Review Article

Conceptual Problems in Quantum Gravity and Quantum Cosmology

Claus Kiefer

Institute for Theoretical Physics, University of Cologne, Zülpicher Strasse 77, 50937 Köln, Germany

Correspondence should be addressed to Claus Kiefer; kiefer@thp.uni-koeln.de

Received 26 May 2013; Accepted 28 June 2013

Academic Editors: M. Montesinos and M. Sebawe Abdalla

Copyright © 2013 Claus Kiefer. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The search for a consistent and empirically established quantum theory of gravity is among the biggest open problems of fundamental physics. The obstacles are of formal and of conceptual nature. Here, I address the main conceptual problems, discuss their present status, and outline further directions of research. For this purpose, the main current approaches to quantum gravity are briefly reviewed and compared.

1. Quantum Theory and Gravity—What Is the Connection?

According to our current knowledge, the fundamental interactions of nature are the strong, the electromagnetic, the weak, and the gravitational interactions. The first three are successfully described by the Standard Model of particle physics, in which a partial unification of the electromagnetic and the weak interactions has been achieved. Except for the nonvanishing neutrino masses, there exists at present no empirical fact that is clearly at variance with the Standard Model. Gravity is described by Einstein’s theory of general relativity (GR), and no empirical fact is known that is in clear contradiction to GR. From a pure empirical point of view, we thus have no reason to search for new physical laws. From a theoretical (mathematical and conceptual) point of view, however, the situation is not satisfactory. Whereas the Standard Model is a quantum field theory describing an incomplete unification of interactions, GR is a classical theory.

Let us have a brief look at Einstein’s theory, see, for example, Misner et al. [1]. It can be defined by the Einstein-Hilbert action

\[ S_{EH} = \frac{c^4}{16\pi G} \int d^4x \sqrt{-g} (R - 2\Lambda) - \frac{c^4}{8\pi G} \int_{\partial \mathcal{M}} d^3x \sqrt{h} K, \]

(1)

where \( g \) is the determinant of the metric, \( R \) is the Ricci scalar, and \( \Lambda \) is the cosmological constant. In addition to the two main terms, which consist of integrals over a spacetime region \( \mathcal{M} \), there is a term that is defined on the boundary \( \partial \mathcal{M} \) (here assumed to be space like) of this region. This term is needed for a consistent variational principle; here, \( h \) is the determinant of the three-dimensional metric, and \( K \) is the trace of the second fundamental form.

In the presence of nongravitational fields, (1) is augmented by a “matter action” \( S_m \). From the sum of these actions, one finds Einstein’s field equations by variation with respect to the metric,

\[ G_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} - \Lambda g_{\mu\nu}. \]

(2)

The right-hand side displays the symmetric (Belinfante) energy-momentum tensor

\[ T_{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{\mu\nu}}, \]

(3)

plus the cosmological-constant term, which may itself be accommodated into the energy-momentum tensor as a contribution of the “vacuum energy.” If fermionic fields are added, one must generalize GR to the Einstein-Cartan theory or to the Poincaré gauge theory, because spin is the source of torsion, a geometric quantity that is identically zero in GR (see e.g., [2]).
As one recognizes from (2), these equations can no longer have exactly the same form if the quantum nature of the fields in $T_{\mu\nu}$ is taken into account. For then we have operators in Hilbert space on the right-hand side and classical functions on the left-hand side. A straightforward generalization would be to replace $T_{\mu\nu}$ by its quantum expectation value, 

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} \langle \Psi \left| \hat{T}_{\mu\nu} \right| \Psi \rangle .$$  \hspace{1cm} (4)$$

These "semiclassical Einstein equations" lead to problems when viewed as exact equations at the most fundamental level, compare Carlip [3] and the references therein. They spoil the linearity of quantum theory and even seem to be in conflict with a performed experiment [4]. They may nevertheless be of some value in an approximate way. Independent of the problems with (4), one can try to test them in a simple setting such as the Schrödinger-Newton equation; it seems, however, that such a test is not realizable in the foreseeable future [5]. This poses the question of the connection between gravity and quantum theory [6].

Despite its name, quantum theory is not a particular theory for a particular interaction. It is rather a general framework for physical theories, whose fundamental concepts have so far exhibited an amazing universality. Despite the ongoing discussion about its interpretational foundations (which we shall address in the last section), the concepts of states in Hilbert space, and in particular the superposition principle, have successfully passed thousands of experimental tests.

It is, in fact, the superposition principle that points towards the need for quantizing gravity. In the 1957 Chapel Hill Conference, Richard Feynman gave the following argument [7, pages 250–260], see also Zeh [8]. He considers a Stern-Gerlach type of experiment in which two spin-1/2 particles are put into a superposition of spinup and spindown and is guided to two counters. He then imagines a connection of the counters to a ball of macroscopic dimensions. The superposition of the particles is thereby transferred to a superposition of the ball being simultaneously at two positions. But this means that the ball's gravitational field is in a superposition, too. In Feynman's own words [7, p. 251]:

"Now, how do we analyze this experiment according to quantum mechanics? We have an amplitude that the ball is up and an amplitude that the ball is down. That is, we have an amplitude (from a wave function) that the spin of the electron in the first part of the equipment is either up or down. And if we imagine that the ball can be analyzed through the interconnections up to this dimension (=1 cm) by the quantum mechanics, then before we make an observation we still have to give an amplitude when the ball is up and an amplitude when the ball is down. Now, since the ball is big enough to produce a real gravitational field . . . we could use that gravitational field to move another ball, and amplify that, and use the connections to the second ball as the measuring equipment. We would then have to analyze through the channel provided by the gravitational field itself via the quantum mechanical amplitudes."

In other words, the gravitational field must then be described by quantum states subject to the superposition principle. The same argument can, of course, be applied to the electromagnetic field (for which we have strong empirical support that it is of quantum nature).

Feynman's argument is, of course, not an argument of logic that would demand the quantization of gravity with necessity. It is an argument based on conservative heuristic ideas that proceed from the extrapolation of established and empirically confirmed concepts (here, the superposition principle) beyond their present range of application. It is in this way that physics usually evolves. Albers et al. [9] contain a detailed account of arguments that demand the necessity of quantizing the gravitational field. It is shown that all these arguments are of a heuristic value and that they do not lead to the quantization of gravity by a logical conclusion. It is conceivable, for example, that the linearity of quantum theory breaks down in situations where the gravitational field becomes strong, see, for example, Penrose [10], Singh [11], and Bassi et al. [12]. But one has to emphasize that no empirical hint for such a drastic modification exists so far.

Alternatives to the direct quantization of the gravitational field include what is called "emergent gravity", compare Padmanabhan [13] and the references therein. Motivated by the thermodynamic properties of black holes (see below), one might get the impression that the gravitational field is an effective thermodynamic entity that does not demand its direct quantization but points to the existence of new so far unknown microscopic degrees of freedom underlying gravity. Even if this was the case (which is far from clear), there may exist microscopic degrees of freedom for which quantum theory would apply. Whether this leads to a quantized metric or not is not clear. In string theory (see below), gravity is an emergent interaction, but still the metric is quantized, and the standard approach of quantum gravitational perturbation theory is naturally implemented. In the following, we restrict ourselves to approaches in which a quantized metric makes sense.

Besides the general argument put forward by Feynman, there exist a couple of further arguments that suggest that the gravitational interaction be quantized [6]. Let me briefly review three of them.

The first motivation comes from the continuation of the reductionist programme. In physics, the idea of unification has been very successful, culminating so far in the Standard Model of strong and electroweak interactions. Gravity acts universally to all forms of energies. A unified theory of all interactions including gravity should thus not be a hybrid theory in using classical and quantum concepts. A coherent quantum theory of all interactions (often called "theory of everything" or TOE) should thus also include a quantum description of the gravitational field.

A second motivation is the unavoidable presence of singularities in Einstein's theory of GR, see, for example, Hawking and Penrose [14] and Rendall [15]. Prominent examples are the cases of the big bang and the interior of
black holes. One would thus expect that a more fundamental theory encompassing GR does not predict any singularities. This would be similar to the case of the classical singularities from electrodynamics, which are avoided in quantum electrodynamics. The fate of the classical singularities in some approaches to quantum gravity will be discussed below.

