ITP-UU-08/78
SPIN-08/60

ERICE LECTURES 2008

Gerard 't Hooft

Institute for Theoretical Physics
Utrecht University

and

Spinoza Institute
Postbox 80.195
3508 TD Utrecht, the Netherlands

e-mail: g.thooft@uu.nl
internet: http://www.phys.uu.nl/~thooft/

1. Opening Lecture. Remembering Sidney Coleman
2. Some Effects of Instantons in QCD
3. Crystalline Gravity

December 2008

1 Opening Lecture: REMEMBERING SIDNEY COLEMAN

The Amsterdam International EPS Conference[1], June 30 - July 6, 1971, was for me the first occasion to present my theories about the renormalizability of non-Abelian gauge theories with spontaneous symmetry breakdown. My then advisor, Martinus Veltman, had given me the opportunity to explain this in 10 minutes to the audience. "Let me introduce you to two American gangsters", he said to me, after the lecture was over. I was puzzled by the remark at that time, but later realized that this was Veltmans way to talk about his adversaries in physics. Anyway, the two gentlemen in question turned out to be Sheldon Glashow and Sidney Coleman. Talking to these two people, made me realize that these must be the smartest gangsters on the planet. Sidney immediately saw the most essential ingredients of the new theory. He later explained that Quantum Field Theory had looked to him as an ugly monster until exactly that moment, when it received a kiss by our work, turning into a beautiful and wealthy prince.

I met Sidney many times since, and I vividly remember the lectures he gave in the 1975 Erice School[2]. At that school, not only the best student and best secretary were chosen, but also best lecturer. This was Sidney Coleman, who, at the farewell party, also came as the most fancy dressed participant, in his fluorescent pink jacket. His lectures had been about topology in QFT. Previously, when renormalizing the theory, we had always limited ourselves to perturbation expansion, where the topology is trivial.

Sidney would illustrate vividly what topology is, using the cord of his microphone, by which he was connected to the loudspeaker system. Winding the cord around his neck he explained the concept of winding. The students, worried as they were that their lecturer might strangle himself, would never forget the definition of winding number.

It was October 1988 when, during one of my later encounters at Harvard, I visited Sidney at his home, with my two little children. There, we would play the game of Charades; he was very good at it. My children loved it.

As emphasized by others here, Sidney was a great story teller: One of his stories was about how he rescued mankind from a pending planetary disaster, while sitting at a bar, with his friend Carl Sagan. With Sagan, Sidney shared his love for science-fiction. Behind a good glass of whiskey, or two, Sagan showed a problem to Sidney: Suppose that there exist alien, totally unknown, life forms on Mars, or, for that matter, on the Moon as well. The Moon would soon be visited by astronauts. What should space agencies do to avoid the danger that these life forms would infect the Earth, and destroy mankind? How do we avoid contamination of one planet by another in general, when space flight becomes routine?

This became a vivid discussion. But a month later, Sidney had almost forgotten this conversation, so that he was highly surprised when a manuscript fell in his mailbox: "Spacecraft Sterilization Standards and Contamination of Mars", by S. Coleman and C. Sagan. The norm for the possibility of cross contamination from one celestial body to another should be less than 0.1%. Further, astronauts should be put in quarantine, and the chances of survival for any individual organism of any life form from one planet to any other should be less than one in 10,000.

The paper was submitted for publication[3]. Now, this was the only existing publication on this topic, so, NASA decided to follow the advice. The Committee of Space Research, COSPAR, determined that astronauts returning from the Moon would have to go into quarantine for a certain amount of time. The "Lunar Receiving Laboratory" (LRL) was erected in Houston, Texas. Upon their return from the Moon, many astronauts spent boring hours there. Fortunately, dangerous life forms were never found.

2 Lecture II: SOME EFFECTS OF INSTANTONS IN QCD

2.1 Construction of instantons

Consider a non-Abelian Yang-Mills gauge theory in 4 space-time dimensions, and the group or subgroup $SU(2)$ of gauge transformations. They may be space-time dependent, so we have a mapping from $\mathbb{R}4$ to the space Ω of $SU(2)$ transformations, to be indicated by the function

$$U(\vec{x},t)=a_0(\vec{x},t)\,\mathbf{1}+i\sum_{i=1}^{3}a_i(\vec{x},t)\,\sigma_i\ , \tag{2.1}$$

where σ_i are the Pauli matrices and the coefficients $a_0,\cdots,\ a_3$ are real and obey

$$\sum_{i=0}^{3}a_i^2=1\ . \tag{2.2}$$

Clearly, these coefficients (2.2) lie on a 3-sphere, $S3$.

Now the boundary of $\mathbb{R}4$ is also a 3-sphere, so suppose that we choose our mapping in such a way that the 3-sphere of the boundary is a 1-to-1 mapping onto the gauge-sphere (2.2). The mapping that we chose, as well as small deformations of it, is said to have winding number one, and is indicated as $U_1(\vec{x},t)$. We also have mappings of winding number n, by raising U_1 to the $n^{\rm th}$ power:

$$U_n(\vec{x},t)=U_1(\vec{x},t)^n\ , \tag{2.3}$$

where n can be any integer.

Only if $n=0$, however, this mapping can be extended to all of $\mathbb{R}4$ without any singularity anywhere. Such singularities are normally not admitted in a gauge transformation. When $n\neq 0$, our gauge transformation can only be well-defined on a space $\mathbb{R}4$ with a small region V cut out. The singularity is then assumed to lie inside V.

Suppose we apply this gauge transformation $U_1(\vec{x},t)$ to the vacuum field configuration that was defined as $A_\mu(\vec{x},t) = 0$. The transformed field,

$$A_\mu^{(1)}(\vec{x},t) = -\frac{i}{g}(\partial_\mu U_1) \cdot U_1^{-1} \ , \tag{2.4}$$

is non-vanishing now, but of course, outside the region V, it still describes the vacuum state.

Now the gauge function $U(\vec{x},t)$ could not be extended inside V, but there is no problem if one wishes to extend the field configuration (2.4) to also obtain values inside V. If V would be the sphere $\vec{x}^2 + t^2 \leq \varrho$, then we could take

$$A_\mu^{(1)}(\vec{x},t) = \frac{|(\vec{x},t)|}{\varrho} A_\mu^{(1)} \left(\frac{(\vec{x},t)}{|(\vec{x},t)|} \right) \ . \tag{2.5}$$

to hold inside V.

