INSTITUTE OF PHYSICS PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **21** (2004) R53–R152 PII: S0264-9381(04)78804-8

TOPICAL REVIEW

Background independent quantum gravity: a status report

Abhay Ashtekar[1,2,3] **and Jerzy Lewandowski**[1,2,3,4]

[1] Physics Department, 104 Davey, Penn State, University Park, PA 16802, USA
[2] Max-Planck-Institut für Gravitationsphysik, Albert Einstein Institut, 14476 Golm, Germany
[3] Erwin Schrödinger Institute, Boltzmanngasse 9, 1090 Vienna, Austria
[4] Instytut Fizyki Teoretycznej, Uniwersytet Warszawski, ul. Hoża 69, 00-681 Warsaw, Poland

Received 5 April 2004
Published 9 July 2004
Online at stacks.iop.org/CQG/21/R53
doi:10.1088/0264-9381/21/15/R01

Abstract

The goal of this review is to present an introduction to loop quantum gravity—a background-independent, non-perturbative approach to the problem of unification of general relativity and quantum physics, based on a quantum theory of geometry. Our presentation is pedagogical. Thus, in addition to providing a bird's eye view of the present status of the subject, the review should also serve as a vehicle to enter the field and explore it in detail. To aid non-experts, very little is assumed beyond elements of general relativity, gauge theories and quantum field theory. While the review is essentially self-contained, the emphasis is on communicating the underlying ideas and the significance of results rather than on presenting systematic derivations and detailed proofs. (These can be found in the listed references.) The subject can be approached in different ways. We have chosen one which is deeply rooted in well-established physics and also has sufficient mathematical precision to ensure that there are no hidden infinities. In order to keep the review to a reasonable size, and to avoid overwhelming non-experts, we have had to leave out several interesting topics, results and viewpoints; this is meant to be an introduction to the subject rather than an exhaustive review of it.

PACS numbers: 04.60.Pp, 04.60.Ds, 04.60.Nc, 03.65.Sq

Contents

1. Introduction

This section is divided into three parts. In the first, we outline the general, conceptual viewpoint that underlies loop quantum gravity; in the second, we recall some of the central physical problems of quantum gravity; and in the third, we summarize the progress that has been made in addressing these issues and sketch the organization of the paper.

1.1. The viewpoint

In this approach, one takes the central lesson of general relativity seriously: gravity *is* geometry whence, in a fundamental theory, there should be no background metric. In quantum gravity, geometry and matter should *both* be 'born quantum mechanically'. Thus, in contrast to approaches developed by particle physicists, one does not begin with quantum matter on a background geometry and use perturbation theory to incorporate quantum effects of gravity. There *is* a manifold but no metric, or indeed any other fields, in the background[5].

At the classical level, Riemannian geometry provides the appropriate mathematical language to formulate the physical, kinematical notions as well as the final dynamical equations of modern gravitational theories. This role is now taken by *quantum* Riemannian geometry, discussed in sections 4 and 5. In the classical domain, general relativity stands out as the best available theory of gravity, some of whose predictions have been tested to an amazing accuracy, surpassing even the legendary tests of quantum electrodynamics. Therefore, it is natural to ask: *Does quantum general relativity, coupled to suitable matter* (or supergravity, its supersymmetric generalization) *exist as a consistent theory non-perturbatively?*

In the particle physics circles, the answer is often assumed to be in the negative, not because there is concrete evidence against non-perturbative quantum gravity, but because of an analogy to the theory of weak interactions. There, one first had a 4-point interaction model due to Fermi which works quite well at low energies but fails to be renormalizable. Progress occurred not by looking for non-perturbative formulations of the Fermi model but by replacing the model with the Glashow–Salam–Weinberg renormalizable theory of electro-weak interactions, in which the 4-point interaction is replaced by $W^{\pm}$ and Z propagators. It is often assumed that perturbative non-renormalizability of quantum general relativity points in a similar direction. However, this argument overlooks the crucial fact that, in the case of general relativity, there is a qualitatively new element. Perturbative treatments pre-suppose that spacetime can be assumed to be a continuum *at all scales* of interest to physics under consideration. This appears to be a safe assumption in theories of electro-weak and strong interactions. In the gravitational case, on the other hand, the scale of interest is given by the Planck length $\ell_{\rm Pl}$ and there is no physical basis to *pre-suppose* that the continuum picture should be valid down to that scale. The failure of the standard perturbative treatments may be largely due to this grossly incorrect assumption and a non-perturbative treatment which correctly incorporates the physical micro-structure of geometry may well be free of these inconsistencies.

Note that, even if quantum general relativity did exist as a mathematically consistent theory, there is no *a priori* reason to assume that it would be the 'final' theory of all known physics. In particular, as is the case with classical general relativity, while requirements of background independence and general covariance do restrict the form of interactions between gravity and matter fields and among matter fields themselves, the theory would not have a built-in principle which *determines* these interactions. Put differently, such a theory would not be a satisfactory candidate for unification of all known forces. However, just as general relativity has had powerful implications in spite of this limitation in the classical domain, quantum general relativity should have qualitatively new predictions, pushing further the existing frontiers of physics. Indeed, unification does not appear to be an essential criterion for usefulness of a theory even in other interactions. QCD, for example, is a powerful theory

[5] In $2+1$ dimensions, although one begins in a completely analogous fashion, in the final picture one can get rid of the background manifold as well. Thus, the fundamental theory can be formulated combinatorially [2, 35]. To achieve this goal in $3+1$ dimensions, one needs a much better understanding of the theory of (intersecting) knots in three dimensions.

even though it does not unify strong interactions with electro-weak ones. Furthermore, the fact that we do not yet have a viable candidate for the grand unified theory does not make QCD any less useful.

Finally, the quantum theory of geometry provides powerful tools to do quantum physics in the absence of a background spacetime. Being kinematical, it is not rigidly tied to general relativity (or supergravity) and may well be useful also in other approaches to quantum gravity.

1.2. Physical questions of quantum gravity

Approaches to quantum gravity face two types of issues: problems that are 'internal' to individual approaches and problems that any approach must face. Examples of the former are: incorporation of physical—rather than half-flat—gravitational fields in twistor theory; mechanisms for breaking of supersymmetry and dimensional reduction in string theory; and issues of spacetime covariance in the canonical approach. In this subsection, we will focus on the second type of issue by recalling some of the long standing issues that *any* satisfactory quantum theory of gravity should address.

Big-bang and other singularities. It is widely believed that the prediction of a singularity, such as the big-bang of classical general relativity, is primarily a signal that the physical theory has been pushed beyond the domain of its validity. A key question to any quantum gravity theory, then is: what replaces the big-bang? Are the classical geometry and the continuum picture only approximations, analogous to the 'mean (magnetization) field' of ferro-magnets? If so, what are the microscopic constituents? What is the spacetime analogue of a Heisenberg quantum model of a ferro-magnet? When formulated in terms of these fundamental constituents, is the evolution of the *quantum* state of the universe free of singularities? General relativity predicts that the spacetime curvature must grow unboundedly as we approach the big-bang or the big-crunch but we expect the quantum effects, ignored by general relativity, to intervene, making quantum gravity indispensable before infinite curvatures are reached. If so, what is the upper bound on curvature? How close to the singularity can we 'trust' classical general relativity? What can we say about the 'initial conditions', i.e., the quantum state of geometry and matter that correctly describes the big-bang? If they have to be imposed externally, is there a *physical* guiding principle?

Black holes. In the early 1970s, using imaginative thought experiments, Bekenstein [118] argued that black holes must carry an entropy proportional to their area. About the same time, Bardeen, Carter and Hawking (BCH) showed that black holes in equilibrium obey two basic laws, which have the same form as the zeroth and the first laws of thermodynamics, provided one equates the black-hole surface gravity κ to some multiple of the temperature T in thermodynamics and the horizon area $a_{\rm hor}$ to a corresponding multiple of the entropy S [119]. However, at first this similarity was thought to be only a formal analogy because the BCH analysis was based on *classical* general relativity and simple dimensional considerations show that the proportionality factors must involve Planck's constant $\hbar$. Two years later, using quantum field theory on a black-hole background spacetime, Hawking [120] showed that black holes in fact radiate quantum mechanically as though they are black bodies at temperature $T = \hbar\kappa/2\pi$. Using the analogy with the first law, one can then conclude that the black-hole entropy should be given by $S_{\rm BH} = a_{\rm hor}/4G\hbar$. This conclusion is striking and deep because it brings together the three pillars of fundamental physics—general relativity, quantum theory and statistical mechanics. However, the argument itself is rather a hotchpotch of classical and semi-classical ideas, reminiscent of the Bohr theory of the atom. A natural question then is as follows: what is the analogue of the more fundamental, Pauli–Schrödinger theory of the

hydrogen atom? More precisely, what is the statistical mechanical origin of the black-hole entropy? What is the nature of a quantum black hole and what is the interplay between the quantum degrees of freedom responsible for entropy and the exterior curved geometry? Can one derive the Hawking effect from first principles of quantum gravity? Is there an imprint of the classical singularity on the final quantum description, e.g., through 'information loss'?

Planck scale physics and the low energy world. In general relativity, there is no background metric, no inert stage on which dynamics unfolds. Geometry itself is dynamical. Therefore, as indicated above, one expects that a fully satisfactory quantum gravity theory would also be free of a background spacetime geometry. However, of necessity, a background-independent description must use physical concepts and mathematical tools that are quite different from those of the familiar, low energy physics. A major challenge then is to show that this low energy description does arise from the pristine, Planckian world in an appropriate sense, bridging the vast gap of some 16 orders of magnitude in the energy scale. In this 'top-down' approach, does the fundamental theory admit a 'sufficient number' of semi-classical states? Do these semi-classical sectors provide enough of a background geometry to anchor low energy physics? Can one recover the familiar description? Furthermore, can one pinpoint why the standard 'bottom-up' perturbative approach fails; that is, what is the essential feature which makes the fundamental description mathematically coherent but is absent in the standard perturbative quantum gravity?

There are of course many more challenges: the issue of time, of measurement theory and the associated questions of interpretation of the quantum framework, the issue of diffeomorphism invariant observables and practical methods of computing their properties, practical methods of computing time evolution and S-matrices, exploration of the role of topology and topology change, etc. However, it is our view that the three issues discussed in detail are more basic from a physical viewpoint because they are rooted in general conceptual questions that are largely independent of the specific approach being pursued; indeed they have been with us longer than any of the current leading approaches.

1.3. Organization

In recent years, a number of these fundamental physical issues were addressed in loop quantum gravity. These include the following: (i) a natural resolution of the big-bang singularity in homogeneous, isotropic quantum cosmology [103–117]; (ii) a statistical mechanical derivation of the horizon entropy, encompassing astrophysically interesting black holes as well as cosmological horizons [122–141]; and (iii) the introduction of semi-classical techniques to make contact between the background-independent, non-perturbative theory and the perturbative, low energy physics in Minkowski space [142–160]. In addition, advances have been made on the mathematical physics front. In particular, these include the following: (iv) a demonstration that all Riemannian geometric operators have discrete eigenvalues, implying that the spacetime continuum is only an approximation [65–80]; (v) a systematic formulation of quantum Einstein equations in the canonical approach [85–102]; and (vi) the development of spin foam models which provide background-independent path integral formulations of quantum gravity [161–173]. These developments are also significant. For example, in contrast to (v), quantum Einstein's equations are yet to be given a precise mathematical meaning in quantum geometrodynamics—a canonical approach that predates loop quantum gravity by two decades or so—because the products of operators involved are divergent.

All these advances spring from a detailed quantum theory of geometry that was systematically developed in the mid-1990s. This theory is, in turn, an outgrowth of two

developments: (a) formulation of general relativity (and supergravity) as a dynamical theory of connections, with the same phase space as in Yang–Mills theories [12–16]; and, (b) heuristic but highly influential treatments of quantum theories of connections in terms of loops [33–38][6]. In this review, we will first provide a brief but self-contained and pedagogical introduction to quantum geometry and then discuss its applications to problems mentioned above.

The review is organized as follows. In section 2 we recall connection formulations of general relativity. (Readers who are primarily interested in quantum geometry rather than dynamical issues of general relativity may skip this section in the first reading.) The next four sections present the basics of quantum theory. In section 3 we summarize the overall strategy used in the construction of quantum kinematics; in section 4, we discuss background-independent formulations of general quantum theories of connections; in section 5, the basics of quantum Riemannian geometry and in section 6 the basics of quantum dynamics. These sections are self-contained and the reader is referred to the original papers only for certain proofs, technical subtleties and details that are interesting in their own right but not essential to follow the general approach. Sections 7–9 are devoted to applications of quantum geometry and a summary of current directions, where the treatment is less pedagogical: while the main ideas are spelled out, the reader will have to go through at least some of the original papers to get a thorough working knowledge. Section 10 contains a summary and the outlook.

For simplicity, most of the discussion in the main body of the review is focused on the gravitational field. There is a large body of work on coupling of gauge, fermionic and scalar fields to gravity where the quantum nature of underlying geometry modifies the physics of matter fields in important ways. Appendix A illustrates these issues using the Einstein–Maxwell theory as an example. Appendix B contains a list of symbols which are frequently used in the review.

For a much more detailed review, at the level of a monograph, see [9]. Less pedagogical overviews, at the level of plenary lectures in conferences, can be found in [6, 8].

2. Connection theories of gravity

General relativity is usually presented as a theory of metrics. However, it can also be recast as a dynamical theory of connections[7]. Such a reformulation brings general relativity closer to gauge theories which describe the other three fundamental forces of Nature in the sense that, in the Hamiltonian framework, all theories now share the same kinematics. The difference, of course, lies in dynamics. In particular, while dynamics of gauge theories of other interactions requires a background geometry, that of general relativity does not. Nonetheless, by a suitable modification, one can adapt quantization techniques used in gauge theories to general relativity. We will see in sections 4–6 that this strategy enables one to resolve the functional analytic difficulties which have prevented 'geometrodynamical' approaches to quantum gravity, based on metrics, to progress beyond a formal level.

In this section, we will present a self-contained introduction to connection formulations of general relativity. However, we will not follow a chronological approach but focus instead only on those aspects which are needed in subsequent sections. For a discussion of other issues, see [2, 12–23].

[6] This is the origin of the name 'loop quantum gravity'. Even though loops play no essential role in the theory now, for historical reasons, this name is widely used. The current framework is based on graphs introduced in [40–42].

[7] Indeed, in the late 1940s both Einstein and Schrödinger had recast general relativity as a theory of connections. However, the resulting theory was rather complicated because they used the Levi-Civita connection. Theory simplifies if one uses spin connections instead.

Our conventions are as follows. $\mathcal{M}$ will denote the four-dimensional spacetime manifold which we will assume to be topologically $M \times \mathbb{R}$, equipped with a fixed orientation. For simplicity, in this section we will assume that M is an oriented, compact 3-manifold without boundary. (Modifications required to incorporate asymptotic flatness can be found in [2] and those needed to allow an isolated horizon as an inner boundary can be found in [133, 134].) For tensor fields (possibly with internal indices), we will use Penrose's abstract index notation. The spacetime metric will be denoted by $g_{\mu\nu}$ and will have signature $-, +, +, +$ (or, occasionally, $+, +, +, +$). In the Lorentzian case, spacetime will be assumed to be time-orientable. The torsion-free derivative operator compatible with $g_{\mu\nu}$ will be denoted by ∇ and its curvature tensors will be defined via $R_{\alpha\beta\gamma}{}^{\delta}K_{\delta} = 2\nabla_{[\alpha}\nabla_{\beta]}K_{\gamma}$; $R_{\alpha\beta} = R_{\alpha\beta\gamma}{}^{\beta}$; and $R = g^{\alpha\beta}R_{\alpha\beta}$. For the tetrad formalism, we fix a four-dimensional vector space V equipped with a fixed metric $\bar{\eta}_{IJ}$ of signature $-, +, +, +$ (or $+, +, +, +$), which will serve as the 'internal space'. Orthonormal co-tetrads will be denoted by e^I_α; thus $g_{\alpha\beta} = \bar{\eta}_{IJ}e^I_\alpha e^J_\beta$. In the passage to the Hamiltonian theory, the metric on a spacelike Cauchy surface M will be denoted by q_{ab} and the spatial co-triads will be denoted by e^i_a. Finally, we will often set $k = 8\pi G$ where G is Newton's constant. Due to space limitation, we will focus just on the gravitational part of the action and phase space. For inclusion of matter, see, e.g., [16, 133]; for extension to supergravity, see, e.g., [17]; and for ideas on extension to higher dimensions, see [25].