A third motivation is known as the "problem of time." In quantum mechanics, time is absolute. The parameter $t$ occurring in the Schrödinger equation has been directly inherited from Newtonian mechanics and is not turned into an operator. In quantum field theory, time by itself is no longer absolute, but the four-dimensional spacetime is; it constitutes the fixed background structure on which the dynamical fields act. GR is of a very different nature. According to the Einstein equations (2), spacetime is dynamical, acting in a complicated manner with energy momentum of matter and with itself. The concepts of time (spacetime) in quantum theory and GR are thus drastically different and cannot both be fundamentally true. One thus needs a more fundamental theory with a coherent notion of time. The absence of a fixed background structure is also called "background independence" (cf. [16]) and often used a leitmotiv in the search for a quantum theory of gravity, although it is not quite the same as the problem of time, as we shall see below.

A central problem in the search for a quantum theory of gravity is the current lack of a clear empirical guideline. This is partly related to the fact that the relevant scales, on which quantum effects of gravity should definitely be relevant, are far remote from being directly explorable. The scale is referred to as the Planck scale, $l_P$, Planck time, $t_P$, and Planck mass, $m_P$, respectively. They are given by the expressions

\[ l_P := \frac{\hbar G}{c^3} \approx 1.62 \times 10^{-33} \text{ cm}, \]
\[ t_P := \frac{l_P}{c} = \frac{\hbar G}{c^4} \approx 5.39 \times 10^{-44} \text{ s}, \]
\[ m_P := \frac{\hbar}{l_P c} = \frac{\hbar c}{G} \approx 2.18 \times 10^{-5} \text{ g} \approx 1.22 \times 10^{19} \text{ GeV}/c^2. \]

It must be emphasized that units of length, time, and mass cannot be formed out of $G$ and $c$ (GR) or out of $\hbar$ and $c$ (quantum theory) alone.

In view of the Planck scale, the Standard Model provides an additional motivation for quantum gravity: it seems that the Standard Model does not exist as a consistent quantum field theory up to arbitrarily high energies, see, for example, Nicolai [17]. The reasons for this failure may be a potential instability of the effective potential and the existence of Landau poles. The Standard Model can thus by itself not be a fundamental theory, although with the recently measured Higgs mass $m_{1H}$ of about 126 GeV, it is in principle conceivable that it holds up to the Planck scale. It is so far an open issue why the Higgs mass is stabilized at such a low value and not driven to high energies by quantum loop corrections ("hierarchy problem"). In fact, one has

\[ \left( \frac{m_p}{m_{1H}} \right)^2 \sim 10^{34}. \tag{6} \]

Possible solutions to the hierarchy problem include supersymmetric models and models with higher dimensions, but a clear solution is not yet available.

In astrophysics, the Planck scale is usually of no relevance. The reason is that structures in the Universe, with the exception of black holes, occur at (length, time, and mass) scales that are different from the Planck scale by many orders of magnitude. This difference is quantified by the "fine-structure constant of gravity" defined by

\[ \alpha_g := \frac{G m^2_{pr}}{\hbar c} = \left( \frac{m_{pr}}{m_p} \right)^2 \approx 5.91 \times 10^{-39}, \tag{7} \]

where $m_{pr}$ denotes the proton mass. The Chandrasekhar mass, for example, which gives the correct order of magnitude for main sequence stars like the Sun, is given by $\alpha_g^{-1/2} m_{pr}$.

Let us have at the end of this section a brief look at the connection between quantum theory and gravity at the level where gravity is treated as a classical interaction [6]. The lowest level is quantum mechanics plus Newtonian gravity, at which many experimental tests exist. Most of them employ atom or neutron interferometry and can be described by the Schrödinger equation with the Hamiltonian given by

\[ H = \frac{\mathbf{p}^2}{2m} + m g r - \omega L, \tag{8} \]

where the second term is the Newtonian potential in the limit of constant gravitational acceleration, and the last term describes a coupling between the rotation of the Earth (or another rotating system) and the angular momentum of the particle. An interesting recent suggestion in this context is the possibility to see the general relativistic time dilatation in the interference pattern produced by the interference of two partial particle beams at different heights in the gravitational field [18]. A more general treatment is based on the Dirac equation and its nonrelativistic ("Foldy-Wouthuysen") expansion.

The next level is quantum field theory in a curved spacetime (or in a flat spacetime, but in noninertial coordinates). Here, concrete predictions are available, although they have so far not been empirically confirmed. Perhaps the most famous prediction is the Hawking effect according to which every stationary black hole is characterized by the temperature [19]

\[ T_{BH} = \frac{\hbar c}{2 \kappa \hbar k_B}, \tag{9} \]

where $\kappa$ is the surface gravity of a stationary black hole, which by the no-hair theorem is uniquely characterized by its mass $M$, its angular momentum $J$, and (if present) its electric charge $q$. In the particular case of the spherically symmetric Schwarzschild black hole, one has $\kappa = c^4/4GM = GM/R_S^2$, where $R_S$ is the Schwarzschild radius.
where $R_S = \frac{2GM}{c^2}$ is the Schwarzschild radius, and therefore

$$T_{BH} = \frac{\hbar c^3}{8\pi k_B GM} \approx 6.17 \times 10^{-8} \left( \frac{M_\odot}{M} \right) \text{ K.} \quad (10)$$

The presence of a temperature for black holes means that these objects have a finite lifetime. Upon radiating away energy, they become hotter, releasing even more energy, until all the mass (or almost all the mass) has been radiated away. For the lifetime of a Schwarzschild black hole, one finds the expression [20]

$$\tau_{BH} \approx 407 \left( \frac{f(M_0)}{15.35} \right) \left( \frac{M}{10^{10} \text{ g}} \right)^3 \text{ s} \approx 6.24 \times 10^{-27} M_0^3 [\text{g}] f^{-1}(M_0) \text{ s}, \quad (11)$$

where $f(M_0)$ is a measure of the number of emitted particle species; it is normalized to $f(M_0) = 1$ for $M_0 \gg 10^{17} \text{ g}$ when only effectively massless particles are emitted. If one sums over the contributions from all particles in the Standard Model up to an energy of about 1 TeV, one finds $f(M) = 15.35$, which motivates the occurrence of this number in (11). Because the temperature of black holes that result from stellar collapse is too small to be observable, the hope is that primordial black holes exist for which the temperature can become high [21].

Since black holes have a temperature, they also have an entropy. It is called “Bekenstein-Hawking entropy” and is found from thermodynamic arguments to be given by the universal expression

$$S_{BH} = \frac{k_B A}{4l_p^2}, \quad (12)$$

where $A$ is the surface of the event horizon. For the special case of the Schwarzschild black hole, it reads

$$S_{BH} = \frac{k_B \pi R^2}{Gh} \approx 1.07 \times 10^{-77} k_B \left( \frac{M}{M_\odot} \right)^2. \quad (13)$$

All these expressions depend on the fundamental constants $\hbar$, $G$, and $c$. They may thus provide a key for quantum gravity.

In flat spacetime, an effect exists that is analogous to the black-hole temperature (9). If an observer moves with constant acceleration through the standard Minkowski vacuum, he/she will perceive this state not as empty but as filled with thermal particles ("Unruh effect"), see Unruh [22]. The temperature is given by the "Davies-Unruh temperature"

$$T_{DU} = \frac{\hbar a}{2\pi k_B g} \approx 4.05 \times 10^{-23} a \left[ \frac{\text{cm}}{\text{s}^2} \right] \text{ K.} \quad (14)$$

The similarity of (14) and (9) is connected with the presence of an event horizon in both cases. A suggestion for an experimental test of (14) can be found in Thirolf et al. [23].

Many of the open questions to be addressed in any theory of quantum gravity are connected with the temperature and the entropy of black holes [24]. The two most important questions concern the microscopic interpretation of entropy and the final fate of a black hole; the latter is deeply intertwined with the problem of information loss.

The Bekenstein-Hawking entropy (12) was derived from thermodynamic considerations, without identifying appropriate microstates and performing a counting in the sense of statistical mechanics. Depending on the particular approach to quantum gravity, various ways of counting states have been developed. In most cases, one can get the behaviour that the statistical entropy $S_{stat} \propto A$, although not necessarily with the desired factor occurring in (12).