This is smooth and singularity-free everywhere inside V (mild singularities in the derivatives of A can be smeared away if so desired). However, inside V, this field $A_\mu^{(1)}$ does not describe the vacuum, since no function U can be found that gauge transforms it away. Indeed, one can compute the associated curvature fields $G_{\mu\nu}(\vec{x},t) = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]$ to find that these do not vanish inside V. This is therefore a non-trivial field configuration, and since it only differs from the vacuum inside a region V that surrounds a point as well as an instant in time[4], it is called an *instanton*. If the winding number n is one, we call it a single instanton, if it is -1, we have an anti-instanton. For values of n greater than 1 (or lower than -1) we have multi-instantons.

2.2 Instantons and fermions

Instantons can exist either in Minkowski spacetime or in Euclidean spacetime, since no reference was made to any Lorentz invariant metric. Instantons are time-dependent, therefore fields such as fermion (quark) fields, can have modes where the energy before and after an instanton is different. In the case of massless fermions, chirality (or, equivalently, the helicity) of the fermion is preserved, since the gauge field interaction (which goes as $ig\gamma_4\gamma_\mu A_\mu(\vec{x},t)$) commutes with γ_5. Thus, left-handed fermions stay left-handed and right-handed ones stay right-handed. In the case of a single instanton, however, exactly one eigen mode of the fermionic Hamiltonian, for each chiral fermion flavor of positive helicity, is lifted from the Dirac sea to a positive energy state, see Fig. 1

The figure explains that this leads to one solution of the Dirac equation,

$$\gamma_\mu(\partial_\mu - igA_\mu)\psi = 0 \ , \tag{2.6}$$

in the presence of an instanton that rapidly approaches zero in all four Euclidean directions. [5][6]

Under a parity transformation,

$$\vec{x} \to -\vec{x} \ , \qquad t \to t \ , \tag{2.7}$$

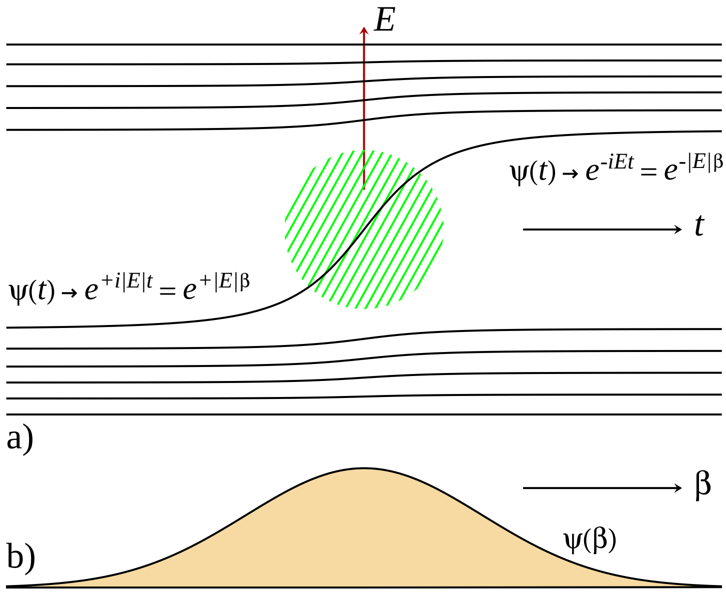

Figure 1: Chiral fermion levels in the presence of an instanton, *a*) as a function of Minkowski time, *b*) After a Wick rotation $t = -i\beta$, we see this Euclidean time dependence of this chiral fermionic mode.

the instanton transforms into an anti-instanton, since

$$(a_0\mathbf{1} + ia_i\sigma_i)^{-1} = a_0\mathbf{1} - ia_i\sigma_i \ , \qquad \text{if} \qquad \sum_{i=0}^{3} a_i^2 = 1 \ . \tag{2.8}$$

Therefore, left-helicity fermions do the opposite of right-helicity fermions: if a left-helicity fermion is created out of the Dirac sea, a right-handed helicity fermion disappears, and *vice versa*.

Since holes in the Dirac sea describe anti-fermions, one can extend the statement to include the possibility of the creation of a chiral antifermion in stead of the annihilation of a chiral fermion, and so on.

An instanton acts on all fermionic flavors the same way; therefore, an instanton effectively acts as a vertex where one unit of positive chirality of each flavor disappears and one unit of negative chirality appears. This vertex can be mimicked by an extra term in the Lagrangian of the theory of the following form:

$$\mathcal{L}^{\text{inst}} = \kappa \det_{a,p}(\overline{\psi}_a^R \psi_p^L) + h.c. \ , \tag{2.9}$$

where an inner product is taken in color space, so that the interaction is color-gauge invariant, and $h.c.$ stands for the anti-instanton's contribution, which is the hermitean conjugate of that of the instanton. a and p are flavor indices. In the case of three flavors, this is written explicitly as

$$\tfrac{1}{3!}\kappa\, \varepsilon_{abc}\varepsilon_{pqr}(\overline{\psi}_a^R \psi_p^L)(\overline{\psi}_b^R \psi_q^L)(\overline{\psi}_c^R \psi_r^L) \ + \ h.c. \ . \tag{2.10}$$

In QCD, this interaction has important effects both for the pseudoscalar mesons and for the scalar meson spectrum.

2.3 Pseudoscalar mesons

In the case of a single chiral fermion flavor, the effective instanton action (2.9) acts just as a mass term. This mass term violates the chiral symmetry

$$\psi^L \to U_L \psi^L \ , \qquad \psi^R \to U_R \psi^R \ , \tag{2.11}$$

if the complex numbers U_L and U_R are chosen to be independent. Global vector gauge invariance, corresponding to $U_L = U_R$ remains a good symmetry. This is exactly the result that one also obtains in perturbation theory, where the Adler-Bell-Jackiw anomaly[7] is found to break the symmetry in exactly the same way. This is not an accident[8].

