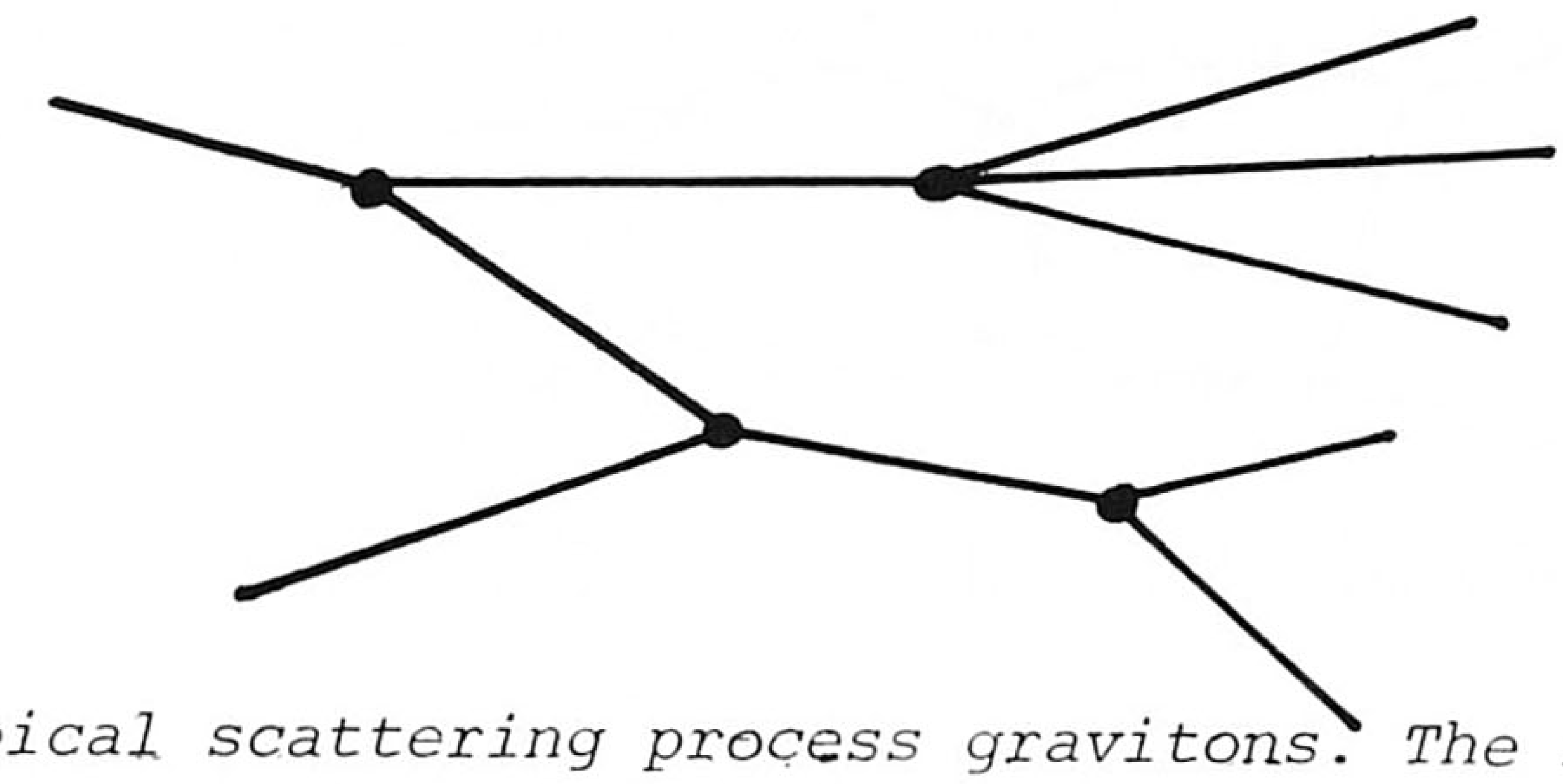


Fig.1 *Typical scattering process gravitons. The lines between dots are propagators (defined by the quadratic parts in (9).*

A quantum theory only differs from the classical theory by allowing in addition also diagrams with closed loops, see Fig.2.

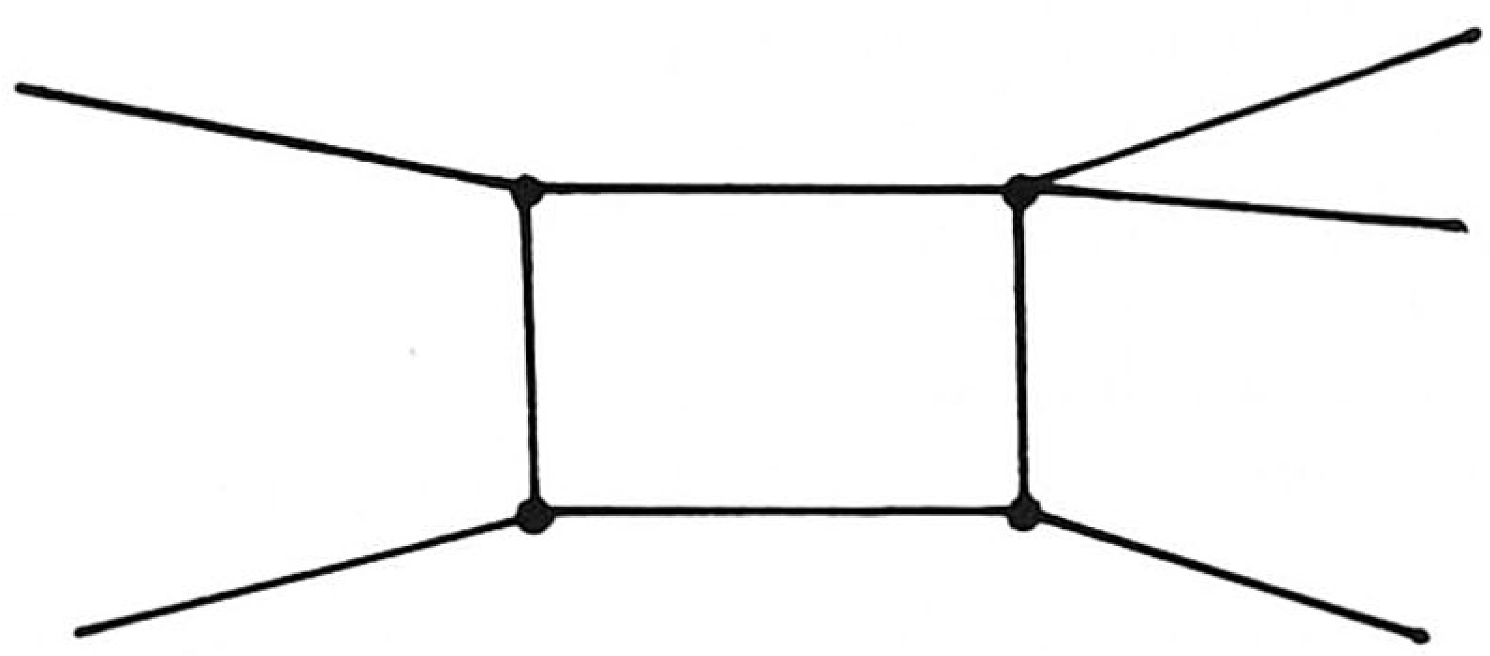


Fig.2 *Feynman Diagram with a loop.*

In simple field theories such as ϕ^4 the rules for the loops follow straightforwardly from the rules for trees such as Fig.1. Here however the rules suggested by the tree diagrams do not give the correct quantum theory. One must add a new so-called ghost field. This field does not correspond to any new physical particle but is needed to make the mathematics sound. It is obtained by adding to the Lagrangian

$$\mathcal{L}^{\text{ghost}} = \varphi_\mu^* \hat{M}_{\mu\nu} \varphi_\nu \; . \tag{10}$$

Here the object $\hat{M}$ is the same $\hat{M}$ as defined in eq.(5). It contains space-time derivatives acting on φ_ν on the right and terms very non-linear in h. Consequently the field equations for φ and φ^* have linear parts of the Laplace form and parts interacting with the h field. Because of the nature of the Feynman rules φ and φ^* can only occur in closed loops (see Fig.3). Finally, the weight factor for these diagrams is not the same as for ordinary theories without ghosts : one must insert an extra factor $(-1)^N$, where N is the number of ghost loops. Such a factor