2.1. Holst's modification of the Palatini action

In the Palatini framework, the basic gravitational variables constitute a pair $\left(e_\mu{}^I, \omega_\mu{}^I{}_J\right)$ of 1-form fields on $\mathcal{M}$ taking values, respectively, in V and in the Lie algebra $so(\bar{\eta})$ of the group $SO(\bar{\eta})$ of the linear transformations of V preserving $\bar{\eta}_{IJ}$. Because of our topological assumptions, the co-frame fields $e_\mu{}^I$ are defined globally; they provide an isomorphism between $T_x\mathcal{M}$ and V at each $x \in \mathcal{M}$. The action is given by

$$\mathbf{S}_{(P)}(e, \omega) = \frac{1}{4k}\int_{\mathcal{M}} \epsilon_{IJKL}e^I \wedge e^J \wedge \Omega^{KL} \tag{2.1}$$

where ϵ_{IJKL} is an alternating tensor on V compatible with $\bar{\eta}_{IJ}$ such that the orientation of $\epsilon_{\alpha\beta\gamma\delta} = \epsilon_{IJKL}e^I_\alpha e^J_\beta e^K_\gamma e^L_\delta$ agrees with that fixed on M and

$$\Omega := \mathrm{d}\omega + \omega \wedge \omega, \tag{2.2}$$

is the curvature of the connection 1-form $\omega_\mu{}^I{}_J$. The co-frame e^I_μ determines a spacetime metric $g_{\mu\nu} = \eta_{IJ}e^I_\mu e^J_\nu$. Thus, in contrast to the more familiar Einstein–Hilbert action, $\mathbf{S}_{(P)}$ depends on an additional variable, the connection ω^{IJ}_μ. However, the equation of motion obtained by varying the action with respect to the connection implies that ω^{IJ}_μ is in fact completely determined by the co-frame:

$$\mathrm{d}e + \omega \wedge e = 0. \tag{2.3}$$

If we now restrict ourselves to histories on which the connection is so determined, $\mathbf{S}_{(P)}$ reduces to the familiar Einstein–Hilbert action,

$$\mathbf{S}_{(P)}(e, \omega(e)) = \frac{1}{2k}\int_{\mathcal{M}} \mathrm{d}^4x\,\sqrt{|\det g|}R. \tag{2.4}$$

where R is the scalar curvature of $g_{\mu\nu}$. Therefore, the equation of motion for the metric is the same as that of the Einstein–Hilbert action.

The action $\mathbf{S}_{(P)}$ is invariant under diffeomorphisms of $\mathcal{M}$ as well as *local* $SO(\bar{\eta})$ transformations

$$(e, \omega) \mapsto (e', \omega') = (b^{-1}e, b^{-1}\omega b + b^{-1}\,\mathrm{d}b). \tag{2.5}$$

It is straightforward but rather tedious to perform a Legendre transform of this action and pass to a Hamiltonian theory [18]. It turns out that the theory has certain second class constraints and, when they are solved, one is led to a triad version of the standard Hamiltonian theory of geometrodynamics; all reference to connection-dynamics is lost. This can be remedied using the following observation: there exists another invariant, constructed from the pair (e, ω), with the remarkable property that its addition to the action does not change equations of motion. The modified action, discussed by Holst [21], is given by[8]

$$\mathbf{S}_{(H)}(e,\omega) = \mathbf{S}_{(P)}(e,\omega) - \frac{1}{2k\gamma}\int_{\mathcal{M}} e^I \wedge e^J \wedge \Omega_{IJ} \tag{2.6}$$

where γ is an arbitrary but fixed number, called the *Barbero–Immirzi* parameter. For applications to quantum theory, it is important to note that γ *cannot be zero*. The purpose of this section is to analyse this action and the Hamiltonian theory emerging from it. In sections 2.2 and 2.3 we will show that the Hamiltonian theory can be naturally interpreted as a background-independent, dynamical theory of connections.

Recall that, in Yang–Mills theories, one can also add a 'topological term' to the action which does not change the classical equations of motion because its integrand can be re-expressed as an exterior derivative of a 3-form. In the present case, while the extra term is not of topological origin, because of the first Bianchi identity it vanishes identically on histories on which (2.3) holds. Therefore the situation is similar in the two theories in some respects: in both cases, the addition of the term does not change the classical equations of motion and it induces a canonical transformation on the classical phase space which fails to be unitarily implementable in the quantum theory. Consequently, the parameter γ is in many ways analogous to the well-known θ parameter in the Yang–Mills theory [83, 84]. Just as the quantum theory has *inequivalent* θ-sectors in the Yang–Mills case, it has *inequivalent* γ-sectors in the gravitational case.

We will conclude this preliminary discussion by exhibiting the symplectic structure in the covariant phase space formulation. Here the phase space $\mathbf{\Gamma}_{\rm cov}$ is taken to be the space of (suitably regular) solutions to the field equations on $\mathcal{M}$. To define the symplectic structure, one follows the following general procedure. Denote by $\bar{\delta} \equiv (\bar{\delta}e, \bar{\delta}\omega)$ tangent vectors in the space of histories. Since field equations are satisfied on $\mathbf{\Gamma}_{\rm cov}$, the change in the Lagrangian 4-form L_4 under a variation along $\bar{\delta}$ is for the form

$$(\bar{\delta}L_4)|_{\mathbf{\Gamma}_{\rm cov}} = \mathrm{d}L_3(\bar{\delta})$$

for some 3-form L_3 on $\mathcal{M}$ which depends linearly on $\bar{\delta}$. One can now define a 1-form $\mathbf{\Theta}$ on the space of histories via $\mathbf{\Theta}(\bar{\delta}) := \int_M L_3(\bar{\delta})$. The symplectic structure $\mathbf{\Omega}$ is simply the pullback to $\mathbf{\Gamma}_{\rm cov}$ of the curl of $\mathbf{\Theta}$ on the space of histories. In our case, $\bar{\delta}L_{(H)}$ is given by

$$\bar{\delta}L_{(H)} = -\frac{1}{2\gamma k}\,\mathrm{d}\left[e^I \wedge e^J \wedge \bar{\delta}\left(\omega_{IJ} - \frac{\gamma}{2}\epsilon_{IJKL}\omega^{KL}\right)\right],$$

whence the symplectic structure is given by

$$\mathbf{\Omega}(\delta_1,\delta_2) = -\frac{1}{k\gamma}\int_M [\delta_{[1}(e^I \wedge e^J)] \wedge \left[\delta_{2]}\left(\omega_{IJ} - \frac{\gamma}{2}\epsilon_{IJKL}\omega^{KL}\right)\right] \tag{2.7}$$

for all tangent vectors δ_1 and δ_2 to $\mathbf{\Gamma}_{\rm cov}$. From general considerations, it follows that the value of the integral is independent of the specific choice of a Cauchy surface M made in its evaluation.

[8] This modification was strongly motivated by the very considerable work on the framework based on (anti-)self-dual connections in the preceding decade (see, e.g., [2]), particularly by the discovery of an action for general relativity using these variables [14]. Also, our presentation contains several new elements which, to our knowledge, have not appeared in the literature before.

Using the fact that the tangent vectors to $\mathbf{\Gamma}_{\rm cov}$ must, in particular, satisfy the linearized version of (2.3) it is easy to verify that the γ-dependent term in (2.7) vanishes identically: not only is $\mathbf{\Gamma}_{\rm cov}$ independent of the value[9] of γ but so is the symplectic structure $\mathbf{\Omega}$ on it. As is clear from (2.7), the momentum conjugate to $e^I \wedge e^J$, on the other hand, does depend on the choice of γ. Thus, as noted above, like the θ term in the Yang–Mills theory, the γ term in (2.6) only induces a canonical transformation on the phase space.

In effect the canonical transformation is induced by the map

$$X_{IJ} \mapsto \frac{1}{2}\left(X_{IJ} - \frac{\gamma}{2}\epsilon_{IJKL}X^{KL}\right)$$

on $so(\bar{\eta})$. It is easy to verify that this map is a vector space isomorphism on $so(\bar{\eta})$ except when

$$\gamma^2 = \sigma := \mathrm{sgn}(\det \bar{\eta}),$$

the sign of the determinant of the metric tensor $\bar{\eta}_{IJ}$ (see footnote 9). At these exceptional values of γ, the map is a projection onto the subspace of $so(\eta)$ corresponding to the eigenvalue $-\gamma\sigma$ of the Hodge-dual operator $\star : X_{IJ} \mapsto \frac{1}{2}\epsilon_{IJ}{}^{KL}X_{KL}$. Furthermore, in this case the map is a *Lie algebra homomorphism*. In the Riemannian case (when $\bar{\eta}$ has signature $+, +, +, +$) this occurs for $\gamma = \pm 1$ while in the Lorentzian case (when $\bar{\eta}$ has signature $-, +, +, +$) it occurs for $\gamma = \pm \mathrm{i}$. In all these exceptional cases the theory has a richer geometrical structure. In particular, the combination $\frac{1}{2}(\omega_{IJ} - \gamma^\star\omega_{IJ})$ that occurs in the symplectic structure is again a (half-flat) connection.

Chronologically, the background-independent approach to quantum gravity summarized in this review originated from a reformulation of general relativity in terms of these half-flat connections [12, 13]. It turns out that all equations in the classical theory simplify considerably and underlying structures become more transparent in these variables. They are also closely related to Penrose's nonlinear gravitons [31] and Newman's H-space constructions [32]. In the Riemannian signature, one can continue to use these variables also in the quantum theory. In the Lorentzian case, on the other hand, the half-flat connections take values in the Lie algebra of non-compact groups and functional analysis on spaces of such connections is still not sufficiently well developed to carry out constructions required in the quantum theory. Therefore, in the Lorentzian case, most progress has occurred by working in sectors with real values of γ where, as we will see, one can work connections with compact structure groups.

In section 2.2 we will summarize the situation with half-flat connections in the Riemannian case and in section 2.3 we will discuss the Lorentzian theory using real-valued γ sectors.

2.2. Riemannian signature and half-flat connections

2.2.1. Preliminaries. Let us then assume that $\bar{\eta}_{IJ}$ is positive definite. Since $\sigma = 1$, the half-flat case corresponds to setting $\gamma = \pm 1$. Let us set

$$\omega^{(+)}{}_{IJ} = \frac{1}{2}\left(\omega_{IJ} - \frac{\gamma}{2}\epsilon_{IJ}{}^{KL}\omega_{KL}\right) \tag{2.8}$$

so that $\omega^{(+)}$ is the anti-self-dual part of ω if $\gamma = 1$ and self dual, if $\gamma = -1$. In these cases, the Holst action simplifies to

$$\mathbf{S}_{(H)}(e, \omega^{(+)}) = -\frac{1}{k\gamma}\int_{\mathcal{M}} \Sigma^{IJ}_{(+)} \wedge \Omega^{(+)}{}_{IJ} \tag{2.9}$$

[9] As a result, for generic values of γ, the equation of motion for the connection resulting from variation of $\mathbf{S}_{(H)}$ with respect to ω_μ^{IJ} is again (2.3). Hence the space of solutions obtained by varying $\mathbf{S}_{(H)}$ is the same as that obtained by varying $\mathbf{S}_{(P)}$. For the exceptional values, this equation states that the (anti-)self-dual part of ω_μ^{IJ} equals the (anti-)self-dual part of the connection compatible with the co-frame e_μ^I. However, it is again true that the spaces of solutions obtained by extremizing $\mathbf{S}_{(H)}$ and $\mathbf{S}_{(P)}$ are the same [2].

where $\Sigma^{IJ}_{(+)}$ is the (anti-)self-dual part of $e^I \wedge e^J$,

$$\Sigma^{IJ}_{(+)} = \frac{1}{2}\left(e^I \wedge e^J - \frac{\gamma}{2}\epsilon^{IJ}{}_{KL} e^K \wedge e^L\right),$$

and $\Omega^{(+)}_{IJ}$ is both the (anti-)self-dual part of Ω_{IJ} *and* the curvature of $\omega^{(+)}_{IJ}$:

$$\Omega^{(+)} = \mathrm{d}\omega^{(+)} + \omega^{(+)} \wedge \omega^{(+)}.$$

Note that the theory under consideration is full (Riemannian) general relativity; we are just describing it in terms of the fields $\left(e^I, \omega^{(+)}_{IJ}\right)$ where $\omega^{(+)}_{IJ}$ is a half-flat (i.e., self-dual or anti-self-dual) connection.

The symplectic form (2.7) now simplifies to

$$\begin{aligned}\boldsymbol{\Omega}(\delta_1, \delta_2) &= -\frac{2}{k\gamma}\int_M \left[\delta_{[1}\Sigma^{IJ}_{(+)}\right] \wedge \left[\delta_{2]}\omega^{(+)}{}_{IJ}\right] \\ &= \int_M \mathrm{d}^3x \left[\delta_1 P^a_{IJ}\delta_2 A^{IJ}_a - \delta_2 P^a_{IJ}\delta_1 A^{IJ}_a\right], \end{aligned} \tag{2.10}$$

where A_{IJ}, the pullback to M of $\omega^{(+)}_{IJ}$, represents the configuration variable and

$$P^a_{IJ} := -\frac{1}{2k\gamma}\eta^{abc}\Sigma^{(+)}_{bcIJ},$$

its canonically conjugate momentum. Here and in what follows η^{abc} will denote the metric-independent Levi-Civita density on M whose orientation is the same as that of the fixed orientation on M. Hence P^a_{IJ} is a pseudo-vector density of the weight 1 on M.[10]

2.2.2. The Legendre transform. Let us introduce on $\mathcal{M}$ a smooth ('time') function t such that $\mathrm{d}t$ is everywhere non-zero and each $t = \text{const}$ slice is diffeomorphic with M. Introduce a vector field t^α such that $t^\alpha \nabla_a t = 1$. Thus, t^α is to be thought of as the 'time-evolution vector field'. Denote by n^α the unit normal to the $t = \text{const}$ slices M and decompose t^α as $t^\alpha = Nn^\alpha + N^\alpha$ with $N^\alpha n_\alpha = 0$. The function N is called the *lapse* and the vector field N^α the *shift*. We will denote by q^a_α and q^α_a the projection operator on to vector and co-vector fields on M. Finally, a tensor field $T^{\alpha\ldots\beta}{}_{\gamma\ldots\delta}$ which is orthogonal in each of its indices to n^μ will be identified with its projection $T^{a\ldots b}{}_{c\ldots d} := q^a_\alpha \ldots q^b_\beta q^\gamma_c \ldots q^\delta_d T^{\alpha\ldots\beta}{}_{\gamma\ldots\delta}$.

With these preliminaries out of the way, it is now straightforward to perform the Legendre transform. The calculation is remarkably short (especially when compared to the Legendre transform in the metric variables; see, e.g., page 47 of [2]). In terms of fields A^{IJ}_a and P^a_{IJ} introduced above, one obtains[11],

$$\mathbf{S}_{(H)} = \int \mathrm{d}t \int_M \mathrm{d}^3x \left(P^a_{IJ}\mathcal{L}_t A^{IJ}_a - h_{(+)}(A, P, N, N^a, \omega^{(+)}\cdot t)\right) \tag{2.11}$$

where the Hamiltonian density $h_{(+)}$ is given by

$$h_{(+)} = -\left(\omega^{(+)}_{IJ}\cdot t\right)G^{IJ} + N^a C^{(+)}_a + NC_{(+)}, \tag{2.12}$$

with

$$\begin{aligned} G_{IJ} &:= \mathcal{D}^{(+)}_a P^a_{IJ} := \partial_a P^a_{IJ} + A_{aI}{}^K P^a_{KJ} + A_{aJ}{}^K P^a_{IK} \\ C^{(+)}_a &:= P^b_{IJ} F^{IJ}_{ab} \\ C_{(+)} &:= -\frac{k}{\sqrt{|\det q|}} P^{aJ}_I P^{bK}_J F_{abK}{}^I. \end{aligned} \tag{2.13}$$

[10] In terms of coordinates, for any smooth field V^{IJ} and 1-form f_a on M, the 3-form $V^{IJ} f_a P^a_{IJ}\, \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3$ is a volume element on M which is independent of the choice of coordinates (x^1, x^2, x^3).

[11] Here, and in the remainder of this paper, in the Lie derivative of a field with internal indices will be treated simply as scalars (i.e., ignored). Thus, $\mathcal{L}_t A^{IJ}_a = t^b \partial_b A^{IJ}_a + A^{IJ}_b \partial_a t^b$.

Here F_{ab}^{IJ} is the curvature of A_a^{IJ}, $F = \mathrm{d}A + A \wedge A$ and q is the determinant of the 3-metric

$$q_{ab} := q_a^\alpha q_b^\beta g_{\alpha\beta}$$

on M. The form of (2.12) confirms that, as suggested by (2.10), we should regard A_a^{IJ} as the configuration variable and P_{IJ}^a as its momentum. The momentum is related in a simple way to the 3-metric:

$$-\mathrm{Tr}\, P^a P^b = P_{IJ}^a P^{bIJ} = \frac{1}{k^2}(\det q) q^{ab}.$$

Note that $\omega \cdot t$, N and N^a are Lagrange multipliers; there are no equations governing them. The basic dynamical variables are only A_a^{IJ} and P_{IJ}^a; all other dynamical fields are determined by them. Variation of $\mathbf{S}_{(H)}$ with respect to these multipliers yields constraints

$$G_{IJ} = 0 \qquad C_a^{(+)} = 0 \qquad \text{and} \qquad C^{(+)} = 0. \tag{2.14}$$

As is always the case (in the spatially compact context) for theories without background fields the Hamiltonian is a sum of constraints. Variations of the action with respect to A_a^{IJ} and P_{IJ}^a yield the equations of motion for these basic dynamical fields. The three constraints (2.13) and these two evolution equations are equivalent to the full set of Einstein's equations.