In the case of more than one flavor, U_R and U_L are complex unitary matrices, and now one only needs to demand that

$$\det(U_R) = \det(U_L) \ . \tag{2.12}$$

This breaks the chiral symmetry

$$U(N_F)_L \otimes U(N_F)_R \tag{2.13}$$

into

$$SU(N_F)_L \otimes SU(N_F)_R \otimes U(1) \ , \tag{2.14}$$

again in agreement with the Adler-Bell-Jackiw anomaly in the perturbative theory.

Long before the advent of QCD, it was noted that the pseudoscalar particles π and K appear to act as the Goldstone bosons of the symmetry (2.14), which seems to be well obeyed by the interactions of these particles, while the symmetry (2.13) would require one more scalar meson, the η particle (in case of isospin $SU(2)$), or the η' (in the case of flavor $SU(3)$) to act as a Goldstone boson. But Goldstone bosons should be massless. The pion is much lighter than the η particle, and the kaons are much lighter than the η' (also sometimes called the *Beppo*[9]). We now understand that the excess masses of η and of η' are both due to the instantons in QCD.

The fact that the interaction (2.10) for three flavors in $SU(3)$ turns into a four-fermion interaction for isospin, $SU(2)$, when the strange quark is assumed to be heavy, can be understood: the strange mass term explicitly mixes the chiral left handed strange quark with the righthanded one, so that $\langle \overline{\psi}_s^R \, \psi_s^L \rangle$ (where s stands for strangeness) can be seen to be proportional to the strange quark mass instead of a dynamical (pseudo)scalar field. In Figure 2 it is illustrated how the mass term for the strange quark s can turn an effective 6-quark vertex into a 4-quark vertex.

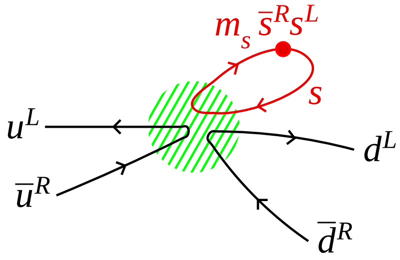

Figure 2: The strange quark s can form a closed loop in the effective quark vertex, an effect that scales with the strange quark mass m, since the left helicity strange quark must turn into a right-helicity one (or vice versa).

2.4 Scalar mesons

The effective instanton action, Eq. (2.9) or (2.10), not only causes shifts in the spectrum of pseusoscalar mesons; it also contributes in the scalar sector. At first sight, it seems to be more difficult to identify its effects there, since, in this channel, no violation takes place of any explicit symmetry pattern. There is, however, a more subtle way in which it manifests itself.

Effective QCD models showed a peculiar difficulty pertinent to the scalar mesons. The scalar mesons, like the pseudovectors, form $SU(3)$ octets or $U(3)$ nonets. One of the scalar spin zero states has sizeable branching ratios in the K - $\overline{K}$ sector, so that one would be tempted to describe as an $\bar{s}\, s$ bound state. However, it is about as light as the other scalar spin zero states that do not decay into strange mesons. The s quark is heavier than the u and d quarks. How do we explain this?

One possibility investigated in the literature is that the scalar mesons are systems containing two quarks and two antiquarks, the so-called *tetraquarks*. Realizing that a color-antisymmetric two-quark system transforms as a $\overline{\mathbf{3}}$ representation of color $SU(3)$, a teraquark is realized if, in a baryon, one quark is replaced by two antiquarks. Therefore, one might imagine that tetraquarks will be about as heavy as baryons are. If the scalar mesons were tetraquarks, their mass spectrum would look a bit better.

This, however, would still raise the question where the ordinary scalar biquark states are. Why would tetraquark states be more important than biquark ones? It is here that the instanton interaction (2.10) sheds new light. It couples tetraquark states such as $\overline{u}\overline{d}ud$ directly to the state $\overline{s}s$. The masses of these two states could be close together, and, as is well known from perturbation theories, when two states are close together, a relatively weak mixing term can give a relatively large splitting. Thus, one new mixed state arises that is relatively light. Biquark states such as $\overline{u}s$ and $\overline{u}d$ will be further away from the tetraquark states with which they are mixed, and so their mixing angles will be smaller and their masses are reduced by smaller amounts. An effective theory of scalar meson states can be constructed along these lines, and it appears that the instanton interaction (2.10) is just what is needed to obtain agreement with the observations.[10]

3 Lecture III:
CRYSTALLINE GRAVITY

Abstract

Matter interacting classically with gravity in 3+1 dimensions usually gives rise to a continuum of degrees of freedom, so that, in any attempt to quantize the theory, ultraviolet divergences are nearly inevitable. Here, we investigate matter of a form that only displays a finite number of degrees of freedom in compact sections of space-time. In finite domains, one has only exact, analytic solutions. This is achieved by limiting ourselves to straight pieces of string, surrounded by locally flat sections of space-time. Next, we suggest replacing in the string holonomy group, the Lorentz group by a discrete subgroup, which turns spacetime into a 4-dimensional crystal with defects.

3.1 Introduction

One way of addressing the problem of quantizing gravity is first to search for its fundamental degrees of freedom. In all quantum systems that we know of, these degrees of freedom are denumerable, at least if one confines oneself to the data inside a box of finite dimensions. In gravity, one expects even more stringent constraints: according to the holographic principle, one expects the total number of degrees of freedom that reside inside a finite box not to exceed the number of microstates of a black hole that would just fit in there — in other words, the total number of degrees of freedom is finite!

Now imagine splitting the box into more and more finite subsystems. If space and time were to be described by coordinates consisting of real numbers, then, at least mathematically, splitting boxes into smaller ones can be done without limits. What this appears to imply then, is that at distance scales small compared to the Planck length, roughly 10^{-33} cm, there should be no degrees of freedom left at all. One way to accommodate for such a situation is to describe space and time there as to be absolutely flat. there is no physically non-trivial structure left. The most one can allow for is some defects when gluing together flat pieces of space-time at their boundaries. In other words: at the Planck scale, gravity must be replaced by some *topological* theory.

"Simplicial gravity", as described by R. Loll in her lectures at this school, is an example of just such a structure, but it is not quite the same as the theory to be discussed now. As will gradually become clear, "gravitational" and "matter" degrees of freedom will be indistinguishable, both being described as *defects* when gluing locally flat pieces of Minkowski space-time together. We assume the pieces not to have a cosmological constant, so, the bare cosmological constant $\Lambda = 0$.