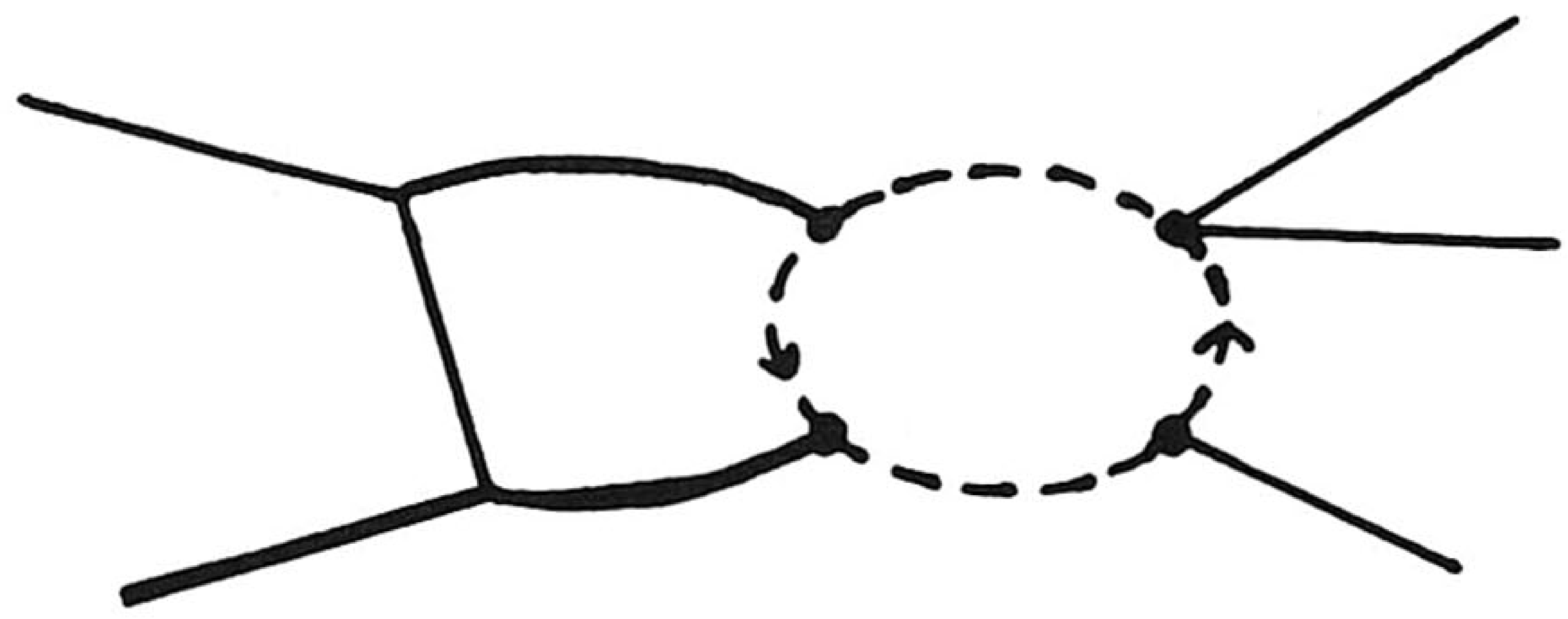


Fig.3 *Example of a diagram with ghost loop (and an ordinary loop).*

corresponds to attributing to this ghost wrong statistics : it is a fermion rather than a boson.

This completes the formal description of the Feynman rules. I don't know how much this has already been discussed in previous lectures, where I could not be present, but this is mainly the result of the investigation of gauge theories in particle physics. This part of it is all well understood. It is also well understood that this prescription is unique; there is very little room for playing around with these rules, or using other rules instead.

The reason why this prescription is so precise is that we want to describe scattering of gravitons by means of an S-matrix. This matrix must satisfy the requirements of causality and unitarity. If it is not unitary it would mean that probability is not conserved and that would make a totally senseless theory out of it. This unitarity condition is so restrictive that it prescribes in a unique fashion what the Feynman rules of this theory should be ; there is no way to play around with it. So that part of the conventional theory is nice and sound and understood. That does not mean that the theory will work all the way and the reason is of course something that you also know of in field theories - the divergencies, to which I now come.

2. Divergencies and Counter terms

Considering the natural unit of energy, $\sqrt{\hbar c^5/G} = 10^{22}$MeV, quantum effects in gravity are unlikely to be ever observed experimentally. Nevertheless, the theory dictates that gravitons can split, form closed loops and rejoin while they scatter. Such processes, of which Figs. 2 and 3 are examples must necessarily occur in a good theory, otherwise the S-matrix will not be unitary and probabilities will not be conserved.

If one computes the contributions coming from such diagrams one finds that they are actually highly divergent. The momenta and energies of the gravitons in the intermediate states must be integrated over. One-loop diagrams typically correspond to integrals of the type

$$\int d^4k \, \frac{Pol_1(k)}{Pol_2(k)} \quad , \tag{11}$$

where the two polynomials can be of the same degree, so that the integral is quartically divergent where $k_\mu \to \infty$. But we have a scheme in gauge theories to deal with that, called dimensional renormalization [6] : in fact there are other renormalization schemes but this is the most convenient one. We replace d^4k by d^nk, saying that we are now going to do the same theory in n dimensions instead of 4. The crazy trick that we are going to do is that we choose n very close to 4 but just a non-integer number (maybe a complex number) near 4. Now complex dimension is of course a totally crazy concept if you are dealing with the full general relativistic theory. However, when you consider these Feynman rules of the system and the formulations of it you find that it makes perfectly good sense to define integrals over n-dimensional momenta instead of 4-dimensional momenta.

The point is that these integrals all have a very special form, and for this class of integrals a unique definition of the symbol $\int d^nk$ can be given. One basic ingredient is the extension of the expression for the surface of an n-dimensional sphere,

$$\int d^nk \quad \delta(k^2 - r^2) = \frac{\pi^{n/2}}{\Gamma(n/2)} \, r^{n-2} \tag{12}$$

to a non-integer n. Another ingredient is a procedure to obtain the "finite part" of an integral, which works only if the degree of divergence is not integer, which is why we first choose n to be non-integer. But what happens if you let n tend to 4 ? You encounter poles and this is just another way of saying that at 4-dimensions you have problems. The prescription then turns out to be that you simply have to add a counter term to your Lagrangian:

$$\mathcal{L} \to \mathcal{L} + \frac{1}{8\pi^2(n-4)} \Delta\mathcal{L} \quad , \tag{13}$$

where $\Delta\mathcal{L}$ is now chosen in such a way that all poles in the physically relevant amplitudes cancel and a finite result is obtained after adding all contributions from all one-loop diagrams.

In gauge theories for elementary particles, the new Lagrangian (13) is of the same type as the previous one, so that only some coupling constants are renormalized this way. We can then preceed to include the effects of two-loop diagrams, which leads to $(n-4)^{-2}$ terms, and so on. We will always require the limit $n \to 4$ to exist for physically observable quantities.

Now this $\Delta\mathcal{L}$ in (13) has been computed for many gravity theories. The first theory for which this $\Delta\mathcal{L}$ has been computed is of course pure gravity [7] :

$$\Delta\mathcal{L} = \sqrt{g}\left[\frac{1}{120} R^2 + \frac{7}{20} R_{\mu\nu}^2\right] . \tag{14}$$

This is the theory with gravitons alone, no matter-fields in addition. Even though $\Delta\mathcal{L}$ is of a different form than $\mathcal{L}$ itself, it is rather harmless, because for pure gravitons $R = o$ and $R_{\mu\nu} = o$ classically. We come back to that in sect. 4.

If matter-fields are coupled to gravity in the form of one real, scalar field, then there is a harmful part in $\Delta\mathcal{L}$:

$$\Delta\mathcal{L} = \sqrt{g} \cdot \frac{203}{80} R^2 . \tag{15}$$

(see ref. 7). Note that here $R \neq o$, even classically.