2.2.3. The Hamiltonian framework. It follows from the Legendre transform (2.11) that the *canonical* phase space $\mathbf{\Gamma}_{\mathrm{can}}$ consists of canonically conjugate pairs of fields $\left(A_a^{IJ}, P_{IJ}^a\right)$ of M. The only non-trivial Poisson bracket is

$$\left\{A_a^{IJ}(x), P_{KL}^b(y)\right\} := \frac{1}{2}\left(\delta_{[K}^I \delta_{L]}^J - \frac{\gamma}{2}\delta_{[M}^I \delta_{N]}^J \epsilon^{MN}{}_{KL}\right)\delta_a^b \delta(x, y). \tag{2.15}$$

A key point is that the configuration variable A_a^{IJ} is again a connection on the 3-manifold M but the structure group is now the *spin group* $SO^{(+)}(\bar{\eta})$ (which, in the Riemannian case now under consideration, is isomorphic to $SU(2)$)[12]. Thus, in the Hamiltonian framework, general relativity has been cast as a dynamical theory of a spin connection.

The basic canonically conjugate variables are subject to three sets of constraints, spelled out in (2.13). It is easy to verify that the Poisson bracket between any two constraints vanishes on the constraint surface; in Dirac's terminology, they are of *first class*. The first constraint, $G_{IJ}^{(+)}$, generates internal gauge transformations in $SO^{(+)}(\bar{\eta})$. Modulo these gauge rotations, the second, $C_a^{(+)}$, generates diffeomorphisms on M, and the third, $C^{(+)}$, generates 'evolutions' along Nn^α. Using the relation between P_{IJ}^a and the 3-metrics q_{ab} on M, one can show that these equations are equivalent to the full set of Einstein's equations. However, one can work just with the connections A_a^{IJ} and their conjugate momenta, without any direct reference to metrics, even when gravity is coupled to matter [16]. In this sense, gravity can be regarded as a 'gauge theory' which has the same phase space $\mathbf{\Gamma}_{\mathrm{can}}$ as that of an $SO^{(+)}(\bar{\eta})$ Yang–Mills theory but is a fully constrained dynamics which does not refer to a background spacetime metric.

2.3. Generic real value of γ

The formulation of general relativity as a dynamical theory of half-flat connections, presented in section 2.2, has been studied in detail also for Lorentzian signature [2, 12–23]. However, in that case, certain subtleties arise because the connection is complex valued and the structure

[12] The full group $SO(\bar{\eta})$ does admit an action on the phase space, given by $(P, A) \mapsto (b^{-1}Pb, b^{-1}Ab + (b^{-1}\,\mathrm{d}b)^{(+)})$, where $^{(+)}$ stands for the projection onto $so^{(+)}(\bar{\eta})$ in $so(\bar{\eta})$. However, because of the projection, A does not transform as an $SO(\bar{\eta})$ connection.

group is non-compact. We have chosen to bypass these issues because, as explained in section 2.1, for passage to quantum theory we have in any case to use compact structure groups, i.e., real values of γ. Therefore, in this subsection we will let γ take any non-zero real value. Although we are now primarily interested in the $-, +, +, +$ signature, our analysis will apply also to the $+, +, +, +$ case.

2.3.1. Preliminaries. It is convenient to first carry out a partial gauge fixing. Let us fix an internal vector field n^I with $n^I n_I = \sigma$ (the signature of $\bar{\eta}_{IJ}$). We will require it to be constant (in the sense that from now on, we will restrict ourselves to flat derivative operators ∂ which annihilate n^I, in addition to $\bar{\eta}_{IJ}$). Let $V_\perp$ be the three-dimensional subspace of V orthogonal to n^I. Elements of $V_\perp$ will carry lower case superscripts, $i, j, \ldots, k$ and the projection operator on to $V_\perp$ will be denoted by q^i_I. In particular, then,

$$\eta_{ij} = q^I_i q^J_j \bar{\eta}_{IJ}$$

is the induced metric on $V_\perp$. Because we have fixed n^I, the group $SO(\bar{\eta})$ is now reduced to its subgroup $SO(\eta)$ which leaves n^I invariant. Finally, the alternating tensor ϵ_{IJKL} on V naturally induces an alternating tensor ϵ_{ijk} on $V_\perp$ via

$$\epsilon_{ijk} = q^I_i q^J_j q^K_k n^L \epsilon_{LIJK}.$$

Next, let us introduce a 'time function' t and the associated structure as in the beginning of section 2.2.2, with the following additional provisos if the signature is Lorentzian: the vector field t^α is future directed and n^α is the future-directed unit timelike normal to M. We will now allow only those co-frame fields e^I_α which are 'compatible' with the fixed n^I in the sense that $n^\alpha := n^I e^\alpha_I$ is the unit normal to the given foliation. (Note that every co-frame is gauge related to that satisfying this condition; see (2.5).) Each of these co-frames e^I_α naturally defines an orthonormal co-triad $e^i_a := e^I_\alpha q^i_I q^\alpha_a$: on each leaf M of the foliation, the induced metric q_{ab} is given by $q_{ab} = e^i_a e^j_b \eta_{ij}$. Similarly, the connection 1-form ω^{IJ}_α naturally defines two $so(3)$-valued 1-forms on M:

$$\Gamma^i_a := \tfrac{1}{2} q^\alpha_a q^i_I \epsilon^{IJ}{}_{KL} n_J \omega^{KL}_\alpha \qquad \text{and} \qquad K^i_a := q^i_I q^\alpha_a \omega^{IJ}_\alpha n_J. \tag{2.16}$$

These 1-forms have natural geometric interpretations. Γ^i_a is an $so(\eta)$ connection on M and it is compatible with e^i_a if ω^{IJ}_α is compatible with e^I_α. Thus, if (2.3) holds, we have

$$\mathrm{d}e^i + \epsilon^i{}_{jk} \Gamma^j \wedge e^k = 0. \tag{2.17}$$

K^i_a is the extrinsic curvature of M if (2.3) holds:

$$K^i_a = \left(q^\alpha_a q^b_\beta \nabla_\alpha n^\beta\right) e^i_b. \tag{2.18}$$

In terms of these fields, the symplectic structure (2.7) can be re-expressed as

$$\mathbf{\Omega}(\delta_1, \delta_2) = \int_M \mathrm{d}^3 x \left(\delta_1 P^a_i \delta_2 A^i_a - \delta_2 P^a_i \delta_1 A^i_a\right) \tag{2.19}$$

where

$$P^a_i := \frac{1}{2k\gamma} e^j_b e^k_c \eta^{abc} \epsilon_{ijk} \qquad \text{and} \qquad A^i_a := \Gamma^i_a - \sigma\gamma K^i_a. \tag{2.20}$$

Note that A^i_a is a connection 1-form on M which takes values in $so(\eta)$. P^a_i is again a vector density of weight 1 on M which now takes values in (the dual of) $so(\eta)$. Geometrically, it represents an orthonormal triad $\tilde{E}^a$ of density weight 1 on M,

$$k\gamma P^a_i = \sqrt{|\det q|} e^a_i \equiv \tilde{E}^a_i \qquad \text{whence} \quad |\det q| q^{ab} = k^2\gamma^2 P^a_i P^b_j \eta^{ij} \tag{2.21}$$

where $\det q$ is the determinant of the 3-metric q_{ab} on M.

Let us summarize. Through gauge fixing, we first reduced the internal gauge group from $SO(\bar{\eta})$ to $SO(\eta)$. The new configuration variable A_a^i is an $so(\eta)$-valued connection on M, constructed from the spin connection Γ_a^i compatible with the co-triad e_a^i and the extrinsic curvature K_a^i. Apart from a multiplicative factor γ, the conjugate momentum P_i^a has the interpretation of a triad with density weight 1. Note that relation (2.20) between the canonical variables A_a^i, P_i^a and the geometrical variables e_a^i and K_a^i holds also in the half-flat case; it is just that there is also an additional restriction, $\sigma^2\gamma^2 = \pm 1$.

2.3.2. The Legendre transform. Let us return to the Holst action (2.6) and perform the Legendre transform as in section 2.2.2. Again, the calculations are simple but the full expression of the resulting Hamiltonian density h is now more complicated. As before one obtains

$$\mathbf{S}_{(H)} = \int \mathrm{d}t \int_M \mathrm{d}^3x \big(P_i^a \mathcal{L}_t A_a^i - h\big(A_a^i, P_i^a, N, N^a, \Gamma\cdot t\big)\big) \tag{2.22}$$

with h given by

$$h = (\omega^i \cdot t)G_i + N^a C_a + NC. \tag{2.23}$$

Again $\omega^i \cdot t := -\frac{1}{2}\epsilon^{ijk}\omega_{jk} \cdot t$, N^a and N are Lagrange multipliers. However, now the accompanying constraints acquire additional terms:

$$\begin{aligned} &G_i = \mathcal{D}_a P_i^a := \partial_a P_i^a + \epsilon_{ij}{}^k A_a^j P_k^a \qquad C_a = P_i^b F_{ab}^i - \frac{\sigma-\gamma^2}{\sigma\gamma} K_a^i G_i \\ &C = \frac{k\gamma^2}{2\sqrt{|\det q|}} P_i^a P_j^b \big[\epsilon^{ij}{}_k F_{ab}^k + (\sigma-\gamma^2) 2K_{[a}^i K_{b]}^j\big] + (\gamma^2-\sigma)k\partial_a\left(\frac{P_i^a}{\sqrt{|\det q|}}\right) G^i. \end{aligned} \tag{2.24}$$

Here, F_{ab}^k is the curvature of the connection A_a^i and $|\det q|$ can be expressed directly in terms of P_i^a:

$$|\det q| = \frac{(k\gamma)^3}{\sqrt{|\det \eta|}} \det P. \tag{2.25}$$

Thus, the overall structure of the constraints is very similar to that in the half-flat case. However, there is a major new complication in the detailed expressions of constraints: now they involve also $K_a^i = (1/\sigma\gamma)\big(\Gamma_a^i - A_a^i\big)$ and Γ_a^i is a non-polynomial function of P_i^a.[13] (Since these terms are multiplied by $(\sigma-\gamma^2)$, they disappear in the half-flat case.)

2.3.3. Hamiltonian theory. Now the canonical phase space $\mathbf{\Gamma}_{\mathrm{can}}$ consists of pairs $\big(A_a^i, P_i^a\big)$ of fields on the 3-manifold M, where A_a^i is a connection 1-form which takes values in $so(\eta)$ and P_i^a is a vector density of weight 1 which takes values in the dual of $so(\eta)$. The only non-vanishing Poisson bracket is

$$\big\{A_a^i(x), P_j^b(y)\big\} := \delta_j^i \delta_a^b \delta(x,y). \tag{2.26}$$

Thus, the phase space is the same as that of a Yang–Mills theory with $SO(\eta)$ as the structure group. There is again a set of three constraints (2.24) which are again of first class in Dirac's terminology. The basic canonical pair evolves via Hamilton's equations,

$$\dot{A}_a^i = \big\{A_a^i, H\big\}, \qquad \dot{P}_i^a = \big\{P_i^a, H\big\}$$

[13] Although the possibility of using real γ was noted already in the mid-1980s, this choice was ignored in the Lorentzian case because the term K_a^i seemed unmanageable in quantum theory. The viewpoint changed with Thiemann's discovery that this difficulty can be overcome; see section 6.

where the Hamiltonian is simply $H = \int_M \mathrm{d}^3x\, h$. The set of three constraints and these two evolution equations are completely equivalent to Einstein's equations. Thus, general relativity is again recast as a dynamical theory of connections.

Before analysing the phase space structure in greater detail, we wish to emphasize two important points. First, note that in the Hamiltonian theory we simply begin with the fields $\big(A_a^i, P_i^a\big)$; neither they nor their Poisson brackets depend on the Barbero–Immirzi parameter γ. Thus, the canonical phase space is manifestly γ independent. γ appears only when we express geometrical fields—the spatial triad e_i^a and the extrinsic curvature K_a^i—in terms of the basic canonical variables (see (2.21) and (2.18)). The second point concerns a conceptual difference between the use of half-flat and general connections. The configuration variables in both cases are connections on M. Furthermore, as noted in section 2.3.1, relation (2.20) between these connections and the fields e_a^i, K_a^i is identical in form. However, while the variable A_a^{IJ} of section 2.2 is the pullback to M of a *spacetime* connection A_α^{IJ}, the variable A_a^i now under consideration is not so obtained [22]. From the spacetime geometry perspective, therefore, A_a^i is less natural. While this is a definite drawback from the perspective of the classical theory, it is not a handicap for canonical quantization. Indeed a spacetime geometry is analogous to a trajectory in particle mechanics and particle trajectories play no essential role in quantum mechanics.

Finally, let us analyse the structure of constraints. As one would expect, the first constraint, $G_i = 0$, is simply the 'Gauss law' which ensures invariance under internal $SO(\eta)$ rotations. Indeed, for any smooth field Λ^i on M which takes values in $so(\eta)$, the function

$$\mathcal{C}_{\mathrm{G}}(\Lambda) := \int_M \mathrm{d}^3x\, \Lambda^i G_i \tag{2.27}$$

on the phase space generates precisely the internal rotations along Λ^i:

$$\big\{A_a^i, \mathcal{C}_{\mathrm{G}}(\Lambda)\big\} = -\mathcal{D}_a\Lambda^i \qquad \text{and} \qquad \big\{P_i^a, \mathcal{C}_{\mathrm{G}}(\Lambda)\big\} = \epsilon_{ij}{}^k\Lambda^j P_k^a. \tag{2.28}$$

To display the meaning of the second constraint C_a of (2.24), it is convenient to remove from it the part which generates internal rotations which we have already analysed. Therefore, for each smooth vector field $\vec{N}$ on M let us define

$$C_{\mathrm{Diff}}(\vec{N}) := \int_M \mathrm{d}^3x \big(N^a P_i^b F_{ab}^i - \big(N^a A_a^i\big)G_i\big). \tag{2.29}$$

This constraint function generates diffeomorphisms along $\vec{N}$:

$$\big\{A_a^i, \mathcal{C}_{\mathrm{Diff}}(\vec{N})\big\} = \mathcal{L}_{\vec{N}} A_a^i \qquad \text{and} \qquad \big\{P_i^a, \mathcal{C}_{\mathrm{Diff}}(\vec{N})\big\} = \mathcal{L}_{\vec{N}} P_i^a. \tag{2.30}$$

Finally, let us consider the third constraint in (2.24). For quantization purposes, it is again convenient to remove a suitable multiple of the Gauss constraint from it. Following Barbero and Thiemann, we will set

$$\mathcal{C}(N) = \frac{k\gamma^2}{2}\int_M \mathrm{d}^3x\, N \frac{P_i^a P_j^b}{\sqrt{|\det q|}}\big[\epsilon^{ij}{}_k F_{ab}^k + 2(\sigma-\gamma^2)K_{[a}^i K_{b]}^j\big]. \tag{2.31}$$

As one might expect, this constraint generates time evolution, 'off' M. The Poisson brackets between these specific constraints are

$$\{\mathcal{C}_{\mathrm{G}}(\Lambda), \mathcal{C}_{\mathrm{G}}(\Lambda')\} = \{\mathcal{C}_{\mathrm{G}}([\Lambda, \Lambda'])\} \qquad \{\mathcal{C}_{\mathrm{G}}(\Lambda), \mathcal{C}_{\mathrm{Diff}}(\vec{N})\} = -\mathcal{C}_{\mathrm{G}}(\mathcal{L}_N\Lambda) \tag{2.32}$$

$$\{\mathcal{C}_{\mathrm{Diff}}(\vec{N}), \mathcal{C}_{\mathrm{Diff}}(\vec{N}')\} = \mathcal{C}_{\mathrm{Diff}}([\vec{N}, \vec{N}']) \tag{2.33}$$

$$\{\mathcal{C}_{\mathrm{G}}(\Lambda), \mathcal{C}(N)\} = 0 \qquad \{\mathcal{C}_{\mathrm{Diff}}(\vec{N}), \mathcal{C}(M)\} = -\mathcal{C}(\mathcal{L}_N M) \tag{2.34}$$

and

$$\{\mathcal{C}(N), \mathcal{C}(M)\} = k^2\gamma^2\sigma(\mathcal{C}_{\rm Diff}(\vec{S}) + \mathcal{C}_{\rm G}(S^a A_a)) + (\sigma - \gamma^2)\mathcal{C}_{\rm G}\left(\frac{[P^a\partial_a N,\, P^b\partial_b M]}{|\det q|}\right). \quad (2.35)$$

In the last equation, the vector field S^a is given by

$$S^a = (N\partial_b M - M\partial_b N)\frac{P^b_i P^{ai}}{|\det q|}. \quad (2.36)$$

As in geometrodynamics, the smearing fields in the last Poisson bracket depend on dynamical fields themselves. Therefore, the constraint algebra is open in the BRST sense; we have structure functions rather than structure constants. We will return to this point in section 6.

To summarize, both the Euclidean and Lorentzian general relativity can be cast as a dynamical theory of (real-valued) connections with compact structure groups. The price in the Lorentzian sector is that we have to work with a real value of the Barbero–Immirzi parameter, for which the expressions of constraints and their Poisson algebra are more complicated.

Remarks

(1) For simplicity, in this section we focused just on the gravitational field. Matter couplings have been discussed in detail in the literature using half-flat gravitational connections in the framework of general relativity as well as supergravity (see, e.g., [16, 17]). In the matter sector, modifications required to deal with generic γ values of the Barbero–Immirzi parameter are minimal.