Hawking's calculation leading to (9) breaks down when the black hole becomes small, and effects of full quantum gravity are expected to come into play. This raises the following question. According to (9), the radiation of the black hole is thermal. What, then, happens in the final phase of the black-hole evolution? If only thermal radiation was left, all initial states that lead to a black hole would end up in one and the same final state—a thermal state, that is, the information about the initial state would be lost. This is certainly in contradiction with standard quantum theory for a closed system, for which the von Neumann entropy $S = -k_B \text{tr}(\rho \ln \rho)$ is constant, where $\rho$ denotes the density matrix of the system. This possible contradiction is called the “information-loss paradox” or “information-loss problem” for black holes [25].

Much has been said since 1976 about the information-loss problem (see, e.g., [26, 27]), without final consensus. This is not surprising, because the final solution will only be obtained if the final theory of quantum gravity is available. Some remarks can be made, though. Hawking's original calculation does not show that black-hole radiation is strictly thermal. It only shows that the expectation value of the particle-number operator is strictly thermal. There certainly exist pure quantum states which lead to such a thermal expression [28, 29]. A relevant example is a two-mode squeezed state, which after tracing out one mode leads to the density matrix of a canonical ensemble; such a state can be taken as a model for a quantum state entangling the interior and exterior of the black hole, see Kiefer [28, 29] and the references therein.

The same can be said about the Unruh effect. The temperature (14) can be understood as arising from tracing out degrees of freedom in the total quantum state, which is pure [30]. In the case of a moving mirror, the pure quantum state can exhibit thermal behaviour, without any information loss (see, e.g., [31, Section 4.4]).

For a semiclassical black hole, therefore, the information-loss problem does not arise. The black hole can, and in fact must, be treated as an open quantum system, so that a mixed state emerges from the process of decoherence [32]. The total state of system and environment (which means everything interacting with the black hole) can be assumed to stay in a pure state. Still, there is some discussion whether something unusual happens at the horizon even for a large black hole [33], but this is a contentious issue.

If the black hole approaches the Planck regime, the question about the information-loss problem is related to the fate of the singularity. If the singularity remains in quantum gravity (which is highly unlikely), information will indeed be
destroyed. Most approaches to quantum gravity indicate that entropy is conserved for the total system, so there will not be an information-loss problem, in accordance with standard quantum theory. How the exact quantum state during the final evaporation phase looks like is unclear. One can make oversimplified models with harmonic oscillators [34], but an exact solution from an approach to quantum gravity is elusive.

2. Main Approaches to Quantum Gravity

Following Isham [35], one can divide the approaches roughly into two classes. In the first class, one starts from a given classical theory of gravity and applies certain quantization rules to arrive at a quantum theory of gravity. In most cases, the starting point is GR. This does not yet lead to a unification of interactions; one arrives at a separate quantum theory for the gravitational field, in analogy to quantum electrodynamics (QED). Most likely, the resulting theory is an effective theory only, valid only in certain situations and for certain scales "it is generally believed today that the realistic theories that we use to describe physics at accessible energies are what are known "effective field theories."

Depending on the method used, one distinguishes between covariant and canonical quantum gravity.

The second class consists of approaches that seek to construct a unified theory of all interactions. Quantum aspects of gravity are then seen only in a certain limit—in the limit where the various interactions become distinguishable. The main representative of this second class is string theory.

In the rest of this section, I shall give a brief overview of the main approaches. For more details, I refer to Kiefer [6]. In most expressions, units are chosen with $c = 1$.

2.1. Covariant Quantum Gravity. In covariant quantum gravity, one employs methods that make use of four-dimensional covariance. Today, this is usually done by using the quantum gravitational path integral, see, for example, Hamber [37]. Formally, the path integral reads

$$Z [g] = \int \mathcal{D} g_{\mu \nu} (x) e^{i S [g_{\mu \nu}]} ,$$

(15)

where the sum runs over all metrics on a four-dimensional manifold $\mathcal{M}$ quotiented by the diffeomorphism group Diff.$\mathcal{M}$. In addition, one may wish to perform a sum over all topologies, because they may also be subject to the superposition principle. This is, however, not possible in full generality, because four manifolds are not classifiable. Considerable care must be taken in the treatment of the integration measure. In order to make it well defined, one has to apply the Faddeev-Popov procedure known from gauge theories.

One application of the path integral (15) is the derivation of Feynman rules for the perturbation theory. One makes the ansatz

$$g_{\mu \nu} = \bar{g}_{\mu \nu} + \sqrt{\frac{32 \pi G}{f_{\mu \nu}},}$$

(16)

where $\bar{g}_{\mu \nu}$ denotes the background field with respect to which covariance is implemented in the formalism and $f_{\mu \nu}$ denotes the quantized field (the "gravitons"), with respect to which the perturbation theory is performed. Covariance with respect to the background metric means that no particular background is distinguished; in this sense, "background independence" is implemented into the formalism.

There is an important difference in the quantum gravitational perturbation theory as compared to the Standard Model: the theory is nonrenormalizable. This means that one encounters a new type of divergences at each order of perturbation theory, resulting in an infinite number of free parameters. For example, at two loops the following divergence in the Lagrangian is found [38]:

$$\mathcal{L}^{(\text{div})}_{2-\text{loop}} = \frac{209 h^2}{2880} \frac{32 \pi G}{(16 \pi^2)^3} e^{-\sqrt{\frac{32 \pi G}{f_{\mu \nu}}}} \frac{e^{\alpha \delta}}{e^{\alpha \delta}} \frac{R_{\mu \nu}}{R_{\mu \nu}},$$

(17)

where $\epsilon = 4 - D,$ with $D$ being the number of spacetime dimensions, and $R_{\mu \nu}$ and so forth, denotes the Riemann tensor corresponding to the background metric.

New developments have given rise to the hope that a generalization of covariant quantum general relativity may not even be renormalizable but even finite—this is $N = 8$ supergravity. Bern et al. [39] have found that the theory is finite at least up to four loops. This indicates that a hitherto unknown symmetry may be responsible for the finiteness of this theory. Whether such a symmetry really exists and what its nature could be are not known at present.

Independent of the problem of nonrenormalizability, one can study covariant quantum gravity at an effective level, truncating the theory at, for example, the one-loop level. At this level, concrete predictions can be made, because the ambiguity from the free parameters at higher order does not enter. One example is the calculation of the quantum gravitational correction to the Newtonian potential between two masses [40]. The potential at one-loop order reads

$$V (r) = -\frac{G m_1 m_2}{r} \times \left(1 + 3 \frac{G (m_1 + m_2)}{r c^2} + \frac{41 G h}{10 \pi r^2 c^5} + \mathcal{O} (G^2)\right),$$

(18)

where the first correction term is an effect from classical GR, and only the second term is a genuine quantum gravitational correction term (which is, however, too small to be measurable).

Apart from perturbation theory, expressions such as (15) for the path integral can, in four spacetime dimensions and higher, only be defined and evaluated by numerical methods. One example is Causal Dynamical Triangulation (CDT) [41]. Here, spacetime is foliated into a set of four-dimensional simplices and Monte Carlo methods are used for the path integral. It was found that spacetime appears indeed effectively as four-dimensional in the macroscopic limit, but becomes two-dimensional when approaching the Planck scale.

This effective two-dimensionality is also seen in another approach of covariant quantum gravity-asymptotic safety (see e.g., [42]). A theory is called asymptotically safe if all
essential coupling parameters approach for large energies a fixed point where at least one of them does not vanish. (if they all vanish, one has the situation of asymptotic freedom). Making use of renormalization group equations, strong indications have been found that quantum GR is asymptotically safe. If true, this would be an example for a theory of quantum gravity that is valid at all scales. The small-scale structure of spacetime is among the most exciting open problems in quantum gravity; see Carlip [43] and the articles collected in the volume edited by Amelino-Camelia and Kowalski-Glikman [44].

2.2. Canonical Approaches. In canonical approaches to quantum gravity, one constructs a Hamiltonian formalism at the classical level before quantization. In this procedure, spacetime is foliated into a family of spacelike hypersurfaces. This leads to the presence of constraints, which are connected with the invariances of the theory. One has four (local) constraints associated with the classical diffeomorphisms. One is the Hamiltonian constraint \( \mathcal{H}_\perp \), which generates hypersurface deformations (many-fingered time evolution); the three other constraints are the momentum or diffeomorphism constraints \( \mathcal{H}_a \), which generate three-dimensional coordinate transformations. If one uses tetrads instead of metrics, three additional constraints (“Gauss constraints”) associated with the freedom of performing local Lorentz transformations are present. Classically, the constraints obey a closed (but not Lie) algebra. For the exact relation of the constraints to the classical spacetime diffeomorphisms, see, for example, Pons et al. [45] and Barbour and Foster [46].