3. Solution of the Measure Problem.

If one does not follow the diagrammatic approach to quantum gravity one might run into an apparent difficulty. I just want to mention it very briefly and that is the measure problem. This comes up every now and then in the literature. This is a problem which seems to occur if you write down the theory in terms of the functional integral. Should we write this as

$$\int \prod_{x,\mu,\nu} dg^{\mu\nu}(x)\, e^{-\int \mathcal{L} d^4x} ,$$

or as

$$\int \prod_{x,\mu,\nu} dg_{\mu\nu}(x)\, e^{-\int \mathcal{L} d^4x} , \tag{16}$$

or some other way ? My claim is that this question is senseless at any order in perturbation expansion because both expressions (16) are acceptable. In fact any local, 10 component function of $g^{\mu\nu}(x)$ may be chosen as integration variable at the point x in the functional integral. The reason is that, after renormalization, all choices give precisely the same amplitudes and are therefore equivalent. Let me illustrate this for the two expressions (16). The difference is a factor in the integrand which is the Jacobian of the transformation

$$g_{\mu\nu}(x) \to g^{\mu\nu}(x) :$$

$$\det_{\substack{x,x' \\ \mu\nu,\alpha\beta}} \left[\frac{\partial g_{\mu\nu}(x)}{\partial g^{\alpha\beta}(x')} \right] = \exp\left(\mathrm{Tr}_{\substack{x,x' \\ \mu\nu,\alpha\beta}} \log \frac{\partial g_{\mu\nu}(x)}{\partial g^{\alpha\beta}(x')} \right) = \exp \sum_x 5 \log g(x) . \qquad (17)$$

The determinant in (17) is that of a huge matrix, whose rows and columns are labeled not only by the indices $\mu\nu$ and $\alpha\beta$ but also by the space-time points x and x'. The matrix is diagonal in x so that finally we obtained the exponent of a sum over all space time points x. One will now be tempted to replace $\sum_x$ by $\int dx$, but that is wrong ! The reason why that is wrong is that for the continuum limit of (17) to exist we need a quantity that produces the infinitesimal dx. Such a quantity is not there in (17). It is correct to say

$$\sum_x \to \frac{1}{d^4x} \int d^4x ,$$

in the continuum limit. This is quartically divergent ! Taking the limit more carefully we write

$$\frac{1}{d^4x} = \delta^4(0) = \int 1\, d^4k \to \infty . \qquad (18)$$

Now this latter expression may remind you of something. We have already seen such divergences before in the theory (eq.11) and that is the message I want to convey to you in this little section - that there is a divergence in this determinant which is exactly comparable to the divergences of the one loop diagram. In fact it can be shown to be indeed, in a certain sense, a one loop divergence (which again I have no time to go into details about). But this is essentially a one loop divergence which is just as bad as that in Fig.2. But now remember for this one loop divergence we have a unique procedure to get the convergent theory. So the same unique procedure can be applied here to replace this infinity by a finite number. To explain what happens I have to go a little bit more into the details of this dimensional regularization procedure.

At n different from but close to 4 expression (18) still diverges. But the "finite part" of the integral is then uniquely defined. In gauge field theories integrals may be replaced by finite integrals (notation : $\int_F d^n k$) if they satisfy the following requirements :

$$\text{(i)} \quad \int_F f(k)\, d^n k = \int f(k)\, d^n k \;,$$

if the right hand side converges.

$$\text{(ii)} \quad \int_F d^n k (A+B) = \int_F d^n k\, A + \int_F d^n k\, B \;.$$

$$\text{(iii)} \quad \int_F d^n k\, (\lambda A) = \lambda \int_F d^n k\, A \;. \tag{19}$$

$$\text{(iv)} \quad \int_F d^n k A(k+p) = \int_F d^n k\, A(k) \;.$$

$$\text{(v)} \quad \int_F d^n k\, A(\lambda k) = \lambda^{-n} \int_F d^n k\, A(k) \;.$$

It follows that

$$\int_F d^n k\, (k^2)^{\alpha} = 0 \qquad \text{if } \alpha \neq -n/2 \;. \tag{20}$$

Note that eqs (19) and (20) allow us to make all divergent integrals finite except if the degree of divergence is integer. For instance $\int d^n k/(k^2+a)^{n/2}$ remains infinite. This is because equation (20) is divergent both at ∞ and at o if $2\alpha = -n$, so using (20) and (19,ii) merely enables us to replace the divergence at infinity by a divergence at the origin and (19,i) cannot be used to obtain a finite expression. This is how, in the dimensional renormalization procedure,poles develop at n = 4.

The procedure immediately tells us what to do with expression (18). The only finite number it can be replaced with is zero. Therefore the factor 17 is one, and the ambiguity disappears.

Less elegant renormalization procedures exist where cutoffs are introduced. The factor (17) then corresponds to an additional local term in $\mathcal{L}$. Whenever gauge-invariance is imposed (by requiring validity of identities between amplitudes called Slavnov-Taylor identities [4], [8]) this ambiguity is fixed up to a gauge invariant term without derivatives. Only one such term exists : the cosmological term $\sqrt{g}$, so also in such procedures, after renormalization, the metrix ambiguity can be absorbed in a redefinition of the cosmological term. Because of power-counting arguments however the cosmological constant is not renormalized in the dimensional renormalization scheme.

The conclusion of this section is that, provided proper gauge-invariant renormalization procedures are used such as dimensional renormalization, the metric of the functional integral may be chosen any way you like. Since the divergences of eq.(18) add to all other equally bad divergencies of the theory it is absolutely senseless to defend one choice of measure against one other.

4. Pure Gravity is One-Loop-Finite

The next point which I am sure many of you have heard before is the statement that pure gravity is one loop finite. Why is it so ? Well remember I had a counter term, and this counter term was of a very dangerous nature because it contains the square of the Riemann curvature (see eq. 14). That is different from the previous Lagrangian which we started off with, which only had a linear expression in the Riemann curvature. To have square terms will lead to having four derivatives in your Lagrangian or three derivatives in your field equations. In such systems it is impossible to write down a Hamiltonian that is bounded from below, and that will make the whole of Minkowski space unstable, and you will get very serious blowups of the system. Energy is no longer positive definite. The field theory, in fact, definitely does not work if you do this. A gauge theory for instance with such terms would be thrown out immediately.

However, we say nevertheless that pure gravity is one loop finite because if you have pure gravity we know that the classical equations will tell you that $R = 0$ and $R_{\mu\nu}$ equals zero as well. So we say

$$\Delta \mathcal{L} = \sqrt{g}\left[\frac{1}{120} R^2 + \frac{7}{20} R_{\mu\nu}^2\right] = 0 . \tag{21}$$

Why ? Usually it is not correct to insert an equation of motion back into a Lagrangian. The reason why it is allowed here is that $\Delta\mathcal{L}$ is only infinitesimal (in spite of the fact that $\frac{1}{n-4}$ is formally infinite, since only terms in a given order of perturbation expansion have been collected).

Theorem : $\mathcal{L}$ and $\mathcal{L} + \varepsilon\Delta\mathcal{L}$ are equivalent up to terms of order ε^2, if $\Delta\mathcal{L}$ vanishes for all those field configurations that satisfy the field equations generated by $\mathcal{L}$.

Proof : The equation of motion reads, in an abbreviated notation :

$$\frac{\eth \mathcal{L}}{\eth h_{\mu\nu}(x)} = 0 , \tag{22}$$

where $\eth$ stands for derivative in the Euler sense. Therefore we may assume

$$\Delta \mathcal{L}(h) = \frac{\eth \mathcal{L}}{\eth h_{\mu\nu}(x)} Q_{\mu\nu}(h,x) \tag{23}$$

where $Q_{\mu\nu}$ is any local function of the fields h at x. Otherwise the equations of motion for h would not imply that $\Delta\mathcal{L}$ vanishes. But then

$$\mathcal{L}(h) + \varepsilon \Delta \mathcal{L}(h) = \mathcal{L}(h_{\mu\nu}(x) + \varepsilon Q_{\mu\nu}(h,x)) + \mathcal{O}(\varepsilon^2) \tag{24}$$

If we now take as our new field variables

$$h'_{\mu\nu}(x) = h_{\mu\nu}(x) + \varepsilon Q_{\mu\nu}(h,x) , \tag{25}$$

then

$$\mathcal{L}(h) \Rightarrow \mathcal{L}(h') + \mathcal{O}(\varepsilon^2) , \tag{26}$$

therefore, the two Lagrangians are equivalent up to terms of order ε^2. Physically, the fact that all terms of order ε can be absorbed by a field redefinition implies that they are not directly observable. In particular, when it comes to computing S matrix elements the results will be identical up to ε^2 terms. But the $\mathcal{O}(\varepsilon^2)$ terms would have to be added to all two-loop processes which are of the same order of magnitude and have been neglected so far.