The most conspicuous difference between the theory to be described here[11] and others is that, in essence, it is a classical theory. Its degrees of freedom, the defects, obey classical equations of motion. One can think of this in two ways: the most conventional attitude to take is to regard the theory as a precursor to a quantized version, as was done with many other theories, but we also have in mind a more radical idea, pursued by this lecturer for some time[12], which is called “prequantization”, the idea that all states of the classical theory can directly be promoted to elements of a basis of Hilbert space, without first replacing Poisson brackets by Dirac brackets. We leave these options open for the time being.

The question first to address is then: what are the dynamical equations of motion for the defects? We have a kind of crystal in Minkowski space. The crystal has defects in it. Ordinary crystals only live in 3 dimensional Euclidean space, and their defects can be anything. The defects in our theory are 4 dimensional, and obey e.o.m. In the next chapter, we illustrate the idea by first studying gravity in only two spacelike dimensions.

3.2 Gravity in $2+1$ dimensions

The importance of the model of gravitating point particles in a locally flat 2+1 dimensional space-time, is still severely underestimated[13]—[15]. In 2+1 dimensions, pure gravity (gravity without matter in some small section of space-time) has no physical degrees of freedom at all (apart from boundary effects if spacetime is topologically non-trivial). This is because the Riemann curvature $R_{\alpha\beta\gamma\delta}$ can be rewritten in terms of a symmetric 3×3 matrix $Q^{\mu\nu}$ as follows:

$$R_{\alpha\beta\gamma\delta}=\varepsilon_{\alpha\beta\mu}\,\varepsilon_{\gamma\delta\nu}\,Q^{\mu\nu}\ , \tag{3.1}$$

so that the Ricci curvature is

$$R_{\alpha\gamma}={R_{\alpha\beta\gamma}}^{\beta}=(g_{\mu\nu}g_{\alpha\gamma}-g_{\mu\gamma}g_{\alpha\nu})Q^{\mu\nu}=Q^{\mu}_{\mu}\,g_{\alpha\gamma}-Q_{\gamma\alpha}\ ; \tag{3.2}$$

$$\text{so that}\qquad Q_{\mu\nu}=\tfrac{1}{2}R^{\alpha}_{\alpha}\,g_{\mu\nu}-R_{\mu\nu}\ . \tag{3.3}$$

Clearly, if matter is absent, $R_{\mu\nu}$ vanishes, and therefore so do $Q_{\mu\nu}$ and $R_{\alpha\beta\gamma\delta}$. Conversely, a point particle represents point curvature. Thus, particles are point singularities surrounded by flat space-time. A particle at rest can be described as in Figure 3 a). The wedge is stitched closed, so that the points A and A' are identified. The defect angle can

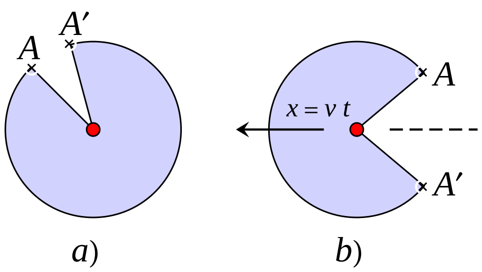

Figure 3: a) 2 dimensional space surrounding a point particle. Points A and A' are identified. b) Particle moving in the x-direction, after applying a Lorentz boost. Only if the defect is arranged in the way indicated, the identification of points A and A' is without a time shift.

directly be identified with the particle's rest mass. One can show that systems containing several point particles in motion, while the total momentum is kept zero, is surrounded by a conical space-time, of which the deficit angle can be identified with the total energy.

When a point particle is set in motion, the surrounding space-time is described by performing Lorentz transformations upon the stationary case, see Fig. 3 *b*) . One then can study systems with N particles, by viewing space-time as a tessellation of locally flat triangles or polygons.[14] If all particles in a $2+1$ dimensional universe would be stationary or nearly stationary, the two-dimensional integrated scalar Ricci tensor, $\int \delta^2 x \sqrt{g}\, R$, would have to be positive, so with a sufficient number of particles the spacelike part of this universe would always close into an $S(2)$ geometry. Its timelike coordinate can form a compact dimension (featuring both a *bang* and a *crunch*), or a semi-infinite one, with either a bang or a crunch. With fast moving particles one can however also form 2-d surfaces with higher genus, without requiring negative mass particles.[15]

Quantization of this system is often carried out 'as usual',[16] but there are delicate problems, having to do with the fact that we are dealing with a strictly finite universe, so that the role of an 'observer' is questionable, and the statistical interpretation of the wave function is dubious because the finiteness of the universe prohibits infinite sequences of experiments to which a statistical analysis would apply. Carrying quantization out with care, one first observes that evidently, time is quantized into 'Planck time' units[17]. This is easily derived from the fact that the hamiltonian is an angle and it is bounded to the unit circle. Consequently, a quantum theory cannot be formulated using differential equations in time, but rather one should use evolution operators that bridge integral time segments. This indeed can be regarded as a first indication of some sort of space-time discreteness, which we will encounter later in a more concrete way. Moreover, a confrontation with foundational aspects of quantum mechanics will appear to be inevitable.

3.3 Three space dimensions: strings and joints

In the work presented here, the question is asked whether a similar "finite" theory can also be formulated in $3+1$ dimensions. This is far from obvious. One first notices that the absence of matter now no longer guarantees local flatness, since the Ricci curvature $R_{\mu\nu}$ can vanish without the total Riemann curvature $R^{\alpha}{}_{\beta\mu\nu}$ being zero. However, one still can decide to view space-time as a tessellation of locally flat pieces. The defects in such a construction again may represent not only matter, but also physical activities such as gravitational waves. The primary defects one finds are direct generalizations of the $2+1$ dimensional case. Take a particle-like defect in 2-space. In 3+1 dimensions, such defects manifest themselves as strings.

It could be that matter is always arranged in such a way that it can be regarded as defects in a locally perfectly flat space-time. As opposed to technical approaches towards solving General Relativity[18], we now regard matter of this form as elementary. The novelty in this idea is that, in spite of matter being distributed on subspaces of measure zero, we still insist that it obeys local laws of causal behavior. Let us see how this looks.