By quantization, the constraints are turned into quantum constraints for physically admissible wave functionals. The exact form depends on the choice of canonical variables. If one uses the three-dimensional metric as the configuration variable, one arrives at quantum geometrodynamics. If one uses a certain holonomy as the configuration variable, one arrives at loop quantum gravity.

2.2.1. Quantum Geometrodynamics. In geometrodynamics, the canonical variables are the three-metric \( h_{ij}(x) \) and its conjugate momentum \( p^{cd}(y) \), which is linearly related to the second fundamental form. In the quantum theory, they are turned into operators that obey the standard commutation rules,

\[
\left[ \hat{h}_{ab}(x), \hat{p}^{cd}(y) \right] = i\hbar \delta_{(a} \delta_{b)} \delta(x, y).
\]

Adopting a general procedure suggested by Dirac, the constraints are implemented as quantum constraints on the wave functionals,

\[
\mathcal{H}_\perp \Psi = 0, \\
\mathcal{H}_a \Psi = 0.
\]

The first equation is called Wheeler-DeWitt equation [47, 48]; the three other equations are called quantum momentum or diffeomorphism constraints. The latter guarantee that the wave functional is invariant under three dimensional coordinate transformations. The configuration space of all three-metrics divided by three-dimensional diffeomorphisms is called superspace.

In the vacuum case, the above equations assume the explicit form

\[
\mathcal{H}_\perp \Psi := \left( -16\pi G h^2 G_{abcd} \frac{\delta^2}{\delta h_{da} \delta h_{cd}} - \frac{\sqrt{h}}{16\pi G} (^{(3)} R - 2\Lambda) \right) \Psi = 0,
\]

\[
\mathcal{H}_a \Psi := -2D_a h_{ac} \frac{\delta}{\delta h_{bc}} \Psi = 0.
\]

Here, \( D_a \) is the three-dimensional covariant derivative, \( G_{abcd} \) is the DeWitt metric (which is an ultralocal function of the three-metric), and \( ^{(3)} R \) is the three-dimensional Ricci scalar. In the presence of nongravitational fields, their Hamiltonians are added to the expressions (21) and (22).

There are various problems connected with these equations. One is the problem to make mathematical sense out of them and to look for solutions. This includes the implementation of the positive definiteness of the three-metric, which may point to the need of an “affine quantization” [49]. Other problems are of conceptual nature and will be discussed below. Attempts to derive the entropy (12) exist in this framework, see, for example, Vaz et al. [50], but the correct proportionality factor between the entropy and \( A \) has not yet been reproduced.

2.2.2. Loop Quantum Gravity. In loop quantum gravity, variables are used that are conceptually closer to Yang-Mills type of variables. The loop variables have grown out of “Ashtekar’s new variables” [51], which are defined as follows. The role of the momentum variable is played by the densitized triad (dreibein)

\[
E^a_i (x) := \sqrt{h}(x) e^a_i (x),
\]

while the configuration variable is the connection

\[
G A^a_i (x) = \Gamma^a_i (x) + \beta K^a_i (x).
\]

Here, \( a(i) \) denotes a space index (internal index); \( \Gamma^a_i (x) \) is the spin connection; and \( K^a_i (x) \) is related to the second fundamental form. The parameter \( \beta \) is called Barbero-Immirzi parameter and can assume any nonvanishing real value. In loop quantum gravity, it is a free parameter. It may be fixed by the requirement that the black-hole entropy calculated from loop quantum cosmology coincides with the Bekenstein-Hawking expression (12). One thereby finds (from numerically solving an equation) the value \( \beta = 0.23753 \ldots \), see Agullo et al. [52] and the references therein. It is of interest to note that the proportionality between entropy and area can only be obtained if the microscopic degrees of freedom (the spin networks) are distinguishable (cf. [53]).

Classically, Ashtekar’s variables obey the Poisson-bracket relation

\[
\{ A^a_i (x), E_b^y (y) \} = 8\pi G \delta^a_b \delta_i^y \delta(x, y).
\]
The loop variables are constructed from these variables in a nonlocal fashion. The new connection variable is the holonomy $U[A, \alpha]$, which is a path-ordered exponential of the integral over the connection (24) around a loop $\alpha$. In the quantum theory, it acts on wave functionals as

$$\tilde{U}[A, \alpha] \Psi_5[A] = U[A, \alpha] \Psi_5[A].$$

(26)

The new momentum variable is the flux of the densitized triad through a two-dimensional surface $\mathcal{S}$ bounded by the loop. Its operator version reads

$$\tilde{E}_i[\mathcal{S}] := -8\pi\beta\hbar \int_{\mathcal{S}} d\sigma^i d\alpha^2 n_\alpha(\mathcal{A}) \frac{\delta}{\delta \mathcal{A}_i[\mathbf{x}(\mathcal{A})]},$$

(27)

where the embedding of the surface is given by $(\alpha^1, \alpha^2) \equiv \mathcal{A}$. The variables obey the commutation relations

$$\left[ \tilde{U}[A, \alpha], \tilde{E}_i[\mathcal{S}] \right] = 8\pi\beta\hbar U[\alpha_1, A] \tau_i U[\alpha_2, A],$$

(28)

where $i(\alpha, \mathcal{S}) = \pm 1, 0$ is the "intersection number," which depends on the orientations of $\alpha$ and $\mathcal{S}$. Given certain mild assumptions, the holonomy-flux representation is unique and gives rise to a unique Hilbert-space structure at the kinematical level, that is, before the constraints are imposed.

At this level, one can define an area operator for which one obtains the following spectrum:

$$\tilde{A}[\mathcal{S}] \Psi_5[A] = 8\pi\beta\hbar \frac{1}{2} \sum_{P \in \mathcal{S}} \sqrt{J_P(J_P + 1)} \Psi_5[A]$$

$$=: A[\mathcal{S}] \Psi_5[A].$$

(29)

Here, $P$ denotes the intersection points between the spin-network $S$ and the surface $\mathcal{S}$, and the $J_P$ can assume integer and half-integer values (arising from the use of the group $SU(2)$ for the triads). There thus exists a minimal "quantum of action" of the order of $\beta$ times the Planck-length squared. Comprehensive discussions of loop quantum gravity can be found in Gambini and Pullin [54], Rovelli [55], Ashtekar and Lewandowski [56], and Thiemann [57].

Here, we have been mainly concerned with the canonical version of loop quantum gravity. But there exists also a covariant version: it corresponds to a path-integral formulation, through which the spin networks are evolved "in time". It is called the spin-foam approach, compare Rovelli [58] and the references therein.

2.3. String Theory. String theory is fundamentally different from the approaches discussed so far. It is not a direct quantization of GR or any other classical theory of gravity. It is an example for a unified quantum theory of all interactions. Gravity, as well as the other known interactions, only emerges in an appropriate limit (this is why string theory is an example of "emergent gravity.") Strings are one-dimensional objects characterized by a dimensionful parameter $\alpha'$ or the string length $l_s = \sqrt{2\alpha'\hbar}$. In spacetime, it forms a two-dimensional surface, the worldsheet. Closer inspection of the theory exhibits also the presence of higher-dimensional objects called D-branes, which are as important as the strings themselves, compare Blumenhagen et al. [59].

String theory necessarily contains gravity, because the graviton appears as an excitation of closed strings. It is through this appearance that a connection to covariant quantum gravity discussed above can be made. String theory also includes gauge theories, since the corresponding gauge bosons are found in the spectrum. It also requires the presence of supersymmetry for a consistent formulation. Fermions are thus an important ingredient of string theory. One recognizes that gravity, other fields, and matter appear on the same footing.

Because of reparametrization invariance on the worldsheet, string theory also possesses constraint equations. The constraints do, however, not close but contain a central term on the right-hand side. This corresponds to the presence of an anomaly (connected with Weyl transformations). The vanishing of this anomaly can be achieved if ghost fields are added that gain a central term which cancels the original one. The important point is that this works only in a particular number $D$ of dimensions: $D = 26$ for the bosonic string, and $D = 10$ for the superstring (or $D = 11$ in M-theory). The presence of higher spacetime dimensions is an essential ingredient of string theory.