5. Field Transformations and Gauge Dependence

To be more explicit in this given example of pure gravity, we have

$$\frac{\eth \mathcal{L}}{\eth g_{\mu\nu}} = R^{\mu\nu} - \tfrac{1}{2} R g^{\mu\nu} , \tag{27}$$

so that the redefinition of the metric tensor has to be

$$g'_{\mu\nu} = g_{\mu\nu} + \delta g_{\mu\nu} \,,$$
$$\delta g_{\mu\nu} \propto \tfrac{7}{20} R_{\mu\nu} - \tfrac{11}{60} R\, g_{\mu\nu} \qquad (28)$$

On the other hand, when matter is present in the form of a scalar field, then $\Delta\mathcal{L}$ cannot be absorbed by a field renormalization and the first order infinity would be directly observable if it were not removed in any other way. This explains the usual statement that pure gravity is one-loop renormalizable and gravity with matter in general not.

A further remark is of order. Since the redefinition in (28) is not directly observable physically, one cannot exclude the possibility that $\delta g_{\mu\nu}$ is actually dependent on the way the procedure is being performed in quantizing gravity. It turns out that this whole procedure is very much dependent on the original choice of gauge. Remember I chose an object called C_μ to fix my gauge; had I chosen another C_μ I could have obtained something else. That's inherent in the procedure. Again, this is not in contradiction with general gauge invariance because such a redefinition is not physically observable - I cannot determine what my metric tensor is from scattering experiments. Anything that is not physically observable might be gauge dependent. That is a general law in gauge theory.

6. The Fate of the Metric Tensor

Ambiguities of the type of eq. (28) in the definition of the metric tensor are to my mind extremely important to observe. The coefficient in $\delta g_{\mu\nu}$ is not only formally infinite. Its finite part may be dependent on scale transformations (dilatations). That implies that after performing a scale transformation in the theory the metric tensor may have a different meaning than before. The conclusion from that is that there may not exist a unique definition of the metric tensor. There exists no formal procedure that distinguishes $g_{\mu\nu}$ as a basic field from say $g_{\mu\nu} + R_{\mu\nu}$ or $g_{\mu\nu} + Rg_{\mu\nu}$, or anything more complicated (at higher orders we encounter R^2 terms and so on).

Now what is the consequence of that? I think that perhaps here we are dealing with a basically new feature of quantum gravity. An illustration of that is given in Fig. 4.
In the conventional theory of general relativity the existence of a well-defined $g_{\mu\nu}$ was used as a starting point. The curvature $R_{\mu\nu}$ was a derived quantity. Now we see that at small distances quantum fluctuations of $R_{\mu\nu}$ (at higher orders, of objects such as

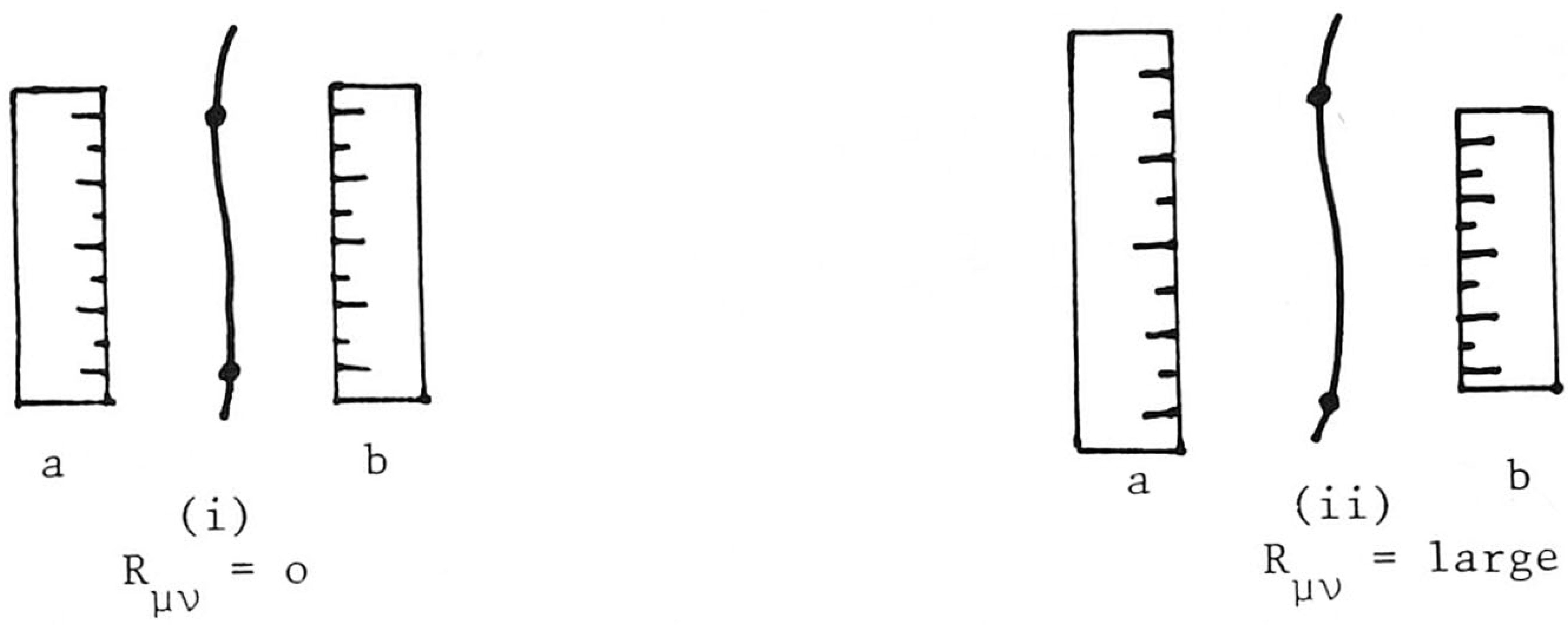


Fig.4 *Measuring distances in curved space. The two units* a *and* b *are sensitive to tidal forces in a different way (one responds to* $g_{\mu\nu}$*, the other, say, to* $g_{\mu\nu} + R_{\mu\nu}$*). Whenever there is a curvature, the metric tensor is ambiguous, since there are no criterea to prefer* a *or* b *as fundamental units of distance.*

$R_{\mu\alpha\beta\gamma} R_{\nu\alpha\beta\gamma}$) become large and the definition of $g_{\mu\nu}$ becomes obscured by admixtures of these curvature tensors. Fundamental criterea to distinguish the pure $g_{\mu\nu}$ from the polluted $g_{\mu\nu}$ are absent. Suppose for instance that the unit of distance is defined by the wavelength of some spectral transition. Quantum effects then give effective couplings of the form $R^{\mu\nu} g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta}$ etc. (and with infinite coefficients !). This indeed implies that not the original metric tensor $g^{\mu\nu}$, but one of the polluted $g^{\mu\nu\prime}$ tensors is observed. My conclusion is that the metric tensor may be not at all as fundamental as is usually supposed. At distances smaller that the Planck length perhaps nothing even remotely resembling a metric tensor can be defined.

7. QGD versus QED and QCD

As we have seen the fact that the counter terms are not the same as the original Lagrangian implies the famous statement that "gravity is not renormalizable". (The arguments presented in sect.4 only work at the one-loop level.) Now nonrenormalizable theories are worse than one might think. Let me explain that by taking some examples. The gravitational coupling constant, G, in units where h and c are put equal to one, has the dimension of an inverse mass squared. That means that at smaller and smaller distance scales (corresponding to higher and higher mass scales) gravitational effects become stronger and stronger. Precisely the same is true for weak interactions in particle physics (Fig.5). The Fermi coupling constant (accidentally also called G) has the same dimensionality, so that that theory also is non-renormalizable. In particle physics the resolution of the problems that arise is that new physics must be going on at the distance scale where G

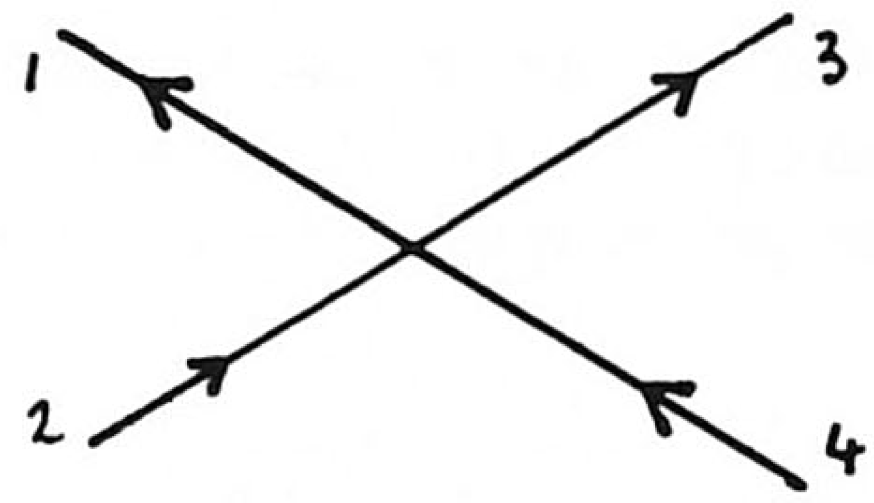


Fig. 5 *4-fermion interaction theory of weak interactions :*

$$\mathcal{L}^{int} = G\,\bar{\psi}_1\,\gamma_\mu(1+\gamma_5)\psi_2\,\bar{\psi}_3\,\psi_\mu(1+\gamma_5)\psi_4 \;.$$