(2) In the purely gravitational sector considered here, the internal group for general real values of γ is $SO(\eta)$. For cases we focused on, η_{ij} is positive definite whence $SO(\eta) = SO(3)$. However, since we also wish to incorporate spinors, in the remainder of the paper we will take the internal group to be $SU(2)$. This will also make the structure group the same in the generic and half-flat cases.

(3) Throughout this section we have assumed that the frames, co-frames and metrics under consideration are non-degenerate. However, the final Hamiltonian framework can be naturally extended to allow degenerate situations. Specifically, by replacing the scalar lapse function N with one of density weight -1, one can allow for the possibility that the fields P^a_i become degenerate, i.e., have $\det P = 0$. Somewhat surprisingly, the dynamics continues to be well defined and one obtains an extension of general relativity with degenerate metrics. For details, see, e.g., [26–30].

3. Quantization strategy

In sections 4 and 5 we will provide a systematic, step-by-step construction of background-independent quantum theories of connections (including general relativity) and a quantum theory of geometry. Since that treatment is mathematically self-contained, the procedure involved is rather long. Although individual steps in the construction are straightforward, the motivation, the goals, and the relation to procedures used in standard quantum field theories may not always be transparent to an uninitiated reader. Therefore, in this section, we will provide the motivation behind our constructions, a summary of the underlying ideas and a global picture that will aid the reader to see where one is headed.

3.1. Scalar field theories

To anchor the discussion in well-established physics, we will begin by briefly recalling the construction of the Hilbert space of states and basic operators for a free massive scalar field in

Minkowski spacetime, within the canonical approach. (For further details, see, e.g., [43].) The classical configuration space $\mathcal{C}$ is generally taken to be the space of smooth functions ϕ which decay rapidly at infinity on a $t = \text{const}$ slice, M. From one's experience in non-relativistic quantum mechanics, one would expect quantum states to be 'square integrable functions' Ψ on $\mathcal{C}$. However, since the system now has an infinite number of degrees of freedom, the integration theory is now more involved and the intuitive expectation has to be suitably modified.

The key idea, which goes back to Kolmogorov, is to build the infinite-dimensional integration theory from the finite-dimensional one. One begins by introducing a space $\mathcal{S}$ of 'probes', typically taken to be real test functions e on the spatial slice M. Elements of $\mathcal{S}$ probe the structure of the scalar field $\phi \in \mathcal{C}$ through linear functions h_e on $\mathcal{C}$,

$$h_e(\phi) = \int_M \mathrm{d}^3x\, e(x)\phi(x) \tag{3.1}$$

which capture a small part of the information in the field ϕ, namely 'its component along e'. Given a set α of probes, $h_{e_1}, \ldots, h_{e_n}$ and a (suitably regular) complex-valued function ψ of n real variables, we can now define a more general function Ψ on $\mathcal{C}$,

$$\Psi(\phi) := \psi\big(h_{e_1}(\phi), \ldots, h_{e_n}(\phi)\big), \tag{3.2}$$

which depends only on the n 'components' of ϕ singled out by the chosen probes. (Strictly, Ψ should be written as Ψ_α but we will omit the suffix for notational simplicity.) Such functions are said to be *cylindrical*. We will denote by Cyl_α the linear space they span. Given a measure $\mu_{(n)}$ on $\mathbb{R}^n$, we define an Hermitian inner product on Cyl_α in an obvious fashion:

$$\langle \Psi_1, \Psi_2 \rangle := \int_{\mathbb{R}^n} \mathrm{d}\mu_{(n)}(\bar{\psi}_1\psi_2)\big(h_{e_1}(\phi), \ldots, h_{e_n}\big). \tag{3.3}$$

The idea is to extend this inner product to the space Cyl of *all* cylindrical functions, i.e., the space of all functions on $\mathcal{C}$ which are cylindrical with respect to *some* set of probes. However, there is an important caveat which arises because a given function Ψ on $\mathcal{C}$ may be cylindrical with respect to *two different* sets of probes. (For example, every $\Psi \in \mathrm{Cyl}_\alpha$ is also in Cyl_β where β is obtained simply by enlarging α by adding new probes.) The inner product will be well defined only if the value of the integral does not depend on the specific set α of probes we use to represent the function. This requirement imposes consistency conditions on the family of measures $\mu_{(n)}$. These conditions are non-trivial. But they *can* be met. The simplest example is provided by setting $\mu_{(n)}$ to be normalized Gaussian measures on $\mathbb{R}^n$. Every family $\{\mu_{(n)}\}$ of measures satisfying these consistency conditions enables us to integrate general cylindrical functions and is therefore said to define a *cylindrical measure* μ on $\mathcal{C}$. The Cauchy completion $\mathcal{H}$ of $(\mathrm{Cyl}, \langle\,,\rangle)$ is then taken to be the space of quantum states.

If this construction were restricted to any one set α of probes, the resulting Hilbert space would be (infinite dimensional but) rather small because it would correspond to the space of quantum states of a system with only a finite number of degrees of freedom. The huge enlargement, accommodating the infinite number of degrees of freedom, comes about because we allow *arbitrary* sets α of probes which provide a 'chart' on all of $\mathcal{C}$, enabling us to incorporate the infinite number of degrees of freedom in the field ϕ.

Let us examine this issue further. Any one cylindrical function is a 'fake' infinite-dimensional function in the sense that its 'true' dependence is only on a finite number of variables. However, in the Cauchy completion, we obtain states which 'genuinely' depend on an infinite number of degrees of freedom. However, in general, these states cannot be realized as functions on $\mathcal{C}$. In the case of free fields, the appropriate measures are Gaussians (with zero mean and variance determined by the operator $\Delta - \mu^2$) and all quantum states can be realized as functions on the space $\mathcal{S}'$ of tempered distributions, the topological dual of the space $\mathcal{S}$ of

probes. In fact, the cylindrical measure can be extended to a regular Borel measure μ on $\mathcal{S}'$ and the Hilbert space is given by $\mathcal{H} = L^2(\mathcal{S}', \mathrm{d}\mu)$. $\mathcal{S}'$ is referred to as the *quantum configuration space*. Finally, as in Schrödinger quantum mechanics, the configuration operators $\hat{\phi}(f)$ are represented by multiplication and momentum operators $\hat{\pi}(f)$ by derivation (plus a multiple of the 'divergence of the vector field $\int \mathrm{d}^3x\, \delta/\delta\phi(x)$ with respect to the Gaussian measure' [160])[14]. This 'Schrödinger representation' of the free field is entirely equivalent to the more familiar Fock representation.

Thus, the overall situation is rather similar to that in quantum mechanics. The presence of an infinite number of degrees of freedom causes only one major modification: the classical configuration space $\mathcal{C}$ of smooth fields is enlarged to the quantum configuration space $\mathcal{S}'$ of distributions. Quantum field theoretic difficulties associated with defining products of operators can be directly traced back to this enlargement.

3.2. Theories of connections

We saw in section 2 that general relativity can be recast in such a way that the configuration variables are $SU(2)$ connections on a 'spatial' manifold M. In this section, we will indicate how the quantization strategy of section 3.1 can be modified to incorporate such background-independent theories of connections. We will let the structure group be an arbitrary compact group G and denote by $\mathcal{A}$ the space of all suitably regular connections on M. $\mathcal{A}$ is the classical configuration space of the theory[15].

The idea again is to decompose the problem into a set of finite-dimensional ones. Hence, our first task is to introduce a set of probes to extract a finite number of degrees of freedom from the connection field. The new element is gauge invariance: now the probes have to be well adapted to extract gauge invariant information from connections. Therefore, it is natural to define cylindrical functions through holonomies h_e along edges e in M. This suggests that we use edges as our probes. Unlike in the case of a scalar field, holonomies are not linear functions of the classical field A; in gauge theories, the duality between the probes and classical fields becomes nonlinear.

Denote by α graphs on M with a finite number of edges e. Then, given a connection A on M, holonomies $h_e(A)$ along the edges e of α contain gauge invariant information in the restriction to the graph α of the connection A. While these capture only a finite number of degrees of freedom, the full gauge invariant information in A can be captured by considering *all possible* graphs α.

The strategy, as in section 3.1, is to first develop the integration theory using single graphs α. If the graph α has n edges, the holonomies $h_{e_1}, \ldots, h_{e_n}$ associate with every connection A an n-tuple $(g_1, \ldots, g_n)$ of elements of G. Therefore, given a (suitably regular) function ψ on G^n, we can define a function Ψ on the classical configuration space $\mathcal{A}$ as follows:

$$\Psi(A) := \psi\big(h_{e_1}(A), \ldots, h_{e_n}(A)\big). \tag{3.4}$$

These functions are said to be cylindrical with respect to the graph α and their space will be denoted by Cyl_α. To define a scalar product on Cyl_α, it is natural to choose a measure $\mu_{(n)}$ on G^n and set

$$\langle \Psi_1, \Psi_2 \rangle := \int_{G^n} \mathrm{d}\mu_{(n)} \bar{\psi}_1 \psi_2. \tag{3.5}$$

[14] For interacting field theories rigorous constructions are available only in low spacetime dimensions. The $\lambda\phi^4$ theory, for example, is known to exist in two spacetime dimensions but now the construction involves non-Gaussian measures. For a brief summary, see [43].

[15] Since the goal of this section is only to sketch the general strategy, for simplicity we will assume that the bundle is trivial and regard connections as globally defined 1-forms which take values in the Lie algebra of G.

This endows Cyl_α with a Hermitian inner product. This analysis is completely analogous to that used in lattice gauge theories, the role of the lattice being played by the graph α.

However, as in section 3.1, elements of Cyl_α are 'fake' infinite-dimensional functions because they depend only on a finite number of 'coordinates', $h_{e_1}, \ldots, h_{e_n}$, on the infinite-dimensional space $\mathcal{A}$. To capture the full information contained in $\mathcal{A}$, we have to allow *all possible* graphs in M.[16] Denote by Cyl functions on $\mathcal{A}$ which are cylindrical with respect to *some* graph α. The main challenge lies in extending the integration theory from Cyl_α to Cyl. Again the key subtlety arises because Ψ_1 and Ψ_2 in Cyl may be cylindrical with respect to many graphs and there is no *a priori* guarantee that the value of the inner product is independent of which of these graphs are used to perform the integral on the right-hand side of (3.5). The requirement that the inner product be well defined imposes severe restrictions on the choice of measures $\mu_{(n)}$ on G^n. However, as discussed in section 5, there is a natural choice compatible with the requirement that the theory be diffeomorphism covariant, imposed by our goal of constructing a background-independent quantum theory [40, 41, 43–45, 56–58].

As in section 3.1, a consistent set of measures $\mu_{(n)}$ on G^n provides a cylindrical measure on $\mathcal{A}$ and a general result ensures that such a measure can be naturally extended to a regular Borel measure on an extension $\bar{\mathcal{A}}$ of $\mathcal{A}$ [43]. The space $\bar{\mathcal{A}}$ is called the *quantum configuration space*. It contains 'generalized connections' which cannot be expressed as continuous fields on M but nonetheless assign well-defined holonomies to edges in M. These are referred to as *quantum connections*. Conceptually, the enlargement from $\mathcal{A}$ to $\bar{\mathcal{A}}$ which occurs in the passage to quantum theory is very similar to the enlargement from $\mathcal{C}$ to $\mathcal{S}'$ in the case of scalar fields. This enlargement plays a key role in quantum theory (especially in the discussion of surface states of a quantum horizon discussed in section 8). It is an imprint of the fact that, unlike in lattice theories, here we are dealing with a genuine field theory with an infinite number of degrees of freedom.

By now, the structure of the quantum configuration space $\bar{\mathcal{A}}$ is well understood [39, 40, 52, 53, 66]. In particular, using an algebraic approach (which has been used so successfully in non-commutative geometry), differential geometry has been developed on $\bar{\mathcal{A}}$ [66]. It enables the introduction of physically interesting operators discussed in sections 4.3, 5 and 6.

Ideas sketched in this section are developed systematically in the next two sections. We begin in section 4.1 by discussing quantum mechanics on a compact Lie group G and use it to introduce the quantum theory of connections on a graph in section 4.2. The quantum theory of connections in the continuum is discussed in section 4.3. This structure is then used in section 5 to introduce quantum geometry.

4. Quantum theories of connections: background-independent kinematics

In this section, we will construct a kinematical framework for background-independent, quantum theories of connections in the abstract, without direct reference to section 2. To bring out the generality of these constructions, we will work with gauge fields for which the structure group is any compact Lie group G. This discussion of theories of connections is divided into three parts. In the first, we provide a gentle introduction to the subject via quantum mechanics of a 'particle' on the group manifold of a compact Lie group G; in the second, we consider the quantum kinematics of a (background-independent) lattice gauge theory with structure group G on an arbitrary graph; and, in the third, we consider connections in the continuum with structure group G.

[16] Note that this strategy is quite different from the standard continuum limit used in lattice approaches to Minkowskian field theories. Our strategy is well suited to background-independent theories where there is no kinematic metric to provide scales. Technically, it involves a 'projective limit' [44, 66].

Constructions based on a general compact Lie group are important, e.g., in the discussion of the Einstein–Yang–Mills theory. However, for quantum geometry and for the formulation of quantum Einstein's equations, as we saw in section 2, the relevant group is $G = SU(2)$. Therefore, we will often spell out the situation for this case in greater detail. We will use the following conventions. The dimension of the G will be d and its Lie algebra will be denoted by $\mathfrak{g}$. Occasionally, we will use a basis τ^i in $\mathfrak{g}$. In the case $G = SU(2)$, the Lie algebra $\mathfrak{g} = su(2)$ will be identified with the Lie algebra of all the complex, traceless, anti-self-adjoint 2 by 2 matrices. Then the Cartan–Killing metric η_{ij} is given by

$$\eta(\xi, \zeta) = -2\,\mathrm{Tr}(\xi\zeta), \tag{4.1}$$

for all $\xi, \zeta \in su(2)$. In this case our τ_i will constitute an orthonormal basis satisfying

$$[\tau_i, \tau_j] = \epsilon^k{}_{ij}\tau_k. \tag{4.2}$$

4.1. Quantum mechanics on a compact Lie group G

Let us consider a 'free' particle on the group manifold of a compact Lie group G. In this subsection, we will discuss (classical and) quantum mechanics of this particle. The quantum Hilbert space and operators will be directly useful to quantum kinematics of theories of connections discussed in the next two subsections. The theory described in this section also has some direct physical applications. For example, in the case $G = SO(3)$, it describes 'a free spherical top' while if $G = SU(2)$, it plays an important role in the description of hadrons in the Skyrme model.

4.1.1. Phase space. The configuration space of the particle is the group manifold of G and the phase space is its cotangent bundle $T^\star(G)$. The natural Poisson bracket between functions on $T^\star(G)$ is given by

$$\{f_1, f_2\} = \frac{\partial f_1}{\partial q^i}\frac{\partial f_2}{\partial p_i} - \frac{\partial f_2}{\partial q^i}\frac{\partial f_1}{\partial p_i} \tag{4.3}$$

where q^i are coordinates on G and (q^i, p_i) are the corresponding coordinates on $T^\star(G)$.

Every smooth function f on G defines a configuration variable and every smooth vector field X^i, a momentum variable $P_X := X^i p_i$ on $T^\star(G)$. As on any cotangent bundle, (non-trivial) Poisson brackets between them mirror the action of vector fields on functions and the Lie bracket between vector fields:

$$\{P_x, f\} = -\mathcal{L}_X f \qquad \text{and} \qquad \{P_X, P_Y\} = -P_{[X,Y]}. \tag{4.4}$$

These configuration and momentum observables are said to be *elementary* in the sense that they admit unambiguous quantum analogues.

Being a Lie group, G admits two natural Lie algebras of vector fields, each of which is isomorphic with the Lie algebra $\mathfrak{g}$ of G. Given any $\xi \in \mathfrak{g}$, we can define a left (respectively, right) invariant vector field $L^{(\xi)}$ (respectively, $R^{(\xi)}$) on G such that

$$L^{(\xi)} f(g) = \frac{\mathrm{d}}{\mathrm{d}t} f(g\,\mathrm{e}^{t\xi}) \qquad \text{and} \qquad R^{(\xi)} f(g) = \frac{\mathrm{d}}{\mathrm{d}t} f(\mathrm{e}^{-t\xi} g). \tag{4.5}$$

(The sign convention is such that $L^{(\xi)} \mapsto R^{(\xi)}$ under $g \mapsto g^{-1}$.)

The corresponding momentum functions on $T^\star(G)$ will be denoted by $J^{(L,\xi)}$, $J^{(R,\xi)}$. These are generalizations of the familiar 'angular momentum functions' on $T^\star SO(3)$. Each set forms a d-dimensional vector space which is closed under the Poisson bracket. Since any vector field X on G can be expressed as a (functional) linear combination of $L^{(\xi)}$ ($R^{(\xi)}$), it suffices to restrict oneself only to this $2d$-dimensional space of momentum observables.

Since the particle is 'free', the Hamiltonian is given just by the kinetic term,

$$H(p,q) = \eta^{ij} p_i p_j, \tag{4.6}$$

where η_{ij} is a metric tensor defined on G and invariant with respect to the left and right action of G on itself. Given an orthonormal basis $\tau_i, i = 1, \ldots, d$, in $\mathfrak{g}$, and we denote $J^{(L,\tau_i)}$ by $J_i^{(L)}$ and $J^{(R,\tau_i)}$ by $J_i^{(R)}$, then the Hamiltonian can be rewritten as

$$H(p,q) = J_i^{(L)} J_j^{(L)} \eta^{ij} = J_i^{(R)} J_j^{(R)} \eta^{ij}. \tag{4.7}$$

We will see that all these basic observables naturally define operators in the quantum theory.