Let us consider, for simplicity, the bosonic string. Its quantization is usually performed through the Euclidean path integral

$$Z = \int \mathcal{D}X \mathcal{D}\sigma e^{-S_P},$$

(30)

where $X$ and $h$ are a shorthand for the embedding variables and the worldsheet metric, respectively. The action in the exponent is the "Polyakov action," which is an action defined on the worldsheet. Besides the dynamical variables $X$ and $h$, it contains various background fields on spacetime, among them the metric of the embedding space and a scalar field called dilaton. It is obvious that this formulation is not background independent. In as much string theory (or M-theory) can be formulated in a fully background independent way is a controversial issue. It has been argued that partial background independence is implemented in the context of the AdS/CFT conjecture, see Section 4 below.

If the string propagates in a curved spacetime with metric $g_{\mu\nu}$, the demand for the absence of a Weyl anomaly leads to consistency equations that correspond (up to terms of order $\alpha'$) to Einstein equations for the background fields. These equations can be obtained from an effective action of the form

$$S_{\text{eff}} \propto \int d^D x \sqrt{-g} \ e^{-2\phi}$$

$$\times \left( R - \frac{2(D - 2)\alpha'}{3\alpha'} - \frac{1}{12} H_{\mu
u\rho} H^{\mu\nu\rho} + 4\nabla_{\mu} \Phi \nabla^{\mu} \Phi + \mathcal{O}(\alpha') \right).$$

(31)
where $\Phi$ is the dilaton, $R$ is the Ricci scalar corresponding to $g_{\mu\nu}$, and $H_{\mu\nu}$ is the field strength associated with an antisymmetric tensor field (which in $D = 4$ would be the axion). This is the second connection of string theory with gravity, after the appearance of the graviton as a string excitation. A recent comprehensive overview of string theory is Blumenhagen et al. [59].

Attempts were made in string theory to derive the Bekenstein-Hawking entropy from a microscopic counting of states. In this context, the D-branes turned out to be of central importance. Counting D-brane states for extremal and close-to-extremal string black holes, one was able to derive (12) including the precise prefactor [60, 61]. For the Schwarzschild black hole, however, such a derivation is elusive.

Originally, string theory was devised as a “theory of everything” in the strict sense. This means that the hope was entertained to derive all known physical laws including all parameters and coupling constants from this fundamental theory. So far, this hope remains unfulfilled. Moreover, there are indications that this may not be possible at all, due to the “landscape problem”: string theory seems to lead to many ground states at the effective level, the number exceeding $10^{500}$ [62]. Without any further idea, it seems that a selection can be made only on the basis of the anthropic principle, compare Carr [63]. If this was the case, string theory would no longer be a predictive theory in the traditional sense.

2.4. Other Approaches. Besides the approaches mentioned so far, there exist further approaches which are either meant to be separate quantum theories of the gravitational field or candidates for a fundamental quantum theory of all interactions. Some of them have grown out of one of the above approaches; others have been devised from scratch. It is interesting to note that most of these alternatives start from discrete structures at the microscopic level. Among them are such approaches as causal sets, spin foams, group field theory, quantum topology, and theories invoking some type of noncommutative geometry. Most of them have not been developed as far as the approaches discussed above. An overview can be found in Oriti [64].

3. Quantum Cosmology

Quantum cosmology is the application of quantum theory to the universe as a whole. Conceptually, this corresponds to the problem of formulating a quantum theory for a closed system from within, without reference to any external observers or measurement agencies. In its concrete formulation, it demands a quantum theory of gravity, since gravity is the dominating interaction at large scales. On the one hand, quantum cosmology may serve as a testbed for quantum gravity in a mathematically simpler setting. This concerns, in particular, the conceptual questions which are of concern here. On the other hand, quantum cosmology may be directly relevant for an understanding of the real Universe. General introductions into quantum cosmology include Coule [65], Halliwell [66], Kiefer [6], Kiefer and Sandhöfer [67], and Wiltshire [68]. A discussion in the general context of cosmology can be found in Montani et al. [69]. Supersymmetric quantum cosmology is discussed at depth in Moniz [70]. An introduction to loop quantum cosmology is Bojowald [71]. A comparison of standard quantum cosmology with loop quantum cosmology can be found in Bojowald et al. [72].

Quantum cosmology is usually discussed for homogeneous models (the models are then called minisuperspace models). The simplest case is to assume also isotropy. Then, the line element for the classical spacetime metric is given by

$$ds^2 = -N(t)^2 dt^2 + a(t)^2 d\Omega_3^2,$$

where $d\Omega_3$ is the line element of a constant curvature space with curvature index $k = 0, \pm 1$. In order to consider a matter degree of freedom, a homogeneous scalar field $\phi$ with potential $V(\phi)$ is added.

In this setting, the momentum constraints (22) are identically fulfilled. The Wheeler-DeWitt equation (21) becomes a two-dimensional partial differential equation for a wave function $\psi(a, \phi)$,

$$\left(\frac{\hbar^2 \kappa^2}{12} \frac{\partial}{\partial a} a \frac{\partial}{\partial a} - \frac{\hbar^2}{2} \frac{\partial^2}{\partial \phi^2} + a^6 \left( V(\phi) + \frac{\Lambda}{\kappa^2} \right) - \frac{3k \alpha^4}{\kappa^2} \right) \Psi(a, \phi) = 0,$$

where $\Lambda$ is the cosmological constant and $\kappa^2 = 8\pi G$. Introducing $\alpha \equiv \ln a$ (which has the advantage to have a range from $-\infty$ to $+\infty$), one obtains the following equation:

$$\left(\frac{\hbar^2 \kappa^2}{12} \frac{\partial^2}{\partial a^2} - \frac{\hbar^2}{2} \frac{\partial^2}{\partial \phi^2} + e^{6\alpha} \left( V(\phi) + \frac{\Lambda}{\kappa^2} \right) - 3 e^{6\alpha} \frac{k}{\kappa^2} \right) \Psi(a, \phi) = 0.$$

Most versions of quantum cosmology take Einstein's theory as the classical starting point. More general approaches include supersymmetric quantum cosmology [70], string quantum cosmology (2003), noncommutative quantum cosmology (see, e.g., [73]), Hořava-Lifshitz quantum cosmology [74], and third-quantized cosmology [75].

In loop quantum cosmology, features from full loop quantum gravity are imposed on cosmological models [71]. Since one of the main features is the discrete nature of geometric operators, the Wheeler-DeWitt equation (33) is replaced by a difference equation. This difference equation becomes indistinguishable from the Wheeler-DeWitt equation at scales exceeding the Planck length, at least in certain models. Because the difference equation is difficult to solve in general, one makes heavy use of an effective theory [76].

Many features of quantum cosmology are discussed in the limit when the solution of the Wheeler-DeWitt equation assumes a semiclassical or WKB form [66]. This holds, in particular, when the no-boundary proposal or the tunneling proposal is investigated for concrete models, see below. It has even been suggested that the wave function of the universe...
be interpreted only in the WKB limit, because only then a time parameter and an approximate (functional) Schrödinger equation are available [77].

This can lead to a conceptual confusion. Implications for the meaning of the quantum cosmological wave functions should be derived as much as possible from exact solutions. This is because the WKB approximation breaks down in many interesting situations, even for a universe of macroscopic size. One example is a closed Friedmann universe with a massive scalar field [78]. The reason is the following. For a classically recollapsing universe, one must impose the boundary condition that the wave function goes to zero for large scale factors, $\Psi \to 0$ for large $a$. As a consequence, narrow wave packets do not remain narrow because of the ensuing scattering phase shifts of the partial waves (that occur in the expansion of the wave function into basis states) from the turning point. The correspondence to the classical model can only be understood if the quantum-to-classical transition in the sense of decoherence (see below) is invoked.

Another example is the case of classically chaotic cosmologies, see, for example, Calzetta and Gonzalez [79] and Cornish and Shellard [80]. Here, one can see that the WKB approximation breaks down in many situations. This is, of course, a situation well known from quantum mechanics. One of the moons of the planet Saturn, Hyperion, exhibits chaotic rotational motion. Treating it quantum mechanically, one recognizes that the semiclassical approximation breaks down and that Hyperion is expected to be in an extremely nonclassical state of rotation (in contrast to what is observed). This apparent conflict between theory and observation can be understood by invoking the influence of additional degrees of freedom in the sense of decoherence, see Zurek and Paz [81] and Section 3.3.4.3 in [82]. The same mechanism should cure the situation for classically chaotic cosmologies [83].