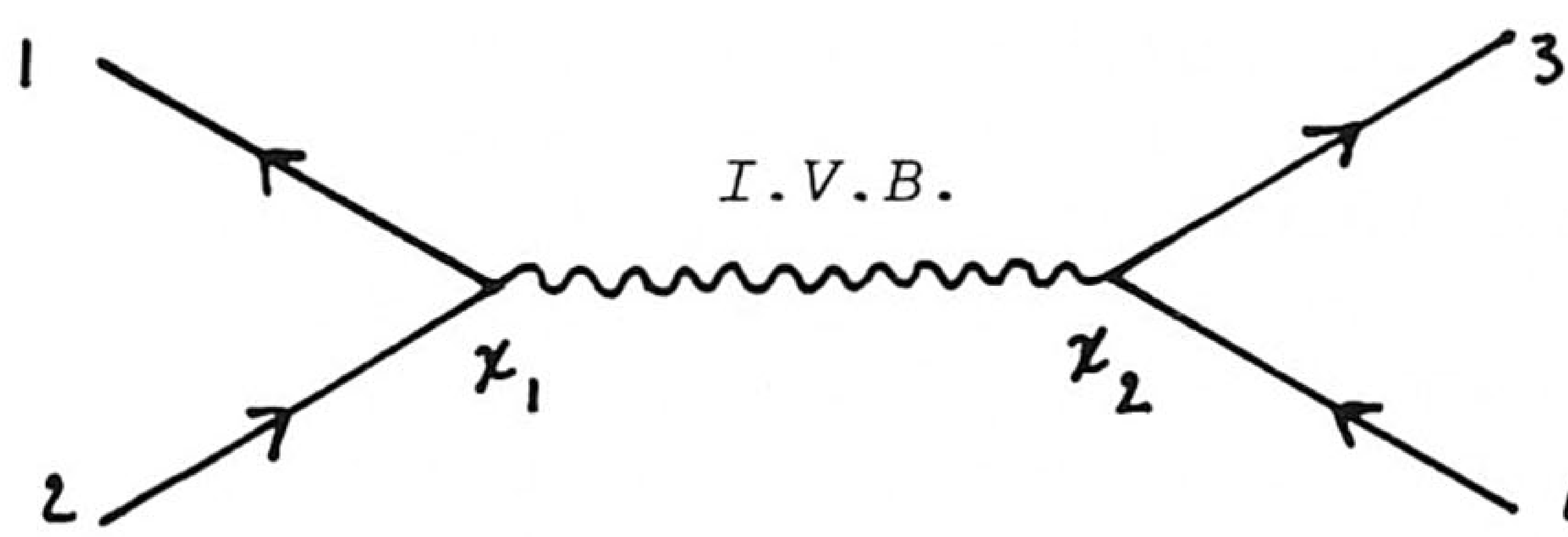


Fig. 6 *Intermediate Vector Boson theory : the interaction points x_1 and x_2 are now separated : the distance travelled by a virtual heavy particle (I.V.B.)*

becomes of order unity. In present-day theories it is generally expected that new heavy particles, called intermediate vector bosons, occur there, whose mass determines G, and they themselves only have dimensionless couplings (these couplings, by the way, have to obey the pattern of a gauge theory). The original 4-fermion interaction has now become a non-local one because it is described by an exchange process of a weak intermediate vector boson.

The well-known theory of quantum-electrodynamics (QED) is in much better shape, since it is described by a dimensionless small parameter α (the fine-structure constant). However, when scale transformations are performed more carefully one finds that α is not entirely dimensionless. At small distances it increases logarithmically, so that at exponentially large mass scales (10^{100} GeV, say) it reaches the value unity. Opinion on the interpretation of this differ, but my conclusion is that here also new physics is required.

The ideal theory in this respect is "quantum chromodynamics", a theory for strong interactions between quarks, where the coupling decreases with the mass-energy scale (though only logarithmically). This means that at small distances I have a less interactive theory and that may be a perfect theory, although mathematicians still put a question mark on that.

Clearly, basically new physics is needed to understand quantum gravitational dynamics at the Planck length-something like an intermediate boson being exchanged or something fundamentally different.

B. An alternative to the continuum theory ?

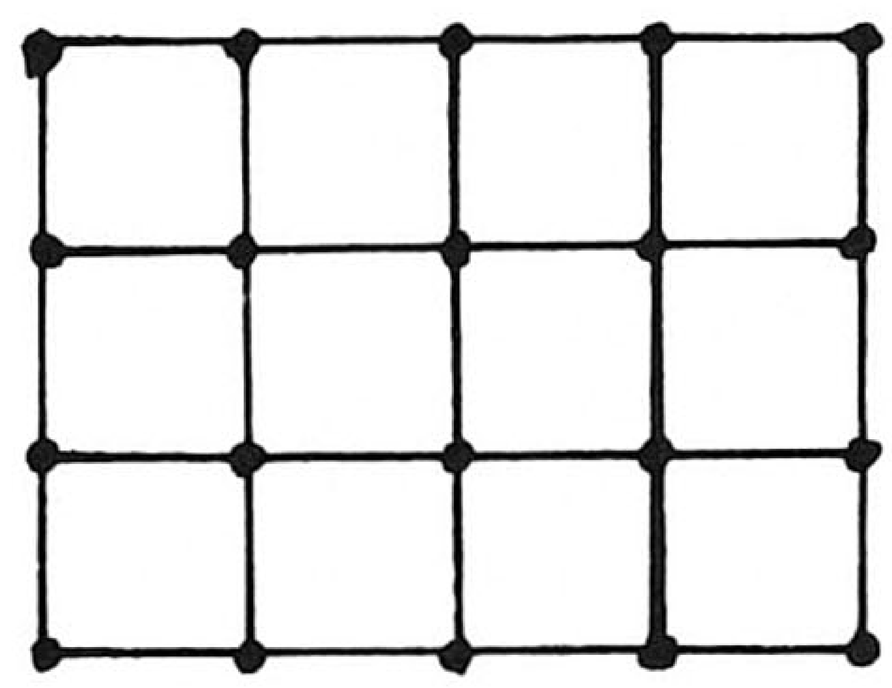
8. Discrete Space-time

So the big question is how could we possibly find an underlying theory that is even more basic that general relativity to explain quantum gravitational dynamics at the Planck length ? Now this is very difficult to conceive of and I just want to use the last part of these lectures to elucidate on the crazy idea about what possibly the physics could be. You see if my metric tensor no longer makes sense, if my momentum integrals diverge, this all means that there may be something basically wrong with working in a continuous space-time, because that is where all the difficulties came from. If I have a continuum, I can have infinitely large wave numbers and infinitely large momenta which give rise to divergences. Also my metric is no longer well-defined if I want to measure distances between two very close points because this distance is no longer a well-defined concept.