By simply adding the third space dimension as a spectator, orthogonal to the first two,

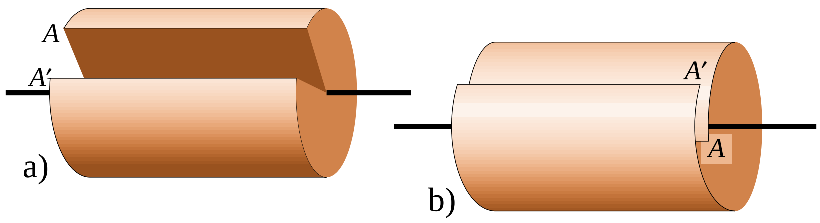

Figure 4: a) 3-space surrounding a positive string, with deficit angle $A\,A'$; b) 3-space surrounding a negative string, showing a surplus angle $A\,A'$.

we find flat 3 dimensional space-time surrounding a string. Indeed, the energy momentum tensor of a straight, infinite, static string pointing in the z-direction, is

$$T_{\mu\nu}=t_{\mu\nu}\,\delta^2(\tilde{x})\ ,\qquad t_{33}=-t_{00}=\varrho\ ,\qquad \text{all other}\ \ t_{\alpha\beta}=0\ ,\tag{3.4}$$

where ϱ is the string tension parameter.

Space-time surrounding a string is then seen to be as sketched in Fig. 4 a . When the string constant ϱ is large, we *redefine* it to be the one generating exactly the deficit angle $\alpha=8\pi G\varrho$. Note that a *positive* string constant leads to a *deficit* angle. A negative string constant would produce a *surplus* angle, see Fig. 4 b . Normally, in string theory, a negative string constant would make a string highly unstable, as it contains positive pressure and negative energy. In our case, however, strings are constrained to form straight lines, and therefore this instability has no effect at small distances. We later may wish to include such negative strings in our models. We leave this open for the time being, but unless indicated specifically, we will usually be discussing strings with positive string constants and thus with positive deficit angles.

In addition to pure, straight stringlike defects, we will need pointlike defects where three or more strings join. If we take the point $\vec{x}$ where the strings join to be standing still, then all strings joining there will be standing still as well. Each string, (i), has a deficit angle that represents a rotation $\Omega_i=e^{2i\vec{\omega}_i\cdot L_i}$, where L_i are the generators of rotations in 3D, and $\omega_i=|\omega_i|$ are half the deficit angles. The condition that the strings match at a joint is that surrounding spacetime has no other deficits, so that

$$\Omega_1\,\Omega_2\,\Omega_3=\mathbb{I}\ .\tag{3.5}$$

3.4 Moving strings

A description of a multitude of static strings is now straightforward, in principle. However, already here, one may expect considerable complications. The system of static strings is not unlike the 2 + 1 dimensional universe with moving point particles, where the time coordinate is replaced by the coordinate z of the static string system. We know that the 2 + 1 dimensional world has either a big bang singularity or a big crunch[14]. Similar "infrared" singularities might show up in the static string system. This question will not be further pursued here. It is the local properties of the model that we will investigate further.

Moving strings impose important questions concerning the internal consistency of the model. A moving string can be characterized in different ways. Firstly, one can specify the orientation vector $\vec{\omega}$ of the string, normalized such that its norm coincides with half the deficit angle, $\frac{1}{2}\alpha = 4\pi G\varrho$ when the string is at rest:

$$|\vec{\omega}| = \tfrac{1}{2}\alpha \ . \tag{3.6}$$

In addition then we specify the velocity vector $\vec{v}$ of the string. But, noticing that the string is invariant under boosts in the direction $\vec{\omega}$, only the component of $\vec{v}$ orthogonal to $\vec{\omega}$ matters, so we limit ourselves to the case

$$\vec{v}\cdot\vec{\omega} = 0 \ . \tag{3.7}$$

Finally, the position of the string at $t = 0$ should be specified. This requires another vector orthogonal to $\vec{\omega}$. All in all, we see that $3+2+2=7$ real parameters are needed, of which 5 are translationally invariant, and 2 can be set to zero by a spacelike translation in 3-space.

Alternatively, we can specify the string's characteristics by giving the element of the Poincaré group that describes the holonomy along a non-contractible cycle C around the string. For static strings through the origin of 3-space, this is just the pure rotation operator, which will be denoted as $U(\vec{\omega})$. For strings moving with velocity $\vec{v}$ through the origin, this is the element $B(\vec{v})\,U(\vec{\omega})\,B(-\vec{v})$ of the Lorentz group, where $B(\vec{v})$ is the element of $SO(3,1)$ that represents a pure Lorentz boost corresponding to the velocity $\vec{v}$. If the string does not move through the origin, we get a more general element of the Poincaré group.

Notice however, that an arbitrary element of the Lorentz group is specified by 6 parameters, not 5, and the elements of the Poincaré group by 10 parameters, not 7. This means that not all elements of the Poincaré group describe the holonomy of a string. Firstly, ignoring the translational part, the pure Lorentz transformation Q associated to the closed curve C has to obey one constraint:

There must be a Lorentz frame such that, in that frame, Q is a pure rotation in 3-space.

Since we plan to describe these Lorentz transformations in terms of their representations in $SL(2,C)$, we identify this as a constraint on the associated $SL(2,C)$ matrices. Write

$$Q = B(\vec{v})\,U(\vec{\omega})\,B(-\vec{v}) \ , \tag{3.8}$$

where $B(\vec{v})$ are 2×2 matrices representing boosts with velocity $\vec{v}$ (we'll see shortly that pure boosts are represented by *hermitean* 2×2 matrices), and U is a unitary matrix representing a pure rotation. When $\vec{\omega}$ points in the z direction, we have

$$U(\vec{\omega}) = \begin{pmatrix} e^{i\omega} & 0 \\ 0 & e^{-i\omega} \end{pmatrix} \ , \tag{3.9}$$

so that

$$\mathrm{Tr}\,(U) = 2\cos\omega\ . \qquad (3.10)$$

Since the trace is invariant under rotations and the boosts (3.8), it follows quite generally that

$$\mathrm{Im}(\mathrm{Tr}\,(Q)) \;=\; 0\ ; \qquad (3.11)$$
$$|\mathrm{Re}(\mathrm{Tr}\,(Q))| \;\le\; 2\ . \qquad (3.12)$$

Eq. (3.11) fixes one of the real variables of the Lorentz transformation Q. Inequality (3.12) is important in a different way. A generic $SL(2,C)$ matrix can de written in a basis where it is diagonal. Because the determinant is restricted to be 1, the diagonal form is then

$$Q = \begin{pmatrix} z & 0 \\ 0 & 1/z \end{pmatrix}\ , \qquad (3.13)$$

where z can be any complex number. Imposing (3.11) leaves two options: either z is on the unit circle – in which case it represents a pure rotation in 3-space, or it is a positive or negative real number. In the latter case, Q is a pure Lorentz boost, and this is when Ineq. (3.12) is violated. It describes the holonomy of something that is quite different from a string.