4. The Problem of Time

The problem of time is one of the major conceptual issues in the search for a quantum theory of gravity [84–86]. As was already emphasized above, time is treated differently in quantum theory and in general relativity. Whereas it is absolute in the first case, it is dynamical in the second. This is the reason why GR is background independent—a major feature to consider in the quantization of gravity.

Background independence is only part of the problem of time. If one applies the standard quantization rules to GR, spacetime disappears and only space remains. This can be understood in simple terms. Classically, spacetime corresponds to what is a particle trajectory in mechanics. Upon quantization, the trajectory vanishes and only the position remains. In GR, spacetime vanishes and only the three-metric remains. Time has disappeared.

This is explicitly seen by the timeless nature of the Wheeler-DeWitt equation (21) or the analogous equation in loop quantum gravity. How can one interpret such a situation? Equation (21) results from a classical constraint in which all momenta occur quadratically. If one could reformulate this constraint in a way where one canonical momentum appears linearly, its quantized form would exhibit a Schrödinger-type equation. Concretely, one would have classically

$$\mathcal{P}_A + A = 0,$$

where $\mathcal{P}_A$ is a momentum for which one can solve the constraint; $A$ simply stands for the remaining terms. Upon quantization one obtains

$$\hat{h} A \frac{\delta \Psi}{\delta q_A} = \hat{A} \Psi.$$

This is of a Schrödinger form. The dynamics of this equation is in general inequivalent to the Wheeler-DeWitt equation. Usually, $A$ is referred to as a “physical Hamiltonian” because it actually describes an evolution in a physical parameter, namely, the coordinate $q_A$ conjugate to $P_A$.

There are many problems with such a “choice of time before quantization” [86]. The constraints of GR cannot be put globally into the form (35) [87]. In most cases, the operator $A$ cannot be defined rigorously. It has therefore been suggested to introduce dust matter with the sole purpose to define a standard of time [88]. More recently, a massless scalar field is used in loop quantum cosmology for this purpose, compare Ashtekar and Singh [89]. One can also define such an internal time from an electric field [90]. At an effective level, one can choose different local internal times within the same model [91]. It is, however, not very satisfactory to adopt the concept of time to the particular model under consideration.

Most of these discussions make use of the Schrödinger picture of quantum theory. An interesting perspective on the problem of time using the Heisenberg picture is presented in Rovelli [92]. It leads to the notion of an “evolving constant of motion”, which unifies the intuitive idea of an evolution with the timelessness of quantum gravity; the evolution proceeds with respect to a physical “clock” variable. An explicit construction of evolving constants in a concrete non-trivial model can be found, for example, in Montesinos et al. [93].

A recent approach to treat the problem of time is shape dynamics, see Barbour et al. [94] and the references therein. In this approach, three-dimensional conformal invariance plays the central role. The configuration space is conformal superspace (the geometrodynamics shape space) times $\mathbb{R}^+$, the second part coming from the volume of three-space. Spacetime foliation invariance at the classical level has been lost. In shape dynamics, the variables are naturally separated into dimensionless true degrees of freedom and a single variable that serves the role of time. This approach is motivated by a similar approach in particle mechanics [95], see also Barbour [96] and Anderson [97]. The consequences of this for the quantum version of shape dynamics have still to be explored.

Independent of these investigations addressing the concept of time in full quantum gravity, it is obvious that the limit of quantum field theory in curved spacetime (in which time as part of spacetime exists) must be recovered in an appropriate limit. How this is achieved is not clear in all of
the above approaches. It is most transparent for the Wheeler-DeWitt equation, as I will be briefly explaining now.

In the semiclassical approximation to the Wheeler-DeWitt equation, one starts with the ansatz

\[ \psi[h_{ab}] = C[h_{ab}] e^{i\int S[h_{ab}] |\psi[h_{ab}]|} \]  

(37)

and performs an expansion with respect to the inverse Planck mass squared \( m_p^{-2} \). This is inserted into (21) and (22), and consecutive orders in this expansion are considered, see Kiefer and Singh [98], Bertoni et al. [99], and Barvinsky and Kiefer [100]. This is close to the Born-Oppenheimer (BO) approximation scheme in molecular physics. In (37), \( h_{ab} \) denotes again the three-metric, and the Dirac bra-ket notation refers to the nongravitational fields, for which the usual Hilbert-space structure is assumed.

The highest orders of the BO scheme lead to the following picture. One evaluates the "matter wave function" \( |\psi[h_{ab}]| \) along a solution of the classical Einstein equations, \( h_{ab}(x,t) \), which corresponds to a chosen solution \( S[h_{ab}] \) of the Hamilton-Jacobi equations. One can then define

\[ \dot{h}_{ab} = NG_{abcd} \frac{\delta S}{\delta h_{cd}} + 2D_{(a}N_{b)}, \]  

(38)

where \( N \) is the lapse function and \( N^a \) is the shift vector; their choice reflects the chosen foliation and coordinate assignment. In this way, one can recover a classical spacetime as an approximation, a spacetime that satisfies Einstein’s equations in this limit. The time derivative of the matter wave function is then defined by

\[ \frac{\partial}{\partial t} |\psi(t)| := \int d^3x \dot{h}_{ab}(x,t) \frac{\delta}{\delta h_{ab}(x)} |\psi[h_{ab}]|, \]  

(39)

where the notation \( |\psi(t)| \) means \( |\psi[h_{ab}]| \) evaluated along the chosen spacetime with the chosen foliation. This leads to a functional Schrödinger equation for quantized matter fields in the chosen external classical gravitational field,

\[ \dot{\tilde{H}}^m |\psi(t)| = \tilde{H}^m |\psi(t)|, \]  

(40)

\[ \tilde{H}^m := \int d^3x \left\{ N(x) \tilde{\Pi}_m^m(x) + N^a(x) \tilde{\Pi}_m^a(x) \right\}, \]  

(41)

where \( \tilde{H}^m \) denotes the matter-field Hamiltonian in the Schrödinger picture, which depends parametrically on the (generally nonstatic) metric coefficients of the curved spacetime background recovered from \( S[h_{ab}] \). The “WKB time” \( t \) controls the dynamics in this approximation.

In the semiclassical limit, one thus finds a Schrödinger equation of the form (36), without using an artificial scalar field or dust matter that is often introduced with the sole purpose of defining a time at the exact level, compare (36). From an empirical point of view, nothing more is needed, because we do not have any access so far to a regime where the semiclassical approximation is not valid.

In this limit, standard quantum theory with all its machinery (Hilbert space, probability interpretation) emerges, and it is totally unclear whether this machinery is needed beyond the semiclassical approximation.

Proceeding to the next order of the \( m_p^{-2} \)-expansion, one can derive quantum gravitational corrections to the functional Schrödinger equation (40). These may, in principle, lead to observable contributions to the anisotropy spectrum of the cosmic microwave background (CMB) anisotropy spectrum [101, 102], although they are at present too tiny.

A major conceptual issue concerns the arrow of time [103]. Although our fundamental laws, as known so far, are time-reversal invariant (or a slight generalization thereof), there are classes of phenomena that exhibit a definite temporal direction. This is expressed by the Second Law of thermodynamics. Is there a hope that the origin of this irreversibility can be found in quantum gravity?

Before addressing this possibility, let us estimate how special our Universe really is; that is, how large its entropy is compared with its maximal possible entropy. Roger Penrose has pointed out that the maximal entropy for the observable Universe would be obtained if all its matters were assembled into one black hole [104]. Taking the most recent observational data, this gives the entropy [105]

\[ S_{\text{max}} \approx 1.8 \times 10^{121}. \]  

(42)

(Here and below, \( k_B = 1 \). This may not yet be the maximal possible entropy. Our Universe exhibits currently an acceleration caused by a cosmological constant \( \Lambda \) or a dynamical dark energy. If it was caused by \( \Lambda \), it would expand forever, and the entropy in the far future would be dominated by the entropy of the cosmological event horizon. This entropy is called the “Gibbons-Hawking entropy” [106] and leads to [105]

\[ S_{\text{GH}} = \frac{3\pi}{\Lambda \ell_p^2} \approx 2.9 \times 10^{122}, \]  

(43)

which is about one order of magnitude higher than (42).