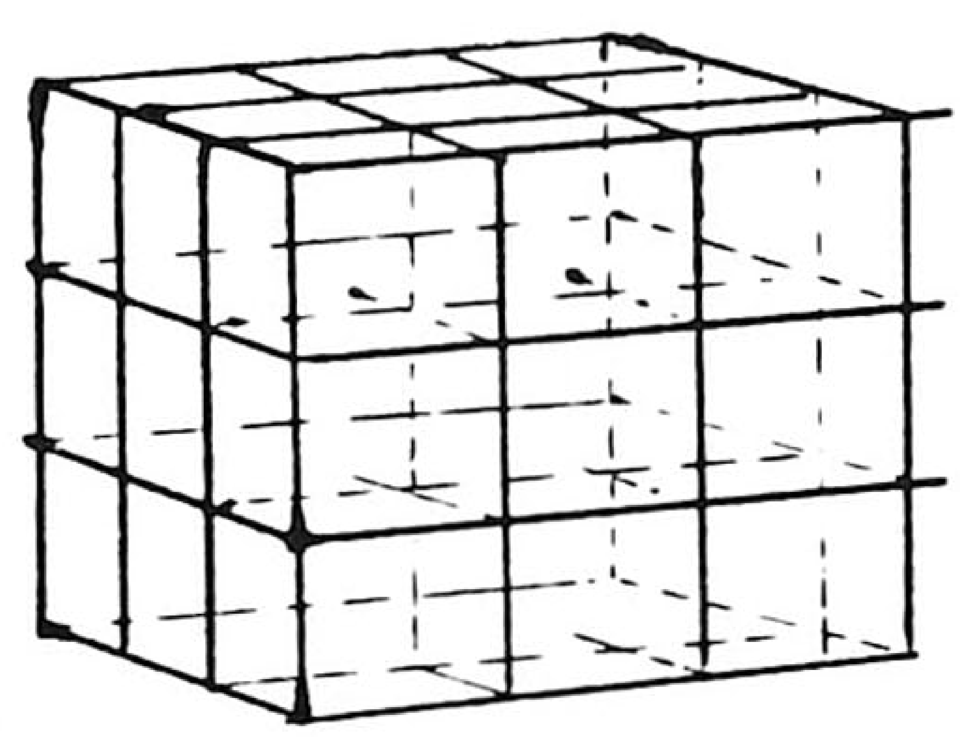
In particle theory we have learnt to be economical.If there is any possibility of reducing the number of physical degrees of freedom then we do. Thus, if a continuum makes no sense, then we replace it by a discrete space-time, called lattice. Lattices have been introduced before in physics. A famous lattice model for statistical mechanics is the exactly soluble Ising model (Fig. 7a) in two dimensions. Wilson [9] introduced a lattice version of gauge theories in four dimensions (fig. 7b). This lattice usually only serves to simplify the calculations and considerations, not because we really think that space-time in which quarks and gluons move is discrete.

In the case of gravity (Fig. 7c) we take the more ambitious point of view that it really does describe the physical situation accurately. The Planck length defines the distance scale on the lattice. However, there is a problem. In gravity the lattice which one would like to introduce would be totally different from the other lattices in field theory, and the reason is that we want to keep this beautiful notion of invariance under coordinate transformations. That would imply invariance under the interchange of two lattice points. That is, if there is interaction between two lattice points then there must be interaction between all other pairs of lattice points as well. (See Fig. 7c).

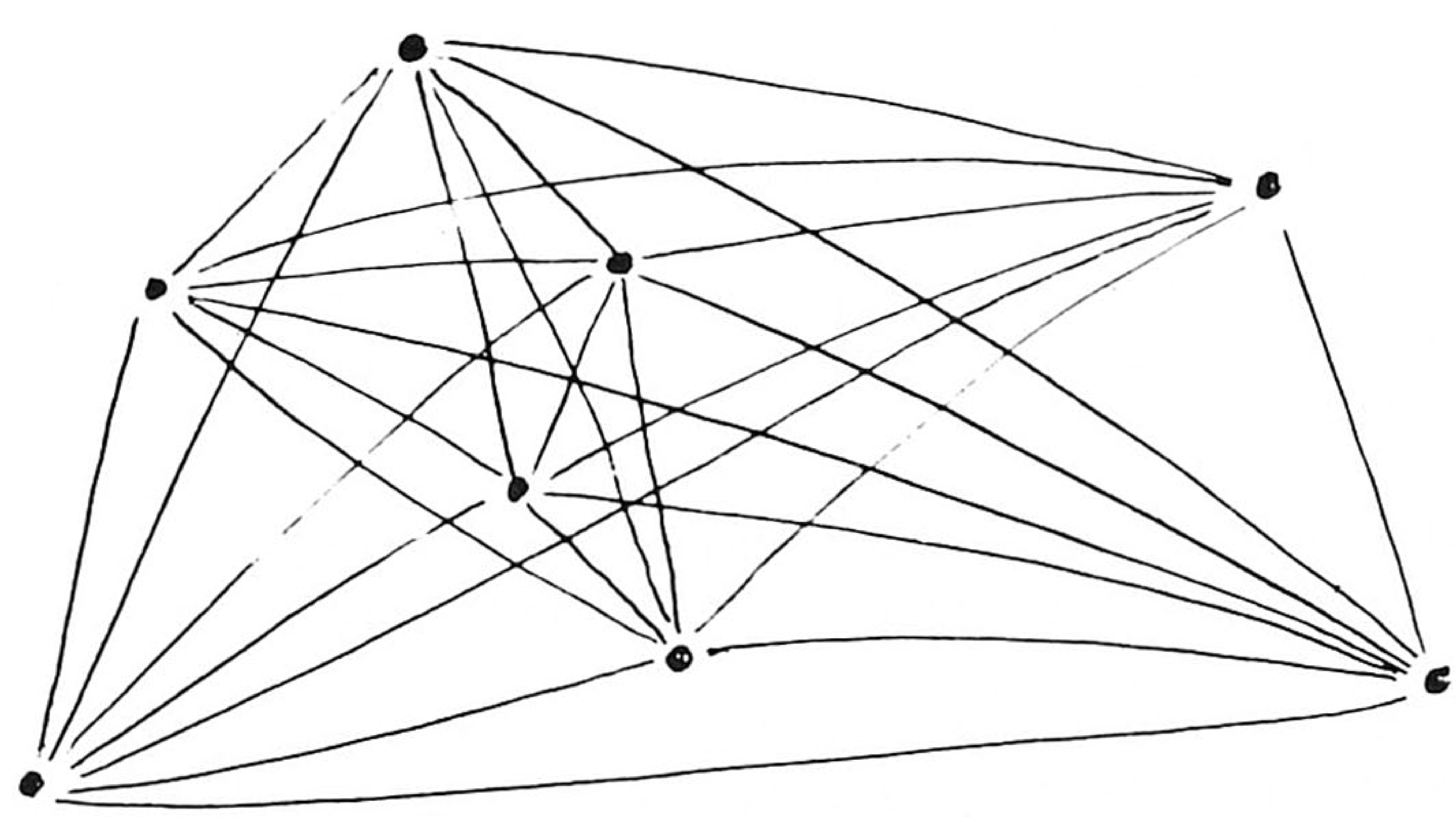
Is it, nevertheless, still possible to introduce some degree of order in this lattice ?



a) *Ising model.*



b) *Wilson's gauge theory.*



c) *Gravity on a lattice.*

Fig. 7

9. Causality on a lattice

I was asked a question when I discussed the ambiguities of the metric tensor in sects. 5 and 6. In general the causal ordering of points in Riemann space is defined by the metric. If one requires conservation of this causal ordering then only the conformal factor in front of the metric tensor is ambiguous. Evidently the redefinitions I described affect causal ordering also. This is a problem, I accept that. Indeed causality is a crucial principle in field theory. In general, physics will become pretty much impossible if one cannot tell whether one event is later or earlier than another. So let us impose a further structure on the lattice as follows. Even though a metric tensor does not make much sense on a lattice, we could at least define a causality relation for each pair of points (x,y). There must be three possibilities:

either $x < y$ (x is a point-event earlier than y)

or $x > y$ (x takes place later than y)

or $x \mid y$ (x and y are space-like). (29)

The latter possibility must be there if we wish to have some analogy to Lorentz invariance on the lattice. The existence of this relation between any pair of points must be dynamically determined. Of course one must require :

$$\text{if} \quad x < y \quad \text{and} \quad y < z \quad \text{then} \quad x < z \tag{30}$$

etc. We now formulate the following statements :

(i) This partial ordering itself again defines a lattice. Suppose that, in some sense to be defined later, a continuum limit exists, then

(ii) This lattice contains sufficient information to define a curved Riemann space in the continuum limit.

(iii) Vice versa, the existence of a continuum limit imposes restrictions on the details of the partial ordering relations (29) between all pairs of lattice points (not only partial ordering will enable us to make a continuum limit).

(iv) A curved Riemann space, if it has signature (+,+,+,-) contains all information on the physical history of the universe it describes.