4.1.2. Quantization. Since G is equipped with the normalized Haar measure μ_H, the Hilbert space of quantum states can be taken to be the space $L^2(G, \mathrm{d}\mu_H)$ of square integrable functions on G with respect to the Haar measure. (For a detailed discussion, see [1, 3].) The configuration and momentum operators can be introduced as follows. With every smooth function f on G, we can associate a configuration operator $\hat{f}$ in the obvious fashion

$$(\hat{f}\psi)(g) = f(g)\psi(g), \tag{4.8}$$

and with every momentum function $X^i p_i$, a momentum operator $\hat{J}_{(X)}$ via

$$(\hat{J}^{(X)}\psi)(g) = \mathrm{i}\left[\mathcal{L}_X \psi + \tfrac{1}{2}(\mathrm{div}\, X)\psi\right](g), \tag{4.9}$$

where div X is the divergence of the vector field X with respect to the invariant volume form on G (and, for later convenience, we have left out the factor of $\hbar$.) It is straightforward to check that the commutators of these configuration and momentum operators mirror the Poisson brackets between their classical counterparts. Of particular interest are the operators associated with the left (and right) invariant vector fields associated with an orthonormal basis τ_i of $\mathfrak{g}$. We will set

$$\hat{L}_i = \hat{J}_i^{(L)} \qquad \text{and} \qquad \hat{R}_i = \hat{J}_i^{(R)}. \tag{4.10}$$

Since the divergence of right and left invariant vector fields vanishes, the action of operators is given just by the Lie-derivative term, i.e., formally, by the Poisson bracket between the momentum functions and ψ. In terms of these operators, the quantum Hamiltonian is given by

$$\hat{H} = \hat{L}_i \hat{L}_j \eta^{ij} = \hat{R}_i \hat{R}_j \eta^{ij} = -\Delta, \tag{4.11}$$

where Δ is the Laplace operator on G.

4.1.3. Spin states. In theories of connections developed in the next two subsections, a 'generalized spin network decomposition' of the Hilbert space of states will play an important role. As a prelude to that construction, we will now introduce an orthogonal decomposition of the Hilbert space $L^2(G, \mathrm{d}\mu_H)$ into *finite*-dimensional subspaces. Let j label inequivalent irreducible representations of G, let V_j denote the carrier space of the j-representation and let $V_j^\star$ be its dual. Then, the Weyl theorem provides the decomposition we are seeking:

$$L^2(G, \mathrm{d}\mu_H) = \oplus_j \mathcal{S}_j \qquad \text{with} \quad \mathcal{S}_j = V_j \otimes V_j^\star. \tag{4.12}$$

In the case $G = SU(2)$ we can make this decomposition more explicit. As is well known from quantum mechanics of angular momentum, in this case the eigenvalues of the operator $\hat{J}^2 = -\Delta$ are given by $j(j+1)$, where j runs through all the non-negative half-integers and labels the irreducible representations. Each carrier space V_j is now $2j+1$ dimensional. We can further decompose each $\mathcal{S}_j = V_j \otimes V_j^\star$ into orthogonal one-dimensional subspaces. Fix an element $\xi \in su(2)$ and consider the pair of commuting operators, $\hat{L}_{(\xi)}$ and $\hat{R}_{(\xi)}$. Given j,

every pair of eigenvalues, $j_{(L,\xi)}$, $j_{(R,\xi)}$, of these operators, each in $-j, -j+1, \ldots, j$, defines a one-dimensional eigensubspace $\mathcal{S}_{j,j_{(L,\xi)},j_{(R,\xi)}}$. Thus, we have

$$L^2(SU(2), \mathrm{d}\mu_H) = \oplus_j \mathcal{S}_j = \oplus_{j,j_{(L,\xi)},j_{(R,\xi)}} \mathcal{S}_{j,j_{(L,\xi)},j_{(R,\xi)}}. \tag{4.13}$$

This fact will lead us to spin network decomposition in the next two subsections.

4.2. Connections on a graph

Before considering (field) theories of connections, let us consider an intermediate quantum mechanical system, that of connections on a fixed graph α with a finite number of edges. This system is equivalent to lattice gauge theory on α [1]. In the next subsection, we will see that field theories of connections in the continuum can be obtained by appropriately 'gluing' theories associated with all possible graphs on the given manifold, in the manner sketched in section 3.2.

A graph may be thought of as a collection of edges and vertices and will serve as a 'floating' lattice[17]. ('Floating', because the edges need not be rectangular. Indeed since we do not have a background metric, terms like 'rectangular' have no invariant meaning.) A graph α' is said to be *larger* than another graph α (or *contain* α), $\alpha \geqslant \alpha'$, if every edge e of α can be written as $e = e'_1 \circ \cdots \circ e'_k$ for some edges $e'_1, \ldots, e'_k$ of α'.

4.2.1. Spaces of connections on a graph. A *G connection A_α on a graph α* is the set of $\mathfrak{g}$-valued 1-forms A_e defined on each edge e of α. For concreteness, we will suppose that each A_α is given by the pullback to α of a smooth $\mathfrak{g}$-valued 1-form on M.[18] Thus, one can think of a connection on α simply as an equivalence class of smooth connections on M where two are equivalent if their restrictions to each edge of α agree. (This concrete representation of A_α will make the passage to section 4.2 more transparent but is not essential in this section.)

Denote the space of G connections on α by $\mathcal{A}_\alpha$. This space is infinite dimensional because of the trivial redundancy of performing local gauge transformations along the edges of α. As in lattice gauge theories it is convenient to remove this redundancy to arrive at a finite-dimensional space $\bar{\mathcal{A}}_\alpha$, which can be taken to be the relevant configuration space for any (background-independent) theory of connections associated with the graph α.

A gauge transformation g_α in $\mathcal{G}_\alpha$ is a map $g_\alpha : x_\alpha \to G$ from all points x_α on α. Thus, g_α can be thought of as the restriction to α of a G-valued function defined on M. Under g_α, connections A_α transform as

$$A_\alpha \mapsto g_\alpha^{-1} A_\alpha g_\alpha + g_\alpha^{-1}\, \mathrm{d}_\alpha g_\alpha, \tag{4.14}$$

where d_α is the exterior derivative along the edges of α. Let us now consider the quotient spaces

$$\bar{\mathcal{A}}_\alpha := \mathcal{A}_\alpha / \mathcal{G}^0_\alpha \quad \text{and} \quad \bar{\mathcal{G}}_\alpha := \mathcal{G}_\alpha / \mathcal{G}^0_\alpha, \tag{4.15}$$

where $\mathcal{G}^0_\alpha$ is the subgroup given by all local gauge transformations g_α which are identity on the vertices of α. Let us choose an arbitrary but fixed orientation of each edge of α. Then,

[17] More precisely, a *graph* α is a finite set of compact one-dimensional sub-manifolds of M called *edges* of α, such that (i) every edge is either an embedded interval with boundary (an open edge with end-points); or, an embedded circle with a marked point (a closed edge with an 'end-point'); or an embedded circle (a loop); and, (ii) if an edge intersects any other edge of α it does so only at one or two of its end-points. The end-points of an edge are called vertices. This precise definition is needed to ensure that our Hilbert space $\mathcal{H}$ of section 4.3 is sufficiently large and admits a generalized 'spin network decomposition.

[18] Throughout this paper, we will work with a fixed trivialization. In this section, lower case Greek letters always refer to graphs and not to indices on spacetime fields; indeed, we do not use spacetime fields in most of the remainder of this review.

every element $\bar{A}_\alpha \in \bar{\mathcal{A}}_\alpha$ can be identified with the G values $\bar{A}_\alpha(e)$ of the parallel transport (i.e., holonomy) defined by any connection A_α in the equivalence class $\bar{A}_\alpha$.[19] Thus, we have natural 1–1 maps between $\bar{\mathcal{A}}_\alpha$ and G^n and between $\bar{\mathcal{G}}_\alpha$ and G^m

$$\mathcal{I}_E : \bar{\mathcal{A}}_\alpha \longrightarrow G^n \qquad \mathcal{I}_E(\bar{A}_\alpha) = (\bar{A}_\alpha(e_1), \ldots, \bar{A}_\alpha(e_1)), \tag{4.16}$$

$$\mathcal{I}_V : \bar{\mathcal{G}}_\alpha \longrightarrow G^m \qquad \mathcal{I}_V(\bar{g}_\alpha) = (\bar{g}(v_1), \ldots, \bar{g}(v_m)) \tag{4.17}$$

where $e_1, \ldots, e_n$ are the edges of α and $v_1, \ldots, v_m$ the vertices. Note that $\mathcal{I}_E$ depends on the orientation of edges. In the next section, this map will play a key role and we will ensure that our final results are insensitive to the choice of the orientation.

Following lattice gauge theory, we will refer to $\bar{A}_\alpha$ as a *configuration variable* of theories of connections on α and $\bar{g}_\alpha$ as a (residual) *gauge transformation* on the configuration variables. Since $\bar{\mathcal{G}}_\alpha$, the group of the (residual) gauge transformations, has a non-trivial action on $\bar{\mathcal{A}}_\alpha$, *physical configuration space* is given by the quotient $\bar{\mathcal{A}}_\alpha/\bar{\mathcal{G}}_\alpha$.

Remark. The quotient $\bar{\mathcal{A}}_\alpha/\bar{\mathcal{G}}_\alpha$ can be characterized in the following way [40]. Fix a vertex v_0 in α. Let $\alpha_1, \ldots, \alpha_h$ be free generators of the first homotopy group of α, based at v_0 (that is every loop in α beginning in v_0 is a product of the generators and their inverses, and this decomposition is unique). The map

$$\bar{A} \mapsto (\bar{A}(\alpha_1), \ldots, \bar{A}(\alpha_h)) \tag{4.18}$$

from $\bar{\mathcal{A}}$ to G^h defines a 1–1 correspondence between $\bar{\mathcal{A}}_\alpha/\bar{\mathcal{G}}_\alpha$ and G^m/G where the quotient is with respect to the residual gauge action $(U_1, \ldots, U_h)g := (g^{-1}U_1 g, \ldots, g^{-1}U_h g)$.

4.2.2. Quantum theory. Since $\bar{\mathcal{A}}_\alpha$ is the configuration space, it is natural to represent quantum states as square integrable functions on $\bar{\mathcal{A}}_\alpha$. This requires that we define a measure on $\mathcal{A}_\alpha$. An obvious strategy is to use the map $\mathcal{I}_E$ of (4.16) to represent $\bar{\mathcal{A}}_\alpha$ by G^n and use the Haar measure on G. This endows $\bar{\mathcal{A}}_\alpha$ with a natural measure which we denote by μ_α^0. Thus, the space of quantum states can be taken to be the Hilbert space $\mathcal{H}_\alpha = L^2\big(\bar{\mathcal{A}}_\alpha, \mathrm{d}\mu_\alpha^0\big)$. Let us denote the pullbacks of functions ψ on G^n to functions on $\bar{\mathcal{A}}_\alpha$ by Ψ:

$$\Psi = \mathcal{I}_E^\star \psi. \tag{4.19}$$

Since $\mathcal{I}_E$ is a *bijection*, every function Ψ on $\bar{\mathcal{A}}_\alpha$ can be so represented, enabling us to think of quantum states Ψ in $\mathcal{H}_\alpha$ as functions ψ on G^n. Then, the inner product can be written as

$$\langle \Psi_1, \Psi_2 \rangle = \int_{G^n} \mathrm{d}\mu_H^0 \bar{\psi}_1 \psi_2, \tag{4.20}$$

where μ_H^0 is the Haar measure on G^n. Since the Haar measure is invariant under $g \mapsto g^{-1}$, the inner product does not depend on the choice of the orientation of edges of α, made in the definition of $\mathcal{I}_E$. It is easy to verify that the inner product is also invariant under the induced action on $\mathcal{H}_\alpha$ of the residual group $\bar{\mathcal{G}}_\alpha$ of the gauge transformations.

We will now introduce a number of interesting operators on $\mathcal{H}_\alpha$, which will turn out to be useful throughout this paper. Clearly, $\mathcal{H}_\alpha$ is the tensor product of the spaces $L^2(G, \mathrm{d}\mu_H)$, each associated with an edge of α. Using the operators $\hat{L}_i$ and $\hat{R}_i$ on $L^2(G, \mathrm{d}\mu_H)$ and the fact

[19] For simplicity, in the main body of the paper we will discuss the case where every edge of α has two vertices. If a graph α admits a closed edge e' without vertices, then $\mathcal{G}_\alpha^0$ contains *all* gauge transformations at points of e' (since e' has no vertex). Hence, $\bar{A}_\alpha(e') \in G/\mathrm{Ad}$, where the quotient is by the adjoint action on G. Thus, for a general graph, if n_0 denotes the number of closed edges without vertices and n_1 the remaining edges (so that $n = n_0 + n_1$), the image of the map $\mathcal{I}_E$ defined below is $[G/\mathrm{Ad}]^{n_0} \times G^{n_1}$. All our constructions and results can be extended to general graphs in a straightforward manner.

that the correspondence (4.16) associates a copy of G with each edge in α, we define certain operators $\hat{J}_i^{(v,e)}$ on $\mathcal{H}_\alpha$. Given a vertex v of α, an edge e with v as an end-point, and a basis τ_i in $\mathfrak{g}$, we set

$$\hat{J}_i^{(v,e)}\Psi = \mathcal{I}_E^\star[(1 \otimes \cdots \otimes 1 \otimes \hat{J}_i \otimes 1 \otimes \cdots \otimes 1)\psi] \tag{4.21}$$

where the non-trivial action is only on the copy of G associated with the edge e, and where $\hat{J}_i = \hat{L}_i$ if the vertex v is the origin of the edge e and $\hat{J}_i = \hat{R}_i$ if v is the target of e. Thus, the edge e dictates the copy of G on which $\hat{J}_i^{(v,e)}$ has non-trivial action while the vertex v determines if the action is through the left or right invariant vector field.

4.2.3. Generalized spin network decomposition. The product structure $\mathcal{H}_\alpha \sim [L^2(G, \mathrm{d}\mu_H)]^{\otimes n}$ enables us to import results of the last subsection on quantum mechanics on G. In particular, using (4.12), $\mathcal{H}_\alpha$ can be decomposed into finite-dimensional subspaces $\mathcal{H}_{\alpha,\mathbf{j}}$ where $\mathbf{j} = \{j_1, \ldots j_n\}$ assigns to each edge of α an irreducible representation of G. The individual subspaces $\mathcal{H}_{\alpha,\mathbf{j}}$ can be further decomposed into irreducible representations of the action of the group of residual gauge transformations. Let $\mathbf{l} = \{l_1, \ldots, l_v\}$ assign to each vertex of α an irreducible representation of G. Then, each $\mathcal{H}_{\alpha,\mathbf{j}}$ can be further decomposed into subspaces $\mathcal{H}_{\alpha,\mathbf{j},\mathbf{l}}$ consisting of all vectors which belong to the irreducible representation $\mathbf{l}$ of the group of residual gauge transformations at every vertex v. Then, we have

$$\mathcal{H}_\alpha = \oplus_{\mathbf{j}}\mathcal{H}_{\alpha,\mathbf{j}} = \oplus_{\mathbf{j},\mathbf{l}}\mathcal{H}_{\alpha,\mathbf{j},\mathbf{l}}. \tag{4.22}$$

The gauge invariant subspace of $\mathcal{H}_\alpha$ corresponds to the labelling of vertices

$$\mathbf{l} = \vec{0} \qquad \text{i.e.} \quad \ell_v = \text{the trivial representation} \quad \text{for all } v \text{ in } \alpha. \tag{4.23}$$

For applications to quantum geometry, let us make this decomposition more explicit in the case when $G = SU(2)$. This discussion will also serve to make the somewhat abstract construction given above by providing a more detailed description of the labels $\mathbf{j}, \mathbf{l}$.

Example. With each edge e of α, we associate an operator $\hat{J}_e^2$:

$$(\hat{J}_e)^2 := \eta^{ij}\,\hat{J}_i^{(v,e)}\,\hat{J}_j^{(v,e)} \tag{4.24}$$

where η_{ij} is again the Cartan–Killing metric (4.1) on $su(2)$, and v is the source or target of e. Since they act on different copies of $SU(2)$, all these operators commute with each other. Since each of these operators has eigenvalues $j_e(j_e+1)$ where j_e is a non-negative half-integer, each simultaneous eigenspace $\mathcal{H}_\mathbf{j}$ of this set of operators is labelled by $\mathbf{j} = \left(j_{e_1}, \ldots, j_{e_n}\right)$. Thus, we have a decomposition of the total Hilbert space:

$$\mathcal{H}_\alpha = \oplus_{\mathbf{j}}\mathcal{H}_{\alpha,\mathbf{j}}. \tag{4.25}$$

The individual subspaces $\mathcal{H}_{\alpha,\mathbf{j}}$ are the natural extensions of spaces $\mathcal{S}_j$ introduced in section 4.1.3 in the case of a single copy of $SU(2)$. They have several interesting properties: (i) each $\mathcal{H}_\mathbf{j}$ is a finite-dimensional subspace of $\mathcal{H}_\alpha$; (ii) it is preserved by the action of every $\hat{J}_i^{(v,e)}$; and (iii) it is preserved by the (induced) action of gauge transformations in $\bar{\mathcal{G}}_\alpha$ (which act non-trivially at vertices of α).