An other restriction to be imposed on the holonomy of a physical string concerns the translational part of the element of the Poincaré group, which we ignore for the time being.

3.5 Collisions

When we were dealing with point particles, in the $2+1$ dimensional case, we could safely assume that the particles will never collide head-on. In general, they will miss one another, and consequently no further dynamical rules are needed to determine how an N particle system will evolve. This will not be true in higher dimensional spaces.[1] Strings in $3+1$ dimensional space-time will in general not be able to avoid one another. They will cross, and in doing so, two straight string sections will not be straight anymore after the collision.

Consider an initial state in which two strings are heading towards one another. We can always work in a Lorentz frame where one of the strings, call it A, is at rest. The conical 3-space surrounding it has a deficit angle $\alpha = 2\omega_A$. In the generic case, in this Lorentz frame, the second string does not have to be oriented orthogonally to the first one. Its string constant, β, does not have to be the same as α. Without loss of generality, we may assume that the velocity vector $\vec{v}$ of the second string is orthogonal to both A and B.

[1]Strings will in general not collide head-on in a space-time of more than 4 dimensions. However, the generalizations of the objects we discuss in this paper, in higher dimensions will be branes, not strings.

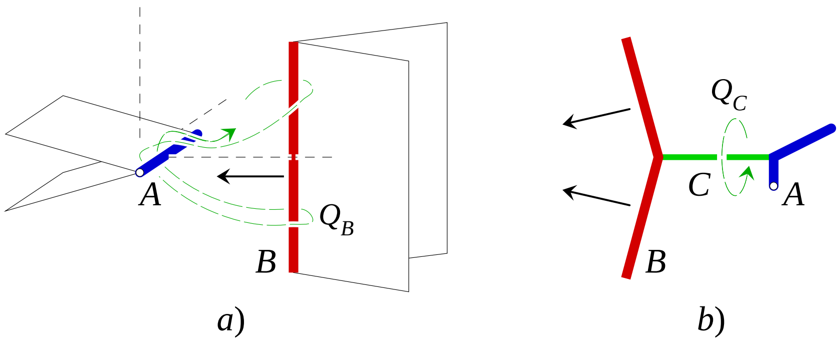

Figure 5: Orthogonal strings scattering. a) Initial state. String B moves towards string A (arrow). The cusps caused by their deficit angles are shown. Dashed lines are an orthogonal frame shown for reference. b) After the scattering, a new string C connecting the first two emerges. Strings A and B now both show a kink. The cusps in the last situation are not shown. The oriented dashed curves explain Eq. (3.14).

However, it is a physical limitation if we also assume A to be orthogonal to B. Just because it is special, we consider this case first. The collision event is sketched in Fig. 5. In 5a, the two strings approach one another. They both drag a space-time cusp with them. Now, what happens when B hits A, is best understood by drawing the cusp of A in the opposite direction. The result of that, however, is that string B is seen to have a kink. The same thing happens to string A itself; it develops a kink due to the cusp of B. After the passage, the two kinks must be connected by a new string, C that stretches with A and B now moving away from one another.

Indeed, we see that, in general, the holonomy of string C is non-trivial; it is obtained from the holonomies Q_A and Q_B of strings A and B as follows (depending on sign conventions for Q_A, Q_B and Q_C):

$$Q_C = Q_B^{-1}\, Q_A^{-1}\, Q_B\, Q_A\ . \tag{3.14}$$

Q_A and Q_B do not commute because they represent rotations along two different axes. Clearly, upon crossing, two strings produce a third stretching between them. We see that the inevitability of string crossings forces us to generalize our model to string *segments*, ending in joints with other segments. All segments may have different string 'constants' ϱ. We will have to address this feature again later.

If the strings approach one another at an angle different from 90°, a complication arises. In this case, one can convince oneself that no solution is possible with a single string stretching between the outgoing strings. This can be understood by studying the geometry, but we can also verify that, in general, the holonomy (3.14) is not of the string type: it violates Eq. (3.11).

To save the model, one can now propose the following. When two strings A and B collide at an angle $\varphi \neq 90^\circ$, not one but two new strings appear[2], both stretching from

[2]Later, in this section, we will argue that even more than two new strings might emerge.

A to B. A single string cannot be associated with a holonomy of the form (3.14), but a pair of strings can. The question is now, whether the data of this pair of strings would be uniquely determined by the initial characteristics of A and B. To investigate this question, the author combined analytical arguments with computer calculations, just to see how things will work out. The topology is defined in Fig. 6.

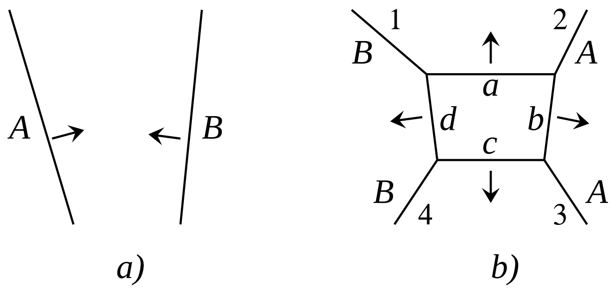

Figure 6: Scattering at an angle produces two new strands. a) A and B enter. b) Since A passes the cusp of B and *vice versa*, strands A_2 and A_3 form angles. These four pieces are labelled 1—4. Four new string pieces must be further specified, here labelled a — d.