Following the arguments in Penrose [104], the "probability" for our Universe can then be estimated as

\[ \frac{\exp(S)}{\exp(S_{\text{max}})} = \frac{\exp(3.1 \times 10^{104})}{\exp(2.9 \times 10^{122})} \approx \exp(-2.9 \times 10^{122}). \]  

(44)

Our Universe is thus very special indeed. It is much more special than what would be estimated from the anthropic principle.

Turning to quantum gravity, the question arises how one can derive an arrow of time from a framework that is fundamentally timeless. A final answer is not yet available, but various ideas exist [103, 105]. The Wheeler-DeWitt equation (21) does not contain any external time parameter, but one can define an intrinsic time from the hyperbolic nature of this equation, which is entirely constructed from the three-metric [107]. In quantum cosmology, this intrinsic time is the scale factor, compare (33) and (34). If one adds small inhomogeneous degrees of freedom to a Friedmann model
The Wheeler-DeWitt equation turns out to be of the form

$$\hat{H} \Psi = \left( \frac{2\pi G \hbar^2}{3} \frac{\partial^2}{\partial \alpha^2} + \sum_i \left[ -\frac{\hbar^2}{2} \frac{\partial^2}{\partial x_i^2} + \frac{\mathcal{V}(\alpha, x_i)}{0 \text{ for } \alpha \to -\infty} \right] \right) \Psi = 0,$$

(45)

where the \{x_i\} denotes the inhomogeneous degrees of freedom as well as neglected homogeneous ones; \mathcal{V}(\alpha, x_i) are the corresponding potentials. One recognizes immediately that this Wheeler-DeWitt equation is hyperbolic with respect to the intrinsic time \alpha. Initial conditions are thus most naturally formulated with respect to constant \alpha.

The important property of (45) is that the potential becomes small for \alpha \to -\infty (where the classical singularities would occur) but complicated for increasing \alpha. In the general case (not restricting to small inhomogeneities), this may be further motivated by the BKL conjecture according to which spatial gradients become small near a spacelike singularity [109]. The Wheeler-DeWitt equation thus possesses an asymmetry with respect to "intrinsic time" \alpha. One can in particular impose the simple boundary condition [103]

$$\Psi_{\alpha \to -\infty} = \psi_0(\alpha) \prod_i \psi_i(x_i),$$

(46)

which means that the degrees of freedom are initially not entangled. They will become entangled for increasing \alpha because then the coupling in the potential between \alpha and the \{x_i\} becomes important. This leads to a positive entanglement entropy. In the semiclassical limit, where a WKB time \tau can be defined, there is a correlation between \tau and \alpha which can explain the standard Second Law, at least in principle. In this scenario, entropy increase is correlated with scale-factor increase. Therefore, the arrow of time would formally reverse at the turning point of a classically recollapsing universe, although in a quantum scenario the classical evolution would end there and no transition to a recollapsing phase could ever be observed [110]. Whether a boundary condition of the form (46) follows as a necessary requirement from the mathematical structure of the full theory or not is an open issue. Alternative ideas to the recovery of the arrow of time in the framework of quantum geometrodynamics. Its principle features should also be applicable, with some modifications, to loop quantum gravity. But what about string theory?

String theory contains general relativity. As we have seen above, the quantum equation that gives back Einstein’s equation in the semiclassical limit is the Wheeler-DeWitt equation (21). One would thus expect that the Wheeler-DeWitt equation can be recovered from string theory in the limit where the string constant \alpha' is small. Unfortunately, this has not been shown so far in any explicit manner.

It is clear that the problem of time is the same in string theory. New insights may be obtained, in addition, for the concept of space. This is best seen in the context of the AdS/CFT correspondence, see, for example, Maldacena [113] for a review. In short words, this correspondence states that nonperturbative string theory in a background spacetime which is asymptotically anti-de Sitter (AdS) is dual to a conformal field theory (CFT) defined in a flat spacetime of one fewer dimension (the boundary of the background spacetime). What really corresponds here are certain matrix elements and symmetries in the two theories; an equivalence at the level of the quantum states is not shown. This correspondence can be considered as an intermediate step towards a background-independent formulation of string theory, because the background metric enters only through boundary conditions at infinity, compare Blau and Theisen [114].

The AdS/CFT correspondence can also be interpreted as a realization of the “holographic principle,” which states that the information of a gravitating system is located on the boundary of this system; the most prominent example is the expression (12) for the entropy, which is given by the surface of the event horizon. In a particular case, laws including gravity in \(d = 3\) are equivalent to laws excluding gravity in \(d = 2\). In a loose sense, space has then vanished, too [115].

Our discussion of the problem of time presented in this section is far from complete. There exist, for example, interesting suggestions of a thermodynamical origin of time in a gravitational context [116]. It has been shown that statistical mechanics of generally covariant quantum theories can be developed without a preferred time and thus without a preferred notion of temperature [117].

5. Singularity Avoidance and Boundary Conditions

As we have mentioned at the beginning, classical GR predicts the occurrence of singularities [14]. What can the above approaches say about their fate in the quantum theory? In order to discuss this, one first has to agree about a definition of singularity avoidance in quantum gravity. Since such an agreement does not yet exist, one has to study heuristic expectations. Already DeWitt [47] has speculated that a classical singularity is avoided if the wave function vanishes at the corresponding region in configuration space,

$$\Psi \left[ (3) \mathcal{G}^{\text{sing}} \right] = 0.$$

(47)

One example where this condition can be implemented concerns the fate of a singularity called "big brake" [118]. This occurs for an equation of state of the form \(p = A/\rho, A > 0\), called "anti-Chaplygin gas." For a Friedmann universe with scale factor \(a(t)\) and a scalar field \(\phi(t)\), this equation of state can be realized by the potential

$$V(\phi) = V_0 \left( \sinh \left( \sqrt{3} \kappa^2 |\phi| \right) - \frac{1}{\sinh \left( \sqrt{3} \kappa^2 |\phi| \right)} \right);$$

(48)

$$V_0 = \sqrt{A/4}.$$
where $\kappa^2 = 8\pi G$. The classical dynamics develops a pressure singularity (only $a(t)$ becomes singular) and comes to an abrupt halt in the future “big brake.”

The Wheeler-DeWitt equation for this model reads

$$\frac{\hbar^2}{2} \left( \frac{\kappa^2}{6} \frac{\partial^2}{\partial \alpha^2} - \frac{\partial^2}{\partial \phi^2} \right) \Psi(\alpha, \phi) + V_0 e^{\alpha} \left( \sinh \left( \sqrt{3} \kappa^2 |\phi| \right) - \frac{1}{\sinh \left( \sqrt{3} \kappa^2 |\phi| \right)} \right)$$

$$\times \Psi(\alpha, \phi) = 0,$$

where $\alpha = \ln a$ and Laplace-Beltrami factor ordering has been used. The vicinity of the big-brake singularity is the region of small $\phi$; we can therefore use the approximation

$$\frac{\hbar^2}{2} \left( \frac{\kappa^2}{6} \frac{\partial^2}{\partial \alpha^2} - \frac{\partial^2}{\partial \phi^2} \right) \Psi(\alpha, \phi) + \overline{V}_0 e^{\alpha} \Psi(\alpha, \phi) = 0,$$

where $\overline{V}_0 = V_0/3\kappa^2$.

It was shown in Kamenshchik et al. [118] that all normalizable solutions are of the form

$$\Psi(\alpha, \phi) = \sum_{k=1}^{\infty} A(k) K_0 \left( \frac{1}{\sqrt{6}} \kappa^2 \right)$$

$$\times \left( \frac{V_a}{k} \right) e^{-V_a/k|\phi|} L_1^{(1)} \left( \frac{2 V_a}{k} |\phi| \right),$$

where $K_0$ is a Bessel function, $L_1^{(1)}$ are Laguerre polynomials, and $V_a \equiv \overline{V}_0 e^{\alpha}$. They all vanish at the classical singularity. This model therefore implements DeWitt’s criterium above. The same holds in more general situations of this type [119]. In a somewhat different approach, the big brake singularity is not avoided [120]. Singularity avoidance in the case of a “big rip,” which can occur for phantom fields, is discussed in Dąbrowski et al. [121].