10. The induced metric tensor

By statement (i) we mean the following. For any relatively time-like pair of points (x,y) with $x < y$ we can count the number of points z between them: $\rho(x,y)$ = number points z with $x < z < y$. If $\rho(x,y) = o$ then we join x and y with a line. In this way a lattice is obtained. We now assume that there is a continuum limit where this lattice is embedded in a four dimensional space. First we define the volume element in this space : $\sqrt{g}$ must be proportional to the number of lattice points inside a unit volume. To find the metric tensor induced in this continuum space we first define time-like distances. In continuous spaces the volume of all points z between two nearby timelike points x, y, with $x < z < y$, is easily found to be (see Fig. 8)

$$\rho(x,y) = \frac{\pi}{24}(dt)^4 .$$

We define now for the lattice

$$dt(x,y) = \sqrt[4]{\frac{24}{\pi}\rho(x,y) + 1} \tag{31}$$

where the +1 is needed for the case that $\rho = o$.

When $x|y$ the spacelike distance is harder to define. We simulate a formal measurement using light rays (see Fig. 9).

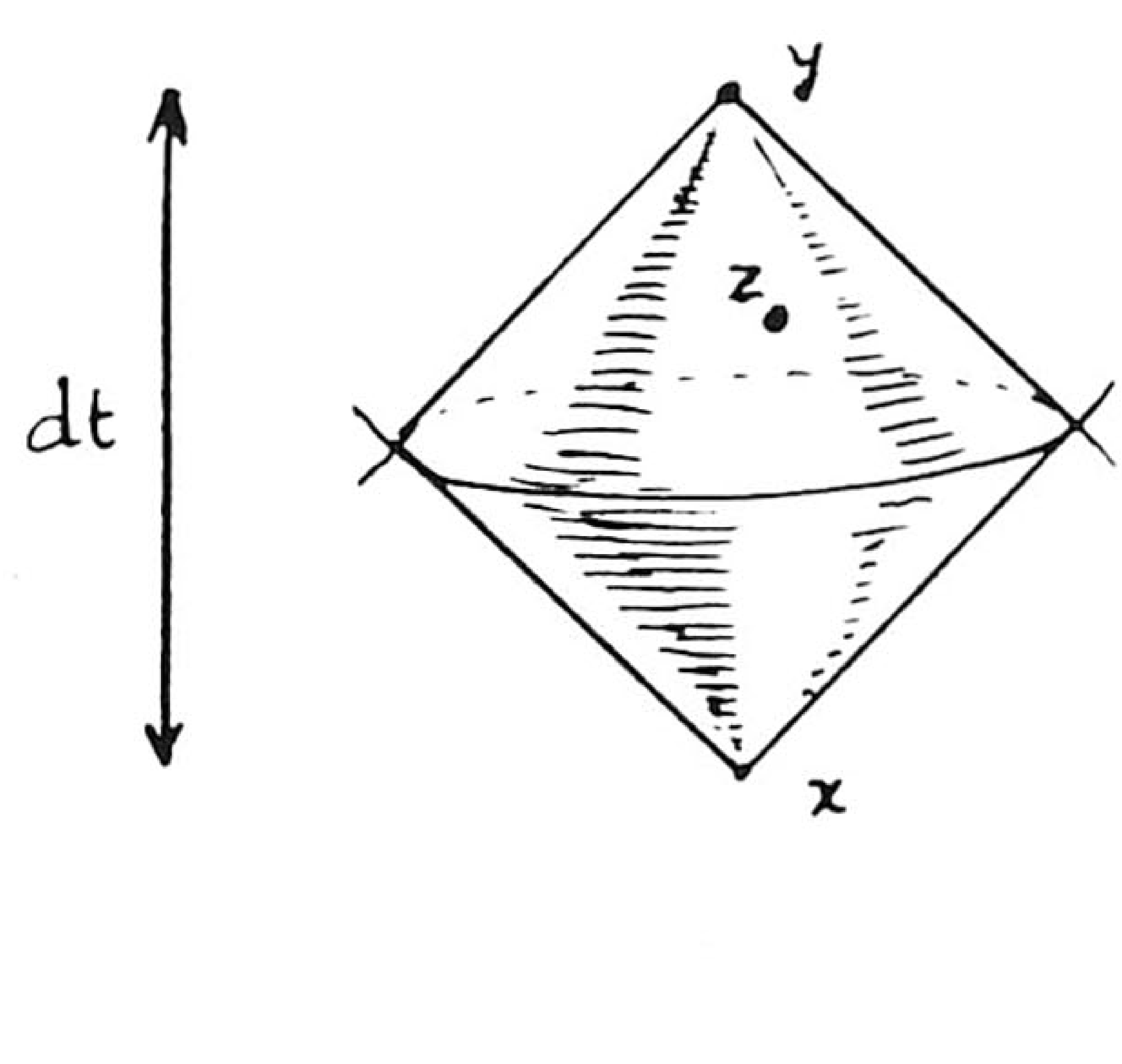


Fig.8 *Time-like distance between x and y (x<y) can be defined by counting the points z with x<z<y.*

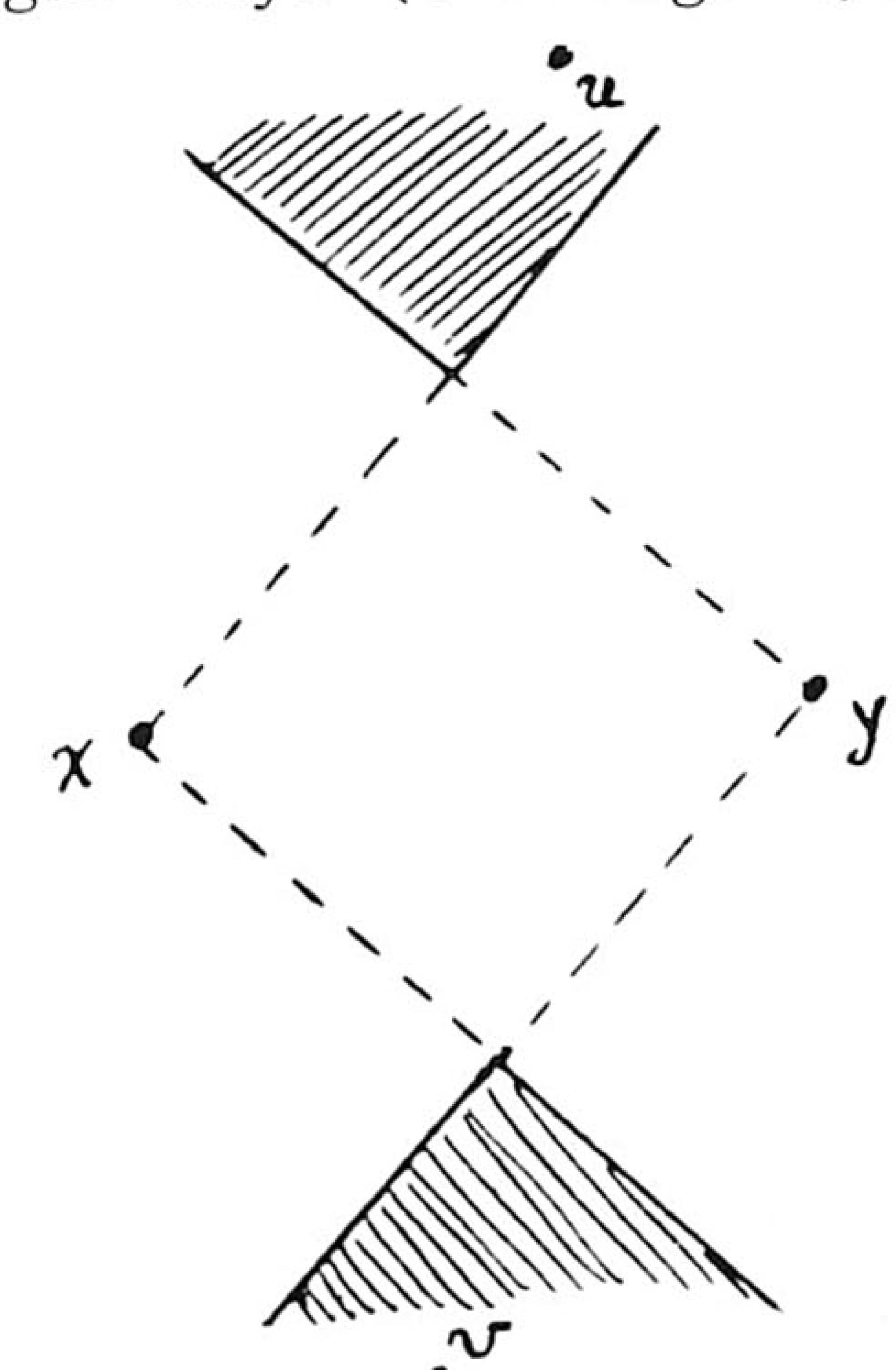


Fig.9 *Measuring spacelike distance for x|y.*
space of points u>x, u>y.
space of points v<x, v<y.