At first sight, it seems that we have considerable freedom to define the orientations, strengths and velocities of the 'internal' strings a — d. However, if we fix one of these, all others are determined since the holonomies at a junction must obey Eq. (3.5). In addition, the strings must be properly attached to one another. As it turns out, the matching of the strings is guaranteed if Eq. (3.5) is obeyed at all junctions, and if in addition the *string conditions* (3.11) and (3.12) are obeyed by all four new holonomies, a — d.

The holonomy matrices at the 'external lines' 1—4 are fixed by the initial conditions. Originally, we had, in one conveniently chosen Lorentz frame, $Q_1 = Q_B$ and $Q_3 = Q_A^{-1}$. Here, the inverse sign arises if we decide to consider the holonomies with respect to observers looking towards the interaction point, and the holonomy curves C are chosen to go clockwise. Then,

$$\begin{aligned} Q_4 &= Q_3^{-1}\, Q_1^{-1}\, Q_3 \ ; \\ Q_2 &= Q_1^{-1}\, Q_3^{-1}\, Q_1 \ . \end{aligned} \tag{3.15}$$

The last equation is actually the one required for consistency with the demand that

$$Q_1\, Q_2\, Q_3\, Q_4 \ = \ \mathbb{I} \ . \tag{3.16}$$

Given the holonomy Q_a of the string section a, the others can be defined as follows:

$$Q_b = Q_a\, Q_2 \ ; \quad Q_c = Q_b\, Q_3 \ ; \quad Q_d = Q_a\, Q_1^{-1} \ , \tag{3.17}$$

which is the most systematic definition, and consistency with Eq. (3.16) is ensured. We emphasize that there is always some ambiguity in defining the Lorentz frames for the holonomies Q_1 — Q_4 and Q_a — Q_d, so we use Eqs. (3.15)—(3.17) also to specify these frames.

Let now all external holonomies $Q_1 — Q_4$ be given. How much freedom is there for $Q_a — Q_d$? We do not need to consider the translation parameters in the Poincaré group; these will be taken care of automatically, since there is a point (0,0,0,0) where the colliding strings A and B first met. This point can be kept at the origin of our coordinate frame. Thus, we consider the equations for the elements of the Lorentz group, which are most conveniently described as $SL(2,C)$ matrices. Each element of the Lorentz group is characterized by 6 real variables: a rotation vector and a velocity vector, or alternatively the four complex numbers in a 2×2 matrix Q, subject to the constraint that the complex number $\det(Q)$ should be set equal to 1.

The junction equations (3.15)—(3.17) leave us the freedom to choose Q_a. This gives a space with 6 real parameters. Then the string equation (3.11) for $Q_a — Q_d$, together gives us 4 real constraints. The surviving 2-dimensional manifold is then further constrained by the demands (3.12). Thus, the manifold of all possibilities is a two-dimensional space.

To obtain somewhat more understanding of this manifold, let us consider the 8 dimensional set of all $L(2,C)$ matrices for Q_a, without the nonlinear constraint concerning the determinant. The conditions (3.11), $\mathrm{Im}\,\mathrm{Tr}\,(Q_{a,b,c,d}) = 0$, in combination with the junction equations (3.17), are 4 linear equations for the matrix elements of Q_a. This leaves us with a linear $8 - 4 = 4$ dimensional space. Then we have the inequalities (3.12), which for the matrix Q_a imply that

$$Q_a = \begin{pmatrix} a_1 + ia_2 & b_1 + ib_2 \\ c_1 + ic_2 & d_1 - ia_2 \end{pmatrix} \quad ; \qquad |a_1 + d_1| \leq 2 \ , \tag{3.18}$$

and similarly for the three other internal holonomies. Realizing that, in our 4 dimensional space, these conditions can be written as

$$|e_i \cdot x| \leq 2 \ , \qquad i = 1, \dots, 4 \ , \tag{3.19}$$

and assuming that, in general, the four vectors e_i will be independent, we see that, in the generic case, the surviving space is a compact one: a four dimensional hypercube. We can be sure that the inequalities (3.19) give us a *non empty* four dimensional space.

Next, however, we have the two constraints

$$\mathrm{Re}(\det(Q_a)) = 1 \ , \qquad \mathrm{Im}(\det(Q_a)) = 0 \ . \tag{3.20}$$

These two equations for Q_a ensure that the same equation will hold for $Q_b — Q_d$, because the determinant is preserved, and because $\det(Q_i) = 1$ also for the external $Q_1 — Q_4$. Now these are quadratic equations for the coefficients of Q_a, so the question whether these two equations are compatible with the inequalities (3.19) and with one another is a more delicate one.

We refer to Ref[11] for further details. What has been discovered is that in a fraction of all cases indeed a class of double-strand solution exists, but when the strings meet at relativistic speeds, and possibly in some other exceptional circumstances, this class is empty: the 'double-strand Ansatz' then fails. In those cases, we have to try something else. More complicated scattering diagrams were considered, but we were not yet able to prove that configurations with either a single internal quadrangle or multiple strands suffice to cover all eventualities, as the space of all possible external holonomies Q_1 — Q_4 is very large[3].

Strings crossing over is not the only kind of "events" that can take place in this model. We can also encounter the situation where a string segment, of the kind that results from collisions of the type described in the above, is reduced to zero length. It is bounded by two other junctions that herewith merge into one. Our first try should be whether the result could again be a one-strand, two-strand, or multiple strand final state, just like the ones described earlier. However, if we allow ourselves strings with negative string constants, then there is a simpler final state: the one where the original string gets a "negative length". This is really a string where the deficit angle has switched sign. Once it was decided to allow their presence, we could allow them here as well.