The vanishing of the wave function at the classical singularity plays a role in the treatment of supersymmetric quantum cosmological billiards [122]. In $D = 11$ supergravity, one can employ near a spacelike singularity such a description based on the Kac-Moody group $E_{10}$ and derive the corresponding Wheeler-DeWitt equation. It was found there, too, that $\Psi \to 0$ near the singularity and that DeWitt’s criterium is fulfilled. Singularity avoidance for the Wheeler-DeWitt equation is also achieved when the Bohm interpretation is used [123]. Singularity avoidance was also discussed using “wavelet quantization” instead of canonical quantization [124]; there the occurrence of a repulsive potential was found.

Singularity avoidance also occurs in the framework of loop quantum cosmology, although in a somewhat different way [71]. The big-bang singularity can be avoided by solutions of the difference equation that replaces the Wheeler-DeWitt equation. The avoidance can also be achieved by the occurrence of a bounce in the effective Friedmann equations. These results strongly indicate that singularities are avoided in loop quantum cosmology, although no general theorems exist [89]. Possible singularity avoidance can also be discussed for black-hole singularities and for naked singularities [125].

The question of singularity avoidance is closely connected with the role of boundary conditions in quantum cosmology. Let us thus here have a brief look at the conceptual side of this issue.

The Wheeler-DeWitt equation (34) is of hyperbolic nature with respect to $\alpha \equiv \ln a$. It makes thus sense to specify the wave function and its derivative at constant $\alpha$. In the case of an open universe, this is fine. In the case of a closed universe, however, one has to impose the boundary condition that the wave function vanishes at $\alpha \to \infty$. The existence of a solution in this case may then lead to a restriction on the allowed values for the parameters of the model, such as the cosmological constant or the mass of a scalar field, or may allow no solution at all.

Another type of boundary conditions makes use of the path-integral formulation (15). In 1982, Hawking formulated his “no-boundary condition” for the wave function [126], which was then elaborated in Hartle and Hawking [127] and Hawking [128]. The wave function is given by the Euclidean path integral

$$\Psi[h_{ab}, \Phi, \Sigma] = \int_M DgD\Phi e^{-S_E[g_{ab}, \Phi]}.$$  (52)

The sum over $M$ expresses the sum over all four manifolds with measure $\nu(M)$ (which actually cannot be performed). The no-boundary condition states that—apart from the boundary where the three metric $h_{ab}$ is specified—there is no other boundary on which initial conditions have to be specified.

Originally, the hope was entertained that the no-boundary proposal leads to a unique wave function (or a small class of wave functions) and that the classical big-bang singularity is smoothed out by the absence of the initial boundary. It was, however, later realized that there are, in fact, many solutions, and that the integration has to be performed over complex metrics (see, e.g., [129, 130]). Moreover, the path integral can usually only be evaluated in a semiclassical limit (using the saddle-point approximation), so it is hard to make a general statement about singularity avoidance.

In a Friedmann model with scale factor $a$ and a scalar field $\phi$ with a potential $V(\phi)$, the no-boundary condition gives the semiclassical solution [128]

$$\psi_{NB} \propto (a^2 V(\phi) - 1)^{1/4} e^{\frac{1}{3V(\phi)}} \times \cos \left( \frac{\pi}{4} - \frac{2V_{(\phi)}}{3V(\phi)} \right).$$

We note that the no-boundary wave function is always real. The form (53) corresponds to the superposition of an expanding and a recollapsing universe.

Another prominent boundary condition is the tunneling proposal (see [131] and the references therein). It was originally defined by the choice of taking “outgoing” solutions
at singular boundaries of superspace. The term “outgoing” is somewhat misleading, because the sign of the imaginary unit has no absolute meaning in the absence of external time [107]. What one can say is that a complex solution is chosen, in contrast to the real solution found from the no-boundary proposal. In the same model, one obtains for the tunnelling proposal instead of (53) the expression

\[ \psi_T \propto \left(a^2V(\phi) - 1\right)^{-1/4} \exp \left(-\frac{1}{3V(\phi)}\right) \times \exp \left(-\frac{i}{3V(\phi)} \left(a^2V(\phi) - 1\right)^{3/2}\right). \]  

(54)

Considering the conserved Klein-Gordon type of current

\[ j = \frac{i}{2} \left(\psi^* \nabla \psi - \psi \nabla \psi^*\right), \quad \nabla j = 0, \]  

(55)

where \( \nabla \) denotes the derivatives in minisuperspace, one finds for a WKB solution of the form \( \psi = C \exp(\mathcal{S}) \) the expression

\[ j = -|C|^2 \nabla \mathcal{S}. \]  

(56)

The tunnelling proposal states that this current should point outwards at large \( a \) and \( \phi \) (provided, of course, that \( \psi \) is of WKB form there). If \( \psi \) was real (as is the case in the no-boundary proposal), the current would vanish. Again, the wave function is of semiclassical form. It has been suggested that it may only be interpreted in this limit [77]. Other boundary conditions include the SIC proposal put forward by Conradi and Zeh [132].

An interesting application of these boundary conditions concerns the prediction of an inflationary phase for the early universe. Here, it seems that the tunnelling wave function favours such a phase, while the no-boundary condition disfavors it [133]. In the context of the string theory landscape, the no-boundary proposal was applied to inflation, even leading to a prediction for the spectral index of the CMB spectrum [134].

6. General Interpretation of Quantum Theory

Quantum cosmology can shed some light on the problem of interpreting quantum theory in general. After all, a quantum universe possesses by definition no external classical measuring agency. It is not possible to apply the Copenhagen interpretation, which presumes the existence of classical realms from the outset.

Since all the approaches discussed above preserve the linearity of the formalism, the superposition principle remains valid and with it the measurement problem. In most investigations, the Everett interpretation [135] is applied to quantum cosmology. In this interpretation, all components of the wave function are equally real. This view is already reflected by the words of Bryce DeWitt in DeWitt [47]:

“Everett’s view of the world is a very natural one to adopt in the quantum theory of gravity, where one is accustomed to speak without embarrassment of the “wave function of the universe.” It is possible that Everett’s view is not only natural but essential.”

Alternative interpretations include the Bohm interpretation, see Pinto-Neto et al. [123] and the references therein.

Quantum cosmology (and quantum theory in general) can be consistently interpreted if the Everett interpretation is used together with the process of decoherence [82]. Decoherence is the irreversible emergence of classical properties from the unavoidable interaction with the “environment” (meaning irrelevant or negligible degrees of freedom in configuration space). In quantum cosmology, the irrelevant degrees of freedom include tiny gravitational waves and density fluctuations [136]. Their interaction with the scale factor and global matter fields transform them into variables that behave classically, see, for example, Kiefer [137] and Barvinsky et al. [138]. An application of decoherence to the superposition of triad superpositions in loop quantum cosmology can be found in Kiefer and Schell [139].

Once the “background variables” such as \( a \) or \( \phi \) have been rendered classical by decoherence, the question arises what happens to the primordial quantum fluctuations out of which—according to the inflationary scenario—all structure in the Universe emerges. Here, again, decoherence plays the decisive role, see Kiefer and Polarski [140] and the references therein. The quantum fluctuations assume classical behaviour by the interaction with small perturbations that arise either from other fields or from a self-coupling interaction. The decoherence time typically turns out to be of the order

\[ t_d \sim \frac{H_1}{g}, \]  

(57)

where \( g \) is a dimensionless coupling constant of the interaction with the other fields causing decoherence and \( H_1 \) is the Hubble parameter of inflation. The ensuing coarse graining brought about by the decohering fields causes an entropy increase for the primordial fluctuations [141]. All of this is in accordance with current observations of the CMB anisotropies.

These considerations can, in principle, be extended to the concept of the multiverse, as it arises, for example, from the string landscape or from inflation (see e.g., [63, 142]), but additional problems emerge (such as the measure problem) that have so far not been satisfactorily been dealt with.

As we have mentioned in Section 1, the linearity of quantum theory may break down if gravity becomes important. In this case, a collapse of the wave function may occur, which destroys all the components of the wave function except one, see, for example, Landau et al. [143]. But as long as there is no empirical hint for the breakdown of linearity, the above scenario provides a consistent and minimalistic (in the sense of mathematical structure) picture of the quantum-to-classical transition in quantum cosmology.

In addition to the many formal and mathematical problems, conceptual problems form a major obstacle for the final construction of a quantum theory of gravity and its application to cosmology. They may, however, also provide the key for the construction of such a theory. Whether or when this will happen is, however, an open question.
References