Finally, we can carry out a further decomposition by introducing additional commuting operators. Of particular importance are vertex operators $[\hat{J}^v]^2$, associated with each vertex v of α. These are defined by

$$[\hat{J}^v]^2 := \eta^{ij}\,\hat{J}_i^v\,\hat{J}_j^v \qquad \text{where} \quad \hat{J}_i^v := \sum_{e' \text{ at } v} \hat{J}_i^{(v,e')}, \tag{4.26}$$

where the sum extends over all edges e' intersecting at v. Heuristically, $\hat{J}^e_i$ can be regarded as angular momentum operators 'living on the edge e' and $\hat{J}^v_i$, as the *total* angular momentum operators 'arriving' at the vertex v. It is easy to check that the operators $[\hat{J}^v]^2$ commute with the operators $[\hat{J}^e]^2$. Hence, if we denote eigenvalues of $[\hat{J}^v]^2$ by $l_v(l_v+1)$ the subspaces $\mathcal{H}_{\mathbf{j}}$ can be further decomposed and we arrive at a finer decomposition of the total Hilbert space,

$$\mathcal{H}_\alpha = \oplus_{\mathbf{j}} \mathcal{H}_{\alpha,\mathbf{j}} = \oplus_{\mathbf{j},\mathbf{l}} \mathcal{H}_{\alpha,\mathbf{j},\mathbf{l}}, \tag{4.27}$$

where $\mathcal{H}_{\mathbf{j},\mathbf{l}}$ is a simultaneous eigenspace of operators $[\hat{J}^e]^2$ and $[\hat{J}^v]^2$.

Remark. One can enlarge the set of commuting operators and further refine the decomposition of $\mathcal{H}_\alpha$. We illustrate the procedure for $G = SU(2)$. At each vertex v, let us first order the intersecting edges, $(e'_1, \ldots, e'_k)$ say. Then, introduce the following (rather large) set of operators,

$$\begin{aligned}
&\big(\hat{J}^{(v,e'_1)}_i + \hat{J}^{(v,e'_2)}_i\big)\eta^{ij}\big(\hat{J}^{(v,e'_1)}_j + \hat{J}^{(v,e'_2)}_j\big), \ldots,\\
&\big(\hat{J}^{(v,e'_1)}_i + \hat{J}^{(v,e'_2)}_i + \hat{J}^{(v,e'_3)}_i\big)\eta^{ij}\big(\hat{J}^{(v,e'_1)}_j + \hat{J}^{(v,e'_2)}_j + \hat{J}^{(v,e'_3)}_j\big), \ldots,\\
&\qquad\qquad \ldots \quad \ldots\\
&\big(\hat{J}^{(v,e'_1)}_i + \cdots + \hat{J}^{(v,e'_{k-1})}_i\big)\eta^{ij}\big(\hat{J}^{(v,e'_1)}_j + \cdots + \hat{J}^{(v,e'_{k-1}i)}_j\big),
\end{aligned} \tag{4.28}$$

where each bracket contains a sum only over operators defined at a given vertex v. These operators commute with each other and with our earlier operators $[\hat{J}^e]^2$ and $[\hat{J}^v]^2$. If we label the eigenvalues of these operators by $\mathbf{s}$, each simultaneous eigenspace $\mathcal{H}_{(\mathbf{j},\mathbf{l},\mathbf{s})}$ can be labelled by the triplet $(\mathbf{j},\mathbf{s},\mathbf{l})$. Each $\mathcal{H}_{(\mathbf{j},\mathbf{l},\mathbf{s})}$ is the irreducible representation of the group of the gauge transformations $\bar{\mathcal{G}}_\alpha$ corresponding to the half-integer values l_v.

Note that, given arbitrary labelling $\mathbf{j}$ of the edges of α, the remaining two labellings $\mathbf{l}$, $\mathbf{s}$ are restricted by some inequalities: at each vertex v, we must have

$$s_{1,2} \in \big\{\big|j_{e'_1} - j_{e'_2}\big|, \big|j_{e'_1} - j_{e'_2}\big| + 1, \ldots, j_{e'_1} + j_{e'_2}\big\}, \tag{4.29}$$

$$s_{1,2,3} \in \big\{\big|s_{1,2} - j_{e'_3}\big|, \ldots, s_{1,2} + j_{e'_3}\big\} \tag{4.30}$$

$$\ldots \tag{4.30}$$

$$l_v \in \big\{\big|s_{1,2,\ldots,k-1} - j_{e'_k}\big|, \ldots, \big|s_{1,2,\ldots,k-1} + j_{e'_k}\big|\big\}. \tag{4.31}$$

When these conditions are met, we obtain the following orthogonal decomposition of $\mathcal{H}_{\mathbf{j}}$,

$$\mathcal{H}_\alpha = \oplus_j \mathcal{H}_{\mathbf{j}}, \quad \mathcal{H}_{\mathbf{j}} = \oplus_{\mathbf{l}} \mathcal{H}_{(\mathbf{j},\mathbf{l})} \quad \text{and} \quad \mathcal{H}_{(\mathbf{j},\mathbf{l})} \oplus_{\mathbf{s}} \mathcal{H}_{\mathbf{j},\mathbf{l},\mathbf{s}} \tag{4.32}$$

where the labellings $\mathbf{j}, \mathbf{l}$ and $\mathbf{s}$ take positive, half-integer values, subject to the inequalities (4.29)–(4.31). Gauge invariant states subspaces are labelled by trivial $\mathbf{l}$.

4.3. Connections on M

Let us now turn to field theories of G connections, such as general relativity, discussed in section 2. Given any graph α on M, each connection A on M defines, just by restriction, a connection $A\,|_\alpha$ on α. Furthermore, A is completely determined by the collection $\{A|_\alpha\}$ defined by considering all possible graphs α on M. Therefore, we will be able to construct a background-independent quantum kinematics for theories of connections on M by weaving together quantum theories of connections on graphs (constructed in section 4.2).

4.3.1. The classical phase space. For simplicity, we will consider G connections on a *trivial* bundle over M. This restriction is motivated by the fact that the main application of this framework will be to quantum geometry where $G = SU(2)$ and all $SU(2)$ bundles over a 3-manifold are trivial. Since all the structures we introduce are gauge covariant, it is convenient to fix a global trivialization once and for all and regard smooth $\mathfrak{g}$-valued 1-forms A on M as connections[20]. The space of all such 1-forms will be the classical configuration space and denoted by $\mathcal{A}$. The phase space will consist of pairs $\left(A_a^i, P_i^a\right)$, where $A \in \mathcal{A}$ and P is a $\mathfrak{g}$-valued vector density defined on M. Following the standard terminology from Yang–Mills theory, we will refer to A_a^i as *connections* and P_a^i as the analogues of Yang–Mills *electric fields*. As we saw in section 2, in the gravitational case, $kP_i^a = (8\pi G\gamma)P_i^a$ also has the interpretation of an orthonormal triad (of density weight 1), where γ is the Barbero–Immirzi parameter. While this fact plays no role in this section, it will be used crucially in section 5 to introduce quantum Riemannian geometry.

The Poisson bracket between any two smooth functions on the phase space is given by

$$\{f_1(A, P), f_2(A, P)\} = \int_M \mathrm{d}^3x \left(\frac{\delta f_1}{\delta A_a^i}\frac{\delta f_2}{\delta P_i^a} - \frac{\delta f_2}{\delta A_a^i}\frac{\delta f_1}{\delta P_i^a}\right). \tag{4.33}$$

The gauge group $\mathcal{G}$ is the group of G-valued functions g on M. This group has a natural action on the phase space, given by

$$(A \cdot g, P \cdot g) = (g^{-1}Ag + g^{-1}\,\mathrm{d}g, g^{-1}Pg). \tag{4.34}$$

The *elementary* classical observables that will have direct quantum analogues are (complex-valued functions of) holonomies $A(e)$ along paths e in M and fluxes $P(S, f)$ of electric fields (smeared by $\mathfrak{g}$-valued functions f) across 2-surfaces S in M. In this subsection, we will fix our convention, introduce precise definitions of these phase space functions and explore some of their properties.

In the main body of this paper, for technical simplicity we will restrict ourselves to oriented, analytic 3-manifolds M and use only *closed-piecewise analytic edges e and closed-piecewise analytic sub-manifolds S* in M.[21] Given an edge $e : [t_2, t_1] \to M$ on M and a connection A, the parallel transport from $e(t_1)$ to $e(t)$ along e is defined by the following differential equation and initial condition:

$$\frac{\mathrm{d}}{\mathrm{d}t}U_e(t, t_1; A) = -A_a(e(t))\dot{e}^a(t)U(t, t_1; A) \qquad \text{and} \qquad U(t_1, t_1; A) = I. \tag{4.35}$$

Given $A \in \mathcal{A}$, the parallel transport along the entire e will be denoted by $A(e)$:

$$A(e) := U_p(t_2, t_1; A). \tag{4.36}$$

Thus, $A(e) \in G$, it is unchanged under orientation-preserving re-parametrizations of e, and has two key properties which will play an important role in the next subsection,

$$A(e_2 \circ e_1) = A(e_2)A(e_1), \qquad A(e^{-1}) = A(e)^{-1}, \tag{4.37}$$

where e^{-1} is obtained from e by simply reversing the orientation.

[20] For background material see, e.g., [3]. It is quite straightforward to generalize the framework introduced in this subsection to non-trivial bundles on n-dimensional manifold M [41, 58].

[21] More precisely, we assume that for each edge $e : [t_n, t_1] \mapsto M$, the interval $[t_n, t_1]$ admits a covering by closed intervals $[t_n, t_{n-1}], \ldots, [t_2, t_1]$ such that the image of each of these closed intervals in M is analytic. Each surface S is an topological sub-manifold of M such that its closure $\bar{S}$ is of the form $\bar{S} = \cup_I \bar{S}_I$ where each $\bar{S}_I$ is a compact, analytic sub-manifold of M (possibly) with boundary. (These assumptions can be relaxed and one can work with just smooth structures; see [47, 63, 64, 92].)

The 'electric flux' is defined using our surfaces S. Fix on S a smooth function f with values in the dual $\mathfrak{g}^\star$ of the Lie algebra $\mathfrak{g}$ and define the (smeared) flux of P through S as

$$P(S, f) := \int_S f_i \Sigma^i, \tag{4.38}$$

where $\Sigma^i_{ab} = \eta_{abc} P^{ci}$ is the 2-form dual of the electric field.

As a prelude to quantization, let us calculate Poisson brackets between these observables. Since the phase space is a cotangent bundle, the configuration observables have vanishing Poisson brackets among themselves. (As in sections 4.1.2 and 4.2.2, this will make it possible to introduce a configuration representation in which quantum states are functions of connections.) The Poisson bracket between configuration observables $A(e)$ and momentum observables $P(S, f)$ can be easily calculated and has a simple, geometrical structure. Any edge e with $e \cap S \neq \emptyset$ can be trivially written as the union of 'elementary' edges which either lie in S, or intersect S in exactly one of their end-points. (This can be achieved simply by introducing suitable new vertices on e.) Then for each of these 'elementary' edges e which intersect S at a point p, we have

$$\{A(e), P(S, f)\} = -\left[\frac{\kappa(S, e)}{2}\right] \times \begin{cases} A(e)\tau^i f_i(p) & \text{if } p \text{ is thesource of } e \\ -f_i(p)\tau^i A(e) & \text{if } p \text{ is the target of } e \end{cases} \tag{4.39}$$

where τ^i is any orthonormal basis in $\mathfrak{g}$ and $\kappa(S, e)$ is 0 or ± 1:

$$\kappa(S, e) = \begin{cases} 0, & \text{if } e \cap S = \emptyset, \text{ or } e \cap S = e \text{ modulo the end-points} \\ +1, & \text{if } e \text{ lies above } S \\ -1, & \text{if } e \text{ lies below } S. \end{cases} \tag{4.40}$$

Thus, the bracket vanishes if e and S do not intersect or e lies within (the closure of) S and, if they have a 'simple' intersection, is given by a linear combination of the configuration observables $A(e)$, where the coefficients are determined by the value of the smearing field f at the intersection point.

The bracket between the momentum observables, by contrast, is not as straightforward because of the following technical complication [48]. The configuration (holonomy) observables $A(e)$ are obtained by smearing A along one-dimensional curves e while the momentum (electric flux) observables are obtained by smearing the electric field 2-forms e on two-dimensional surfaces. Since connections A are 1-forms and dual-electric fields Σ 2-forms, this smearing is geometrically most natural, particularly when there is no background metric. However, in contrast to the standard practice in field theories where smearing is done in three dimensions, our smearing fields are themselves 'distributional' from the full three-dimensional perspective. Therefore one has to exercise due care in evaluating Poisson brackets. The one- and two-dimensional smearings in the definitions of A_e and $P(S, f)$ are 'just right' for the calculation of the Poisson bracket (4.39) to go through. However, technical subtleties arise in the evaluation of the Poisson brackets between smeared electric fields. If one naively uses Poisson brackets (4.33) to conclude that the Lie bracket between the momentum operators $P(S, f)$ must vanish, then (4.39) implies that the Jacobi identity between $A(e)$, $P(S, f)$, $P(\tilde{S}, \tilde{f})$ fails to be satisfied. The correct procedure is to use the fact that the momentum variables $P(S, f)$ are of the form $X_{(S,f)} \cdot P$ for some vector fields $X_{(S,f)}$ on the configuration space $\mathcal{A}$,[22] whence the (non-trivial) Lie bracket between $P(\tilde{S}, \tilde{f})$ is dictated

[22] More precisely, $X_{(S,f)}$ are derivations on the ring of functions of holonomies (i.e., on the space Cyl defined below). From the perspective of textbook treatments of field theories, these functions are 'singular', being supported on one-dimensional edges rather than three-dimensional open sets of M. This is the origin of the counter-intuitive result on Jacobi identity.

by the action of the vector fields $X_{S,f}$ on the ring of functions of holonomies. Now, in general, these vector fields fail to commute. Hence (as on a general cotangent bundle, see (4.4)) the Poisson bracket between momentum variables fails to vanish in general. As in section 4.1.1, the correct Lie algebra between our elementary configuration and momentum observables is given by the geometric Lie algebra of functions and vector fields on the configuration space $\mathcal{A}$. This Lie algebra naturally incorporates (4.39) and provides non-trivial Poisson brackets between the momentum observables. It will be mirrored in the commutators of the corresponding elementary *quantum* operators.

4.3.2. Quantum configuration space $\bar{\mathcal{A}}$ and Hilbert space $\mathcal{H}$. In quantum mechanics of systems with a finite number of degrees of freedom, states are represented by functions on the classical configuration space. By contrast, as explained in section 3, in field theories quantum states are functions on a larger space—the quantum configuration space. Only certain 'nice' functions on the classical configuration space admit an extension to the larger space. In our case, these are the so-called cylindrical functions on $\mathcal{A}$. Fix a graph α with n edges. Then, given a C^∞ complex-valued function ϕ on G^n, we can define a function Φ_α on $\mathcal{A}$ via

$$\Phi_\alpha(A) = \phi(A(e_1), \ldots, A(e_n)). \tag{4.41}$$

(Strictly, Φ should carry a subscript (α, ϕ). However, for notational simplicity, we will drop ϕ.) The space of such functions will be denoted by Cyl_α.[23] A function Φ on $\mathcal{A}$ is said to be *cylindrical* if it arises from this construction for *some* graph α. Note that (i) there is a natural isomorphism between Cyl_α and the space of functions on $\bar{\mathcal{A}}_\alpha$ (see equation (4.16)); and (ii) every function which is cylindrical with respect to a given graph α is automatically cylindrical with respect to a larger graph α'; $\mathrm{Cyl}_\alpha \subseteq \mathrm{Cyl}_{\alpha'}$. These facts are used repeatedly in this section. We will denote the space of all cylindrical functions by Cyl; thus

$$\mathrm{Cyl} = \cup_\alpha \mathrm{Cyl}_\alpha.$$

Given any one graph α, using (4.20), we can introduce a natural inner product on Cyl_α

$$\langle \Phi_\alpha, \Psi_\alpha \rangle = \int_{G^n} \mathrm{d}\mu_H^0 \bar{\phi}\psi. \tag{4.42}$$

The Cauchy completion of this space provides a Hilbert space which is naturally isomorphic with $\mathcal{H}_\alpha = L^2\big(\bar{\mathcal{A}}_\alpha, \mathrm{d}\mu_\alpha^0\big)$ of section 4.2.2 and for notational simplicity, we will denote it also by $\mathcal{H}_\alpha$.