3.6 Crystalline gravity

The gravity model we arrived at is dynamical, finite and Lorentz invariant. It consists of straight string segments, attached to one another at string joints. Space-time surrounding the strings and the joints is locally flat and Minkowskian. The strings form cusp singularities with deficit angles equal to the string constants. These string constants must be allowed to be different for the different strings. Matching conditions at the joints automatically ensure that the tensile forces of these strings balance out. The dynamical rules are then as follows:

1. When two strings or string segments collide, new straight sting segments form, either in the from of a quadrangle (Fig. 6), or in a more complicated setting.
2. When a string segment shrinks to zero length, a similar rule will prescribe the newly formed string segments.
3. There is a considerable amount of freedom in choosing the details of these rules, which is comparable to the freedom one has in choosing the matter Lagrangian in a gravity theory.
4. There is no independent gravitational degree of freedom. "Gravitons" are best regarded as composites of matter. The smoothness of gravity fields and waves that we normally consider, can only be reproduced in a continuum limit of the model.
5. It has not been checked whether the string constants can be kept positive when collisions take place. Most likely, this is not the case, so that we do have negative energy configurations.
6. The model cannot be quantized in the usual fashion, since
   a) strings with a continuous spectrum of different string constants appear;

[3] Note added in proof: we now think that configurations where the joints connect 4 strings rather than 3, might provide complete solution sets.

b) there appears to be no time reversal (or CPT) symmetry: strings can be seen to break up but they do not tend to rejoin.

An entirely different picture however arises if we consider imposing one more constraint: let us assume that not all elements of the Poincaré group are admitted, but instead only a discrete subgroup. The easiest way to think of this is to assume space and time to form a lattice in Minkowski space, preferably a rectangular lattice. Our stringlike defects are then much more like crystal defects. This is why we call the resulting theory "crystalline gravity". It differs from a crystal however in three important ways: our crystal now exists not only in space but also in time, and furthermore, our crystal defects obey equations of motion, as in the above. And finally, our crystal is not imbedded in a flat background spacetime, but it itself defines spacetime, which is curved along the defects.

The defect angles of stationary strings are now restricted to multiples of 90°. In an ordinary crystal, this would cause a lot of strain away from the defect, but that is because it would be imbedded in a flat background space, which is absent here. Unlike ordinary crystals, our spacetime crystal has an infinite rotation group in its crystal group: our set of discrete Lorentz transformations may contain the matrix

$$B = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix} , \tag{3.21}$$

since it is a legitimate Lorentz transformation, mapping integral coordinates onto integral coordinates. Combining this with space rotations over 90°, gives us the largest subgroup of the Lorentz group that leaves a cubical spacetime lattice invariant.[4]

One can find an infinity of further subgroups of this subgroup that could be considered, but we will not dwell on such features any further.

There are reasons why we will *not* advocate "quantization" of this model along the usual procedures. Perhaps, quantization can go by means of the "pre-quantization" procedure proposed earlier[12]. There are quite a few open questions apart from quantization. For instance, one might suspect an instability such as the formation of a black hole horizon, although precisely in this model one might also suspect the converse, that black holes cannot form. If all string constants are kept positive, localized matter configurations cannot exist, whereas all gravitational curvature must be associated with strings — there is no pure gravity in this model. So, if there is a black hole, strings will have to stick out from it.

The absence of pure gravity degrees of freedom is intriguing. In a sense, matter here is "unified" with gravity, not, as in many models, because gravity generates particle-like degrees of freedom, but the converse, because the matter degrees of freedom, here the string segments, carry around all the space-time curvature there is.

[4] Other discrete subgroups exist leaving different types of lattices invariant.

References

[1] IProceedings of the Amsterdam International Conference on Elementary Particles, held at Amsterdam, The Netherlands, June 30 - July 6, 1971 (North-Holland Pub. Co., 1972)

[2] International School of Subnuclear Physics, Director A. Zichichi, 13th Course: New Phenomena in Subnuclear Physics 11 July - 1 August 1975.

[3] S. Coleman and C. Sagan, *Spacecraft Sterilization Standards and Contamination of Mars*, Journal of Astronautics and Aeronautics **3**, 22 (1965).

[4] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu.S. Tyupkin, *Phys. Lett.* **59B** (1975) 85.

[5] R. C.G. Callan Jr., R.F. Dashen and D.J. Gross, *Phys. Lett.* **63B** (1976) 334.

[6] G.'t Hooft, *Phys. Rev. Lett.* **37** (1976) 8; *Phys. Rev.* **D14** (1976) 3432.

[7] S.L. Adler, *Phys. Rev.* **177** (1969) 2426; J.S. Bell and R. Jackiw, *Nuovo Cim.* **60A** (1969) 47 S.L. Adler and W.A. Bardeen, *Phys. Rev.* **182** (1969) 1517; W.A. Bardeen, *Phys. Rev.* **184** (1969) 1848.

[8] G. 't Hooft, textitThe Physics of Instantons in the Pseudoscalar and Vector Meson Mixing, SPIN-1999/01, hep-th/9903189.

[9] A. Zichichi, *The Beppo Particle*, Int. Symp. "Highlights in Physics Today: One hundred Years after the Birth of Beppo Occhialini", Milan,February 16, 2007.

[10] G. 't Hooft, G. Isidori, L. Maiani, A.D. Polosa and V. Riquer, *Phys. Lett.* **B662** (2008) 424, arXiv:0801.2288.

[11] G. 't Hooft, *Found. Phys.* **38** (2008) 733, ITP-UU-08/17, SPIN-08/16, arXiv:0804.0328.

[12] G. 't Hooft, *J. Phys: Conference Series* **67** (2007) 012015, quant-ph/0604008.

[13] S. Deser, R. Jackiw and G. 't Hooft, *Ann. Phys.* **152** (1984) 220.

[14] G. 't Hooft, *Nucl. Phys.* **B30** (Proc. Suppl.) (1993) 200; id., *Class. Quantum Grav.* **10** (1993) 1023.

[15] Z. Kadar, *Class. Quant. Grav.* **22** (2005) 809, e-Print: gr-qc/0410012.

[16] E. Witten, *Nucl. Phys.* **B311** (1988) 46; S. Carlip, *Canadian Gen.Rel.* 1993:0215-234 (QC6:C25:1993), gr-qc/9305020.

[17] G. 't Hooft, *Class. Quantum Grav.* **13** 1023, gr-qc/9607022.

[18] J.W. Barret et al, gr-qc/9411008, *Int.J.Theor.Phys.* **36** (1997) 815, and references therein.

[19] R. Brandenberger, H. Firouzjahi and J. Karouby, *Lensing and CMB Anisotropies by Cosmic Strings at a Junction*, arXiv:0710.1636 (gr-qc).