Consider all points u with $u > x$ and $u > y$.
Consider all points v with $v < x$ and $v < y$.

Then for all pairs (u,v) , $u > v$. An observer may move from v to u. He can emit a signal to both x and y and wait until both signals return. The minimal time he has to wait corresponds to the distance between x and y. Therefore,

$$ds(x,y) = \underset{u,v}{\text{minimum}}\ dt(u,v)$$

$$\text{for all } u,v \text{ satisfying } u > x,\ u > y,\ v < x,\ v < y\ . \tag{32}$$

Now that distances are well defined we can formulate more precisely the conditions for a continuum limit : the lattice must be embedded in a 4 dimensional continuous space such that a metric tensor $g_{\mu\nu}$ can be written down that describes to some degree of accuracy

$$\begin{aligned} & ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu && (\text{if } x|y) \\ \text{or} \quad & dt^2 = -g_{\mu\nu}\, dx^\mu\, dx^\nu && (\text{if } x > y) \end{aligned} \tag{33}$$

11. The Cosmic Tree

At this point we could make connections with cosmology. Is the universe infinite, or does it have a beginning or an end ? I feel that at least a beginning is desirable. We could add as an axiom to the ordering relations :

There is a point o with

$$o < x \quad \text{for all} \quad x \neq o.$$

(Note added : in discussions the point was raised that this seems to be in contradiction with present-day big bang theories ; these give an initial singularity consisting of an infinity of points that are not causally connected.)

With this axiom our lattice looks like a tree (Fig. 10).

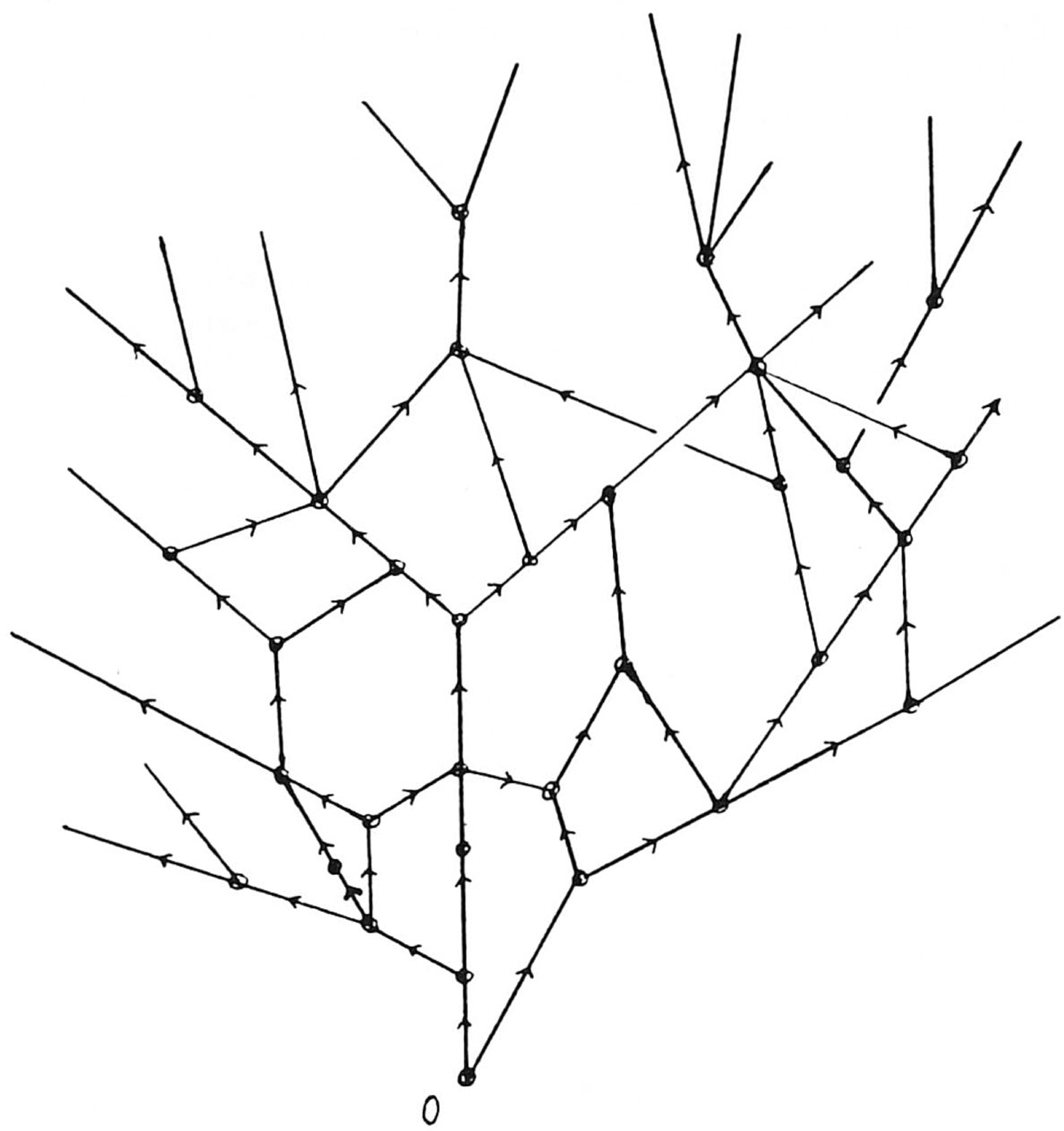


Fig. 10 *The Cosmic Tree.*

12. The Action ?

Perhaps we have here the seeds for a good theory. But we are not even half-way. Suppose that we do have the relevant variables here, in the sense that the details of the lattice are somehow determined by dynamics (the Einstein equations in the continuum limit). We still have to formulate this dynamics. That will be extremely hard. The functional integral formulation of field theories suggests that we have to sum the exponent of some action integral over all configurations. We have no idea how this action will look, but somehow it must approach the Einstein action in the continuum limit. There are two reasons why we expect such an action to be extremely non-local, however. One is that we insist on general coordinate invariance, implying that interactions between distant lattice points should be equal to that between close lattice points. The second is the absence of a cosmological term. We know that quantum fluctuations generate an energy-momentum tensor, even for the vacuum. The cancellation of this vacuum contribution by a "bare" cosmological term must be some mechanism that anticipates the (in itself non-local) quantum fluctuations. The question whether or not there is a vacuum at a point x can only be answered by looking at points surrounding x, even at fairly large distances. Thus the mechanism that renormalizes this cosmological constant must be non-local.

The above suggestions for a discrete gravity theory should not be taken for more than they are worth. The main message, and that is something I am certain of, is that it will not be sufficient to just improve our mathematical formalism of fields in a continuous Riemann space but that some more radical ideas are necessary and that totally new physics is to be expected in the region of the Planck length.

References

1) J. Schwinger, Proc. Nat. Acad. Sci. 44, 956 (1958).

K. Symanzik in Proc. Int. School "Enrico Fermi", R. Jost ed. Acad. Press (1969).

K. Osterwalder, R. Schrader, Comm. Math. Phys. 31, 83 (1973).

2) M. Veltman, Invited talk presented at the International Symposium on Electron and Photon Interactions at High Energies, Bonn 27-31 August 1973.

3) C.N. Yang and R. Mills, Phys. Rev. 96, 191 (1954).

R. Feynman, Acta Phys. Polonica 24, 697 (1963).

4) G. 't Hooft and M. Veltman, Nucl. Phys. B50, 318 (1972).

5) J.D. Bjorken and S.D. Drell, Relativistic Quantum Fields, McGraw-Hill (1965).

G. 't Hooft and M. Veltman, "Diagrammar", CERN report 73/9 (1973).

6) G. 't Hooft and M. Veltman, Nucl. Phys. B44, 189 (1972).

C.G. Bollini and J.J. Giambiagi, Phys. Lett. 40B, 566 (1972).

J.F. Ashmore, Lett. Nuovo Cim. 4, 289 (1972).

7) G. 't Hooft and M. Veltman, Ann. Inst. Henri Poincaré, 20, 69 (1974).

8) A.A. Slavnov, Teor i Mat. Fizika, 10, 153 (1972).

J.C. Taylor, Nucl. Phys. B33, 436 (1971).

9) K.G. Wilson, Phys. Rev. D10, 2245 (1974).

K.G. Wilson, Lectures given at the 1975 Intern. Summer School of Physics "Ettore Majorana".