The idea is to introduce an inner product on the space of *all* cylindrical functions via (4.42). Suppose we are given two cylindrical functions Φ_{α_1} and Ψ_{α_2} based on two distinct graphs α_1 and α_2. Then, we can introduce a third graph α_3 which contains all the edges and vertices of α_1 and α_2, regard the two functions as cylindrical with respect to α_3, and attempt to define the inner product between Φ_{α_1} and Ψ_{α_2} using (4.42) with $\alpha = \alpha_3$. The key question now is whether the resulting number is independent of the specific α_3 used in this construction. Fortunately, the right and left invariance properties of the Haar measure and the fact that we have chosen it to be normalized (so that $\int_G \mathrm{d}\mu_H = 1$) imply that the answer is in the affirmative [40, 41]. Thus, thanks to the Haar measure on G, Cyl has a natural Hermitian inner product. Denote its Cauchy completion by $\mathcal{H}$. This is our Hilbert space for quantum kinematics of background-independent theories of connections.

Since the Cauchy completion $\mathcal{H}_\alpha$ of Cyl_α is simply $L^2\big(\bar{\mathcal{A}}_\alpha, \mathrm{d}\mu_\alpha^0\big)$, every element of $\mathcal{H}_\alpha$ can be represented as a function on $\bar{\mathcal{A}}_\alpha$ (more precisely, an equivalence class of functions on

[23] Occasionally, we will need to let ϕ be only C^n. The space of resulting cylindrical functions Φ will be denoted by $\mathrm{Cyl}_\alpha^{(n)}$.

$\bar{\mathcal{A}}_a$, where two are equivalent if they differ on a set of measure zero). Unfortunately, this is *not* true of $\mathcal{H}$: while every element of Cyl is a function (indeed a very simple one!) on $\mathcal{A}$, in the Cauchy completion, one picks up limit points which *cannot* be represented as functions on $\mathcal{A}$. As noted in section 3, this is a standard occurrence in systems with an infinite number of degrees of freedom (in particular, field theories). It is then natural to ask: is there an enlargement $\bar{\mathcal{A}}$ of $\mathcal{A}$ such that $\mathcal{H}$ is isomorphic with the space of all square integrable functions on $\bar{\mathcal{A}}$ (with respect to some regular Borel measure)? The answer is in the affirmative [39–41]. This $\bar{\mathcal{A}}$ is called the *quantum configuration space*.

Surprisingly, one can give a rather simple characterization of $\bar{\mathcal{A}}$, which will turn out to be extremely useful [44, 66]. Let us denote an element of $\bar{\mathcal{A}}$ by $\bar{A}$ and call it a *quantum connection*. $\bar{A}$ assigns to each edge e in M an element $\bar{A}(e)$ of G such that

$$\bar{A}(e_2 \circ e_1) = \bar{A}(e_2)\bar{A}(e_1) \qquad \text{and} \qquad \bar{A}(e^{-1}) = (\bar{A}(e))^{-1}. \tag{4.43}$$

Thus, every smooth connection A automatically defines a generalized connection (see (4.37)); in this case $\bar{A}(e)$ is just the standard holonomy. However, a general $\bar{A}$ can be *arbitrarily discontinuous*; there are no requirements on it other than (4.43). This is why $\bar{\mathcal{A}}$ is much larger than $\mathcal{A}$. Nonetheless, in a natural topology (due to Gel'fand), $\mathcal{A}$ is *densely* embedded in $\bar{\mathcal{A}}$ and $\bar{\mathcal{A}}$ is compact [40]. Thus, the quantum configuration space $\bar{\mathcal{A}}$ can be naturally thought of as a completion of the classical configuration space $\mathcal{A}$. Finally, we note an important fact that will be used often in regularization procedures in quantum theory: given any quantum connection $\bar{A}$ and any graph α, there exists a smooth connection A such that $\bar{A}(e) = A(e)$ for all edges e of the graph.

Let us now consider quantum states. Now, one can show that the family of induced Haar measures μ^0_α on $\bar{\mathcal{A}}_\alpha$ defines a regular, Borel measure on $\bar{\mathcal{A}}$ [40, 41, 44, 45, 66]. We will denote it by μ^0. As discussed in section 4.3.5, diffeomorphisms on M have a natural induced action on $\bar{\mathcal{A}}$. The measure μ^0 is *invariant* under this action[24]. This is why $\mathcal{H} \equiv L^2(\bar{\mathcal{A}}, \mathrm{d}\mu_0)$ is an appropriate Hilbert space of states for background-independent theories of connections.

Finally, let us highlight the essential steps that lead to μ^0 as this series of steps will be used repeatedly to introduce other structures, such as operators, on $\mathcal{H}$. We begin with spaces $\bar{\mathcal{A}}_\alpha$. Each $\bar{\mathcal{A}}_\alpha$ admits a measure μ^0_α. This family of measures is *consistent* in the sense that, given a function f on $\bar{\mathcal{A}}$ which is cylindrical with respect to graphs α and α',

$$\int_{\bar{\mathcal{A}}_\alpha} \mathrm{d}\mu^0_\alpha f_\alpha = \int_{\bar{\mathcal{A}}_{\alpha'}} \mathrm{d}\mu^0_{\alpha'} f_{\alpha'}. \tag{4.44}$$

It is this consistency that ensures the existence of μ^0 on $\bar{\mathcal{A}}$. More generally, regular Borel measures (as well as geometrical structures) on $\bar{\mathcal{A}}$ are defined by 'gluing together' consistent structures on finite spaces $\bar{\mathcal{A}}_a$ (via the so-called projective techniques [44, 45, 66]). Several families of such measures have been constructed, μ^0 being the simplest and, because of certain uniqueness results [55–58] (discussed at the end of section 4.3.4), the most useful of them.

As emphasized in section 3, the passage from $\mathcal{A}$ to $\bar{\mathcal{A}}$ is highly non-trivial and comes about *because we are going beyond lattice gauge theories and incorporating the infinite number of degrees of freedom in the connection A*. As in Minkowskian field theories, while the classical configuration space $\mathcal{A}$ is densely embedded in the quantum configuration space $\bar{\mathcal{A}}$ (in the natural Gel'fand topology on $\bar{\mathcal{A}}$), measure theoretically, $\mathcal{A}$ is sparse: $\mathcal{A}$ is contained in a set of

[24] Initially, this came as a surprise because, in the mathematical community, there was a widespread expectation that non-trivial diffeomorphism invariant measures do not exist. Note, however, that our μ_0 is defined on $\bar{\mathcal{A}}$ rather than $\mathcal{A}$, the space used in those heuristic arguments. We have included this brief discussion of measure theory to highlight the fact that our constructions are on a sound mathematical footing; in contrast to the habitual situation in the physics literature, our functional integrals are not formal but well defined and finite. The measure $\mathrm{d}\mu_0$ was introduced in [40, 41] and discussed from different perspectives in [45, 44, 66].

zero μ_0 measure [44]. The fact that general quantum states have support on 'non-classical' connection is not a 'mere mathematical technicality': in the Hilbert space language, this is the origin of field theoretic infinities, e.g., the reason why we cannot naively multiply field operators. Hence, to ensure that there are no hidden infinities, it is necessary to pay due attention to the quantum configuration space $\bar{\mathcal{A}}$.

4.3.3. Generalized spin networks. It seems natural to attempt to decompose $\mathcal{H}$ as a direct sum of the Hilbert spaces $\mathcal{H}_\alpha$ associated with various graphs and then use the constructions introduced in section 4.2.2 to carry out further orthogonal decompositions into *finite*-dimensional Hilbert spaces. However, this idea encounters an elementary obstruction. Recall that each function on $\bar{\mathcal{A}}$ which is cylindrical with respect to a graph α is also cylindrical with respect to every larger graph. Thus, regarded as subspaces of $\mathcal{H}$, the Hilbert spaces $\mathcal{H}_\alpha$ cannot be mutually orthogonal. To get around this obstacle, let us introduce new Hilbert spaces: given a graph α, let $\mathcal{H}'_\alpha$ be the subspace of $\mathcal{H}_\alpha$ which is *orthogonal* to the subspace $\mathcal{H}_{\tilde{\alpha}}$ associated with *every* graph $\tilde{\alpha}$ which is *strictly* contained in α. Through the introduction of these $\mathcal{H}'_\alpha$, we remove the undesired redundancy; if $f \in \mathcal{H}'_\alpha$, it cannot belong to $\mathcal{H}'_\beta$ for any graph β distinct from α. While the definition of $\mathcal{H}'_\alpha$ may seem unwieldy at first, using (4.22) it is easy to provide an explicit description of $\mathcal{H}'_\alpha$ which we now describe.

Consider assignments $\mathbf{j}' = \{j'_1, \ldots, j'_n\}$ of irreducible representations of G to edges of e such that each representation is *non-trivial*. Next, let $\mathbf{l}' = \{l'_1, \ldots, l'_{n_v}\}$ denote assignments of irreducible representations to vertices of α which are non-trivial at each *spurious* vertex of α, where a vertex v is spurious if it is bivalent, and if the edges e_i and e_{i+1} which meet at v are such that $e_i \circ e_{i+1}$ is itself an analytic edge (so that v 'just serves to split an edge'). Then, $\mathcal{H}'_\alpha$ is given by

$$\mathcal{H}'_\alpha = \oplus_{\mathbf{j}',\mathbf{l}'} \mathcal{H}'_{\alpha,\mathbf{j}',\mathbf{l}'}. \tag{4.45}$$

The condition on $\mathbf{j}'$ in the definition of $\mathcal{H}'_{\alpha,\mathbf{j}'}$ is necessary because functions in Cyl which result from allowing any of the $(j_1, \ldots, j_n)$ to vanish belong to $\mathcal{H}'_{\tilde{\alpha}}$ where $\tilde{\alpha}$ is a smaller graph obtained by 'removing those edges for which j_i vanished'. The condition on $\mathbf{l}'$ removes the redundancy that would otherwise arise because a function in $\mathcal{H}_{\alpha_1,\mathbf{j}'}$ also defines a function in $\mathcal{H}_{\alpha_2,\mathbf{j}'}$ where α_2 is obtained merely by splitting one or more edges of α_1 just by insertion of new vertices. We can now write the desired decomposition of $\mathcal{H}$:

$$\mathcal{H} = \oplus_\alpha \mathcal{H}'_\alpha = \oplus_{\alpha,\mathbf{j}'} \mathcal{H}'_{\alpha,\mathbf{j}'}. \tag{4.46}$$

Example. Again, for $G = SU(2)$, we can make this procedure more concrete. Given a graph α, the space $\mathcal{H}'_\alpha$ is the subspace of the space $\mathcal{H}_\alpha$ spanned by all the simultaneous eigenvectors of the operators $[\hat{J}^v]^2$ and $[\hat{J}^e]^2$ such that (i) the eigenvalue $j'_e(j'_e+1)$ of $[J^e]^2$ is non-zero for *each* edge e in α; and (ii) the eigenvalue ℓ'_v of $[\hat{J}^v]^2$ is non-zero at each spurious vertex v.

When $G = SU(2)$, the $\mathbf{j}$ and the $\mathbf{l}$ are sets of half-integers, or spins, and the decomposition (4.46) is referred to as the *spin network decomposition* of $\mathcal{H}$. For a general gauge group, it is called the *generalized spin network decomposition* of $\mathcal{H}$. The subspaces $\mathcal{H}_{\alpha,\mathbf{j}'}$ are *finite* dimensional and their elements are referred to as *generalized spin network states* of quantum theories of connection[25] on M. As we will see, spin network subspaces $\mathcal{H}_{\alpha,\mathbf{j}'}$ are left invariant by interesting geometric operators. In this sense, the decomposition (4.46) has a direct physical significance. Because of the sum over all graphs in (4.46), the Hilbert space $\mathcal{H}$ is very large.

[25] Sometimes, the term is used to refer to an orthonormal basis of states in $\mathcal{H}_{\alpha,\mathbf{j}'}$, chosen in a specific calculation. However, the introduction of such a basis requires additional structure. What is naturally available on $\mathcal{H}$ is only the decomposition (4.46) rather than a generalized spin network basis.

Indeed, when it was first constructed, it seemed to be 'too large to be controllable'. However, the later introduction of projective techniques [44, 45, 66], spin networks [60, 61] and the orthogonal decomposition [89] of the Hilbert space showed that quantum theory can, in fact, be developed relatively easily by importing techniques from lattice gauge theories and quantum mechanics of spin systems.

Remark. Trivalent spin networks were already introduced by Penrose in 1971 in a completely different approach to quantum gravity [59]. He expressed his general view of that construction as follows: "I certainly do not want to suggest that the universe 'is' this picture ... But it is not unlikely that essential features of the model I am describing could still have relevance in a more complete theory applicable to more realistic situations". We will see in section 4, trivalent graphs are indeed 'too simple' for semi-classical considerations but Penrose's overall vision *is* realized in a specific and precise way in quantum geometry.

4.3.4. Elementary quantum operators. Recall that in non-relativistic quantum mechanics, typically, one first defines operators on the space $\mathcal{S}$ of smooth functions with rapid decay at infinity and then extends them to self-adjoint operators on the full Hilbert space, $L^2(R^3)$. We will follow a similar strategy; now the role of $\mathcal{S}$ will be played by Cyl.

Let us begin with the configuration operators. Classical configuration variables are represented by complex-valued, cylindrical function f on $\bar{\mathcal{A}}$. We define corresponding quantum operators $\hat{f}$ which also act by multiplication:

$$(\hat{f}\Psi)(\bar{A}) = f(\bar{A})\Psi(\bar{A}). \tag{4.47}$$

Next, let us define momentum operators $\hat{P}_{(S,f)}$, labelled by a 2-surface S and $\mathfrak{g}$-valued smearing fields f^i on S. As with operators $\hat{L}_i$ and $\hat{R}_i$ in section 4.1.2, this action is given just by the Poisson brackets between the classical momentum and configuration observables. For all $\Psi \in \mathrm{Cyl}$, we have

$$(\hat{P}_{(S,f)}\Psi)(\bar{A}) = \mathrm{i}\hbar\{P(S,f),\Psi\}(\bar{A}). \tag{4.48}$$

For later use, let us make the action of the momentum operators explicit. If $\Psi \in \mathrm{Cyl}_\alpha$, we have

$$\hat{P}_{(S,f)}\Psi = \frac{\hbar}{2}\sum_v f^i(v)\left[\sum_{e\,\mathrm{at}\,v}\kappa(S,e)\hat{J}_i^{(v,e)}\Psi\right] \tag{4.49}$$

in terms of $\kappa(S,e)$ of (4.40). With domain $\mathrm{Cyl}^{(2)}$, consisting of twice differentiable cylindrical functions, these operators are essentially self-adjoint (i.e., admit a unique self-adjoint extension) on $\mathcal{H}$. An alternate expression, which brings out the interpretation of $\hat{P}_{(S,f)}$ as the 'flux of the electric field through S', can be given in terms of operators $\hat{J}^{S,v}_{i(u)}$ and $\hat{J}^{S,v}_{i(d)}$ on Cyl_α, associated with a 2-surface S and vertices v of α at which α intersects S (where u stands for 'up' and d for 'down'). If the edges $e_1,\ldots,e_u$ of α lie 'above' S and $e_{u+1},\ldots,e_{u+d}$ lie 'below' S, then we set

$$\begin{aligned}\hat{J}^{S,v}_{i(u)} &= \hat{J}_i^{(v,e_1)} + \cdots + \hat{J}_i^{(v,e_u)},\\ \hat{J}^{S,v}_{i(d)} &= \hat{J}_i^{(v,e_{u+1})} + \cdots + \hat{J}_i^{(v,e_{u+d})}.\end{aligned} \tag{4.50}$$

In terms of these operators, we have

$$\hat{P}_{(S,f)} = \frac{\hbar}{2}\sum_{v\in S} f^i(v)\big(\hat{J}^{S,v}_{i(u)} - \hat{J}^{S,v}_{i(d)}\big), \tag{4.51}$$

where the sum is over all points in S. (The operator is well defined on Cyl because, when acting on a cylindrical function, only a finite number of terms in the uncountable sum are non-zero.)

4.3.5. Gauge and diffeomorphism symmetries. In the classical domain, automorphisms of the bundle on which connections are defined are symmetries of the theory. The group of these symmetries is the semi-direct product of the group of smooth local gauge transformations with the group of smooth diffeomorphisms on M. In this subsection we will examine these symmetries from the quantum perspectives. Modifications arise because on the one hand quantum connections can be arbitrarily discontinuous and on the other hand they are associated only with closed-piecewise analytic edges (see section 4.3.1).

Let us begin with gauge transformations. Given a local G rotation $\bar{g} : M \to G$ there is an active mapping on $\bar{\mathcal{A}}$ defined by

$$\bar{g} \cdot \bar{A}(e) = g(v_+)\bar{A}(e)(g(v_-))^{-1}, \tag{4.52}$$

for all edges e in M with source v_- and target v_+. Note that $\bar{g}$ can be an arbitrarily discontinuous G-valued function on M. We will denote the group of these gauge transformations by $\bar{\mathcal{G}}$. The natural measure μ_0 on $\bar{\mathcal{A}}$ is invariant under $\bar{\mathcal{G}}$, whence the corresponding action of $\bar{\mathcal{G}}$ on $\mathcal{H}$ is unitary:

$$(U_{\bar{g}}\Psi_\alpha)(\bar{A}) = \Psi(\bar{g} \cdot \bar{A}). \tag{4.53}$$

Each of the subspaces $\mathcal{H}'_\alpha$, $\mathcal{H}'_{\alpha,\mathbf{j}'}$ and $\mathcal{H}'_{\alpha,\mathbf{j}',\mathbf{I}'}$ is left invariant by this action. Furthermore, each quantum state in $\mathcal{H}'_{\alpha,\mathbf{j}',\mathbf{I}'=\vec{0}}$ is gauge invariant, i.e., mapped to itself by all $U_{\bar{g}}$. This observation will be useful in section 6.1 to obtain a characterization of the Hilbert space of solutions to the quantum Gauss constraint.

Let us now turn to diffeomorphisms on M. Since we have restricted ourselves to closed-piecewise analytic edges, analytic diffeomorphisms on M have a natural action on $\bar{\mathcal{A}}$. However there is a larger group of maps $\varphi : M \to M$ which has a natural action on Cyl [58]. Let φ be a C^n diffeomorphism of M such that every permissible graph on M is mapped to a permissible graph[26]. Then we can define the action of φ in the space $\bar{\mathcal{A}}$ of the quantum connections, namely

$$\varphi \cdot \bar{A}(e) := \bar{A}(\varphi(e)) \tag{4.54}$$

for all paths e in M. Denote the group of such diffeomorphisms by Diff. Each element φ of this group naturally defines an isomorphism in Cyl. Moreover, the measure μ_0 is Diff invariant, therefore an operator $U_{\bar{\varphi}}$ defined in $\mathcal{H}$ by each $\bar{\varphi}$, namely

$$U_{\bar{\varphi}}\Psi(\bar{A}) := \Psi(\bar{\varphi} \cdot (\bar{A})) \tag{4.55}$$

is unitary. However, under that induced action of Diff, none of the subspaces $\mathcal{H}'_\alpha$, $\mathcal{H}'_{\alpha,\mathbf{j}'}$ and $\mathcal{H}'_{\alpha,\mathbf{j}',\mathbf{I}'}$ is left invariant; they transform covariantly.

The group Diff is a subgroup of all C^n diffeomorphisms but it is considerably larger than the group of the entire analytic diffeomorphisms. The crucial difference is the local character of Diff: for every point $x \in M$ and every open neighbourhood $\mathcal{U}_x$ containing x, there is a $\varphi \in$ Diff which moves x non-trivially within $\mathcal{U}_x$ but is trivial outside $\mathcal{U}_x$ [58]. A generic element of Diff fails to be analytic. Roughly, Diff can be thought of as the group of piecewise analytic, C^n diffeomorphisms. This group will play an important role in the imposition of the diffeomorphism constraint in section 6.2.

Remark. In the kinematic description constructed so far, $\Psi(\bar{A}) = 1$ is the only gauge and diffeomorphism invariant state in $\mathcal{H}$.[27] From this symmetry considerations, one can regard it

[26] Recall that each edge e of a permissible graph is closed-piecewise analytic. Therefore, the requirement is that for every analytic embedding $e : [0, 1] \to M$ of the interval, the image $\varphi(e([0, 1]))$ is a finite sum $\cup_I e_I([0, 1])$, where each $e_I : [0, 1] \to M$ is again an analytic embedding.

[27] Since diffeomorphisms on M are generated by a first class constraint, one would expect that *all* physical states should be diffeomorphism invariant. We will see in section 6 that this expectation is indeed borne out but the physical states belong to $\mathrm{Cyl}^\star$ which is considerably larger than the kinematical Hilbert space $\mathcal{H}$.

as the 'ground state'. It is annihilated by all the momentum (or triad) operators. Elements of Cyl_α represent 'excited states', where the geometry is excited only along the edges of α; the smeared triad $\hat{E}_{S,f}$ has a non-trivial action on these states only if S intersects at least one edge of α. Since these basic excitations are one dimensional, the quantum geometry is said to be *polymer-like*. If the graph has just a few edges, we have a highly quantum mechanical state—the analogue of a state of the quantum Maxwell field with just a few photons. To approximate a classical geometry, one needs a *highly* excited state, with a huge number of edges, criss-crossing M 'very densely'.

Let us summarize our discussion of quantum kinematics for background-independent theories of connections. In section 4.3.1, we introduced a Lie algebra of holonomy and flux functions on the classical phase space [48]. In the subsequent subsections, we constructed a natural, diffeomorphism covariant representation of the quantum analogue of this holonomy–flux algebra. For pedagogical reasons, we chose a constructive approach and developed the theory step by step starting from quantum mechanics on a compact Lie algebra G and passing through the quantum theory of connections on graphs. The actual development of the subject, on the other hand, began with a broader perspective and first principles [39–41]). The main problem is that of finding the physically appropriate representation of the holonomy–flux algebra. The starting point was the observation [39] that Cyl, which serves as the algebra of configuration variables, has the structure of an Abelian $\star$-algebra. By completing it in the sup-norm one obtains the $C^\star$-algebra $\overline{\mathrm{Cyl}}$ of quantum configuration operators. The strategy was to first seek its representations and then represent the momentum operators on the resulting Hilbert spaces. A general theorem due to Gel'fand ensures that *every* representation of $\overline{\mathrm{Cyl}}$ is of the following type: the Hilbert space is the space of square integrable functions on a compact Hausdorff space—called the Gel'fand spectrum of the $C^\star$-algebra—with respect to a regular Borel measure, and the configuration operators act on it by multiplication. The non-trivial fact is that the structure of Cyl is such that the spectrum is easy to exhibit: it is precisely our space $\bar{\mathcal{A}}$ [40]. Thus, the representation of the algebra of elementary variables we constructed step by step is in fact rooted in the general Gel'fand representation theory.

Even though this procedure is quite general and well motivated, one can nonetheless ask why we did not adopt the more general algebraic approach but focused instead on a specific representation. Interestingly, several partial uniqueness theorems have been established indicating that the requirement of general covariance suffices to select a unique cyclic representation of the kinematical quantum algebra [55–58]! This is the quantum geometry analogue of the seminal results by Segal and others that characterized the Fock vacuum in Minkowskian field theories. However, while that result assumed not only Poincaré invariance but also specific (namely free) dynamics, it is striking that the present uniqueness theorems make no such restriction on dynamics. Thus, the quantum geometry framework is surprisingly tight. These results seem to suggest that, for background-independent theories, the full generality of the algebraic approach may be unnecessary: if there is a unique diffeomorphism invariant representation, one might as well restrict oneself to it. For non-trivially constrained systems such as general relativity, this is fortunate because a satisfactory and manageable algebraic treatment of theories with such constraints is yet to become available.

5. Quantum Riemannian geometry

In this section, we will introduce simple geometric operators on $\mathcal{H}$. Recall from section 2 that the internal group for the phase space of general relativity is $SU(2)$ and the Riemannian geometry is coded in the triad field $\tilde{E}^a_i = k\gamma P^a_i \equiv 8\pi G\gamma P^a_i$ (of density weight 1), where

$\gamma > 0$ is the Barbero–Immirzi parameter (see (2.21)). Therefore, in the quantum theory we set $G = SU(2)$ and our geometric operators are built from the (smeared) triad operators $\hat{P}_{(S,f)}$. Because of space limitation, we will only discuss the area and the volume operators [65–78] which have had direct applications, e.g., in the entropy calculations and quantum dynamics. For the length operator, see [79].

5.1. Area operators

Let S be either a closed two-dimensional sub-manifold of M or an open two-dimensional sub-manifold without boundary. In the classical theory, its area is a function on the phase space given by $A(S) = \int_S \mathrm{d}^2x\,\sqrt{h}$, where h is the determinant of the intrinsic 2-metric h_{ab} on S. Our task is to construct the quantum operator corresponding to this phase space function and analyse its properties. (For further details, see [72]).

5.1.1. Regularization. A natural strategy is to first re-express A_S in terms of the 'elementary' observables $P(S, f)$, and then replace each $P(S, f)$ by its unambiguous quantum analogue. This strategy naturally leads to a regularization procedure which we now summarize.

Let us divide S into a large number of elementary cells, S_I, with $I = 1, 2, \ldots, N$. On each cell, introduce an internal triad τ^i and, using its elements as test fields f^i, set $P(S_I, \tau^i) = P^i(S_I)$. Next, recall from (2.21) that the orthonormal triad $\tilde{E}^a_i$ of density weight 1 is related to the momentum field P^a_i via $\tilde{E}^a_i = 8\pi G\gamma P^a_i$. Set

$$[A_S]_N = 8\pi G\gamma \sum_{I=1}^{N} \sqrt{P^i(S_I)P^j(S_I)\eta_{ij}} \tag{5.1}$$

where η_{ij} is again the Cartan–Killing metric on $su(2)$. Then, $[A_S]_N$ is an approximate expression of the area $A(S)$ in the following sense: as the number of cells goes to infinity such that the coordinate size of the cells S_I goes to zero uniformly in I, we have

$$\lim_{N\to\infty} [A_S]_N = A_S. \tag{5.2}$$

Since each $P^i(S_I)$ gives rise to an unambiguously defined quantum operator, $[A_S]_N$ represents a suitable 'regularized area function' and the limit $N \to \infty$ corresponds to the operation of removing the regulator. In the quantum theory, then, we first define an approximate area operator by first noting that, for each I, $\hat{P}^i(S_I)\hat{P}^j(S_I)\eta_{ij}$ is a positive definite self-adjoint operator on $\mathcal{H}$ with a well-defined (positive) square-root, and setting

$$[\hat{A}_S]_N := 8\pi G\gamma \sum_{I=1}^{N} \sqrt{\hat{P}^i(S_I)\hat{P}^j(S_I)\eta_{ij}}. \tag{5.3}$$

To obtain an explicit expression of this operator, let us restrict its action to Hilbert space $\mathcal{H}_\alpha$ associated with any one graph α. Let us first refine the partition sufficiently so that every elementary cell S_I intersects α transversely at most at one point (or contains a segment of α). Then, using expression (4.49) of the smeared triad operators, we conclude that a non-zero contribution to the sum in (5.3) comes only from those S_I which intersect α and, furthermore, a subsequent refinement of the partition does not change the result. Thus, for any given α, the limit $N \to \infty$ is reached already at a finite step; somewhat surprisingly, the removal of the regulator can be achieved rather easily in the quantum theory. It is straightforward to verify that the resulting operator $\hat{A}_{S,\alpha}$ is given by

$$\hat{A}_{S,\alpha} = 4\pi\gamma\ell^2_{\mathrm{Pl}} \sum_v \sqrt{-\Delta_{S,v,\alpha}} \tag{5.4}$$

where v ranges through all the vertices of α which lie on S and the 'vertex Laplace operator' $\Delta_{S,v,\alpha}$ is defined on Cyl_α as

$$\Delta_{S,v,\alpha} = -\big(\hat{J}^{S,v}_{i(u)} - \hat{J}^{S,v}_{i(d)}\big)\big(\hat{J}^{S,v}_{j(u)} - \hat{J}^{S,v}_{j(d)}\big)\eta^{ij}. \tag{5.5}$$

(The 'up' and 'down' operators $\hat{J}^{S,v}_{i(u)}$ and $\hat{J}^{S,v}_{i(d)}$ are defined in (4.50).) Thus, on each $\mathcal{H}_\alpha$ we have obtained a non-negative, self-adjoint area operator $\hat{A}_{S,\alpha}$.

The question is if these operators can be glued together to obtain a well-defined area operator on the full Hilbert space $\mathcal{H}$. As in the definition of measures on $\bar{\mathcal{A}}$, this is a question about *consistency* of the family. More precisely, suppose an element Ψ of Cyl belongs to both Cyl_{α_1} and Cyl_{α_2}, for two different graphs α_1 and α_2. The question is whether $\hat{A}_{S,\alpha_1}\Psi$ equals $\hat{A}_{S,\alpha_2}\Psi$ as an element of Cyl. The answer is in the affirmative. Thus, there is a non-negative, self-adjoint operator $\hat{A}_S$ on $\mathcal{H}$ whose restriction to Cyl_α is given by (5.4) for any graph α.

In fact, we can also define an *area element operator* corresponding to $\sqrt{h(x)}$ whose integral over S gives the total area operator $\hat{A}_S$. Fix a point $x \in S$, and consider a refinement such that x is contained in the interior of a cell S_I. Introduce an approximate area element $[\sqrt{h(x)}]_N$ via

$$\sqrt{[h(x)]_N} = \frac{8\pi G\gamma}{\epsilon^2}\sqrt{P^i(S_I)P^j(S_I)\eta_{ij}} \tag{5.6}$$

where ϵ^2 is the coordinate area of the cell S_I. As we let N tend to infinity, shrinking the coordinate size of the cells uniformly, $\sqrt{h(x)_N}$ tends to $\sqrt{h(x)}$. As before, we can pass to a regularized quantum operator

$$\widehat{[\sqrt{h(x)}]_N} = \frac{8\pi G\gamma}{\epsilon^2}\sqrt{\hat{P}^i(S_I)\hat{P}^j(S_I)\eta_{ij}} \tag{5.7}$$

simply by replacing the smeared P with corresponding operators. Finally, we remove the regulator. The result is a well-defined operator-valued distribution on $\mathcal{H}$, whose action on Cyl_α is given by

$$\widehat{\sqrt{h_S(x)}} = 4\pi\gamma\ell^2_{\mathrm{Pl}}\sum_v \delta^2_{(S)}(x, v)\sqrt{-\Delta_{S,v,\alpha}} \tag{5.8}$$

where $\delta^2_{(S)}(x, v)$ is the two-dimensional Dirac distribution on S and the sum is over intersections v of α and S. Again, this family of operators is consistent and therefore defines an operator $\widehat{\sqrt{h_S(x)}}$ on $\mathcal{H}$.

Remark. In the definition of the momentum operators $\hat{P}(S, f)$ and area operators $\hat{A}_S$, we only considered 2-manifold S without boundary. Now, if we subdivide S as $S = S' \cup I \cup S''$, where S' and S'' are two two-dimensional sub-manifolds without the boundaries and I is a one-dimensional sub-manifold without boundary, then while classically $A_S - (A_{S'} + A_{S''}) = 0$, because of the distributional nature of quantum geometry, $\hat{A}_S - (\hat{A}_{S'} + \hat{A}_{S''})$ is non-zero since its action on graphs with edges passing through I is non-trivial. To obtain additivity of areas, it is then natural to regard this operator as defining the quantum area $\hat{A}_{S,I}$ of I, although I is a 1-manifold. Proceeding in this manner, one is led to assign a quantum area operator $\hat{A}_{V,S}$ also to a point v of S. Detailed examination shows that this operator is just $4\pi\gamma\ell^2_{\mathrm{Pl}}\sqrt{-\Delta_{S,v}}$. From this perspective, then, $\hat{A}_S = \sum_v \hat{A}_{S,v}$; the quantum area of a surface is obtained by summing up the 'areas associated with all points' in it!

5.1.2. Properties of area operators. In each Cyl_α, the area operator is defined by

$$\hat{A}_{S,\alpha} = \int_S \mathrm{d}^2x \widehat{\sqrt{h_{S,\alpha}}}(x) = 4\pi\gamma\ell_{\mathrm{Pl}}^2 \sum_v \sqrt{-\Delta_{S,v,\alpha}}. \tag{5.9}$$

With domain $\mathrm{Cyl}_\alpha^{(2)}$, consisting of twice differentiable functions on $\bar{\mathcal{A}}_\alpha$, this operator is essentially self-adjoint on $\mathcal{H}_\alpha$. Since this family of operators is consistent, the resulting area operator, with domain $\mathrm{Cyl}^{(2)}$, is also essentially self-adjoint on $\mathcal{H}$. By inspection, the operator is gauge invariant (i.e. commutes with the vertex operators $\hat{J}_i^v$ generating $SU(2)$ gauge rotations at vertices v). Since its definition does not require a background structure, it is diffeomorphism covariant.

The eigenvalues of the operator are given by finite sums

$$a_S = 4\pi\gamma\ell_{\mathrm{Pl}}^2 \sum_I \sqrt{-\lambda_I}, \tag{5.10}$$

where λ_I are arbitrary eigenvalues of the operators Δ_{S,v_I}. Now, this operator can be cast in a convenient form as a sum of three commuting operators,

$$-\Delta_{S,v} = 2\big(\hat{J}_{S,v}^{(d)}\big)^2 + 2\big(\hat{J}_{S,v}^{(u)}\big)^2 - \big(\hat{J}_{S,v}^{(u)} + \hat{J}_{S,v}^{(d)}\big)^2, \tag{5.11}$$

which makes its eigenvalues transparent. These are given by

$$-\lambda = 2j^{(u)}(j^{(u)}+1) + 2j^{(d)}(j^{(d)}+1) - j^{(u+d)}(j^{(u+d)}+1), \tag{5.12}$$

where $j^{(u)}$, $j^{(d)}$ and $j^{(u+d)}$ are arbitrary half-integers subject to the standard condition

$$j^{(u+d)} \in \{|j^{(u)} - j^{(d)}|, |j^{(u)} - j^{(d)}| + 1, \ldots, j^{(u)} + j^{(d)}\}. \tag{5.13}$$

Thus, the general eigenvalues of the area operator are given finite sums,

$$a_S = 4\pi\gamma\ell_{\mathrm{Pl}}^2 \sum_I \sqrt{2j^{(u)}(j^{(u)}+1) + 2j^{(d)}(j^{(d)}+1) - j^{(u+d)}(j^{(u+d)}+1)} \tag{5.14}$$

where the j are subject to the constraint (5.13). Thus, all eigenvalues are discrete and the area gap—the smallest non-zero eigenvalue a_S—is given by

$$\Delta a_S = 4\pi\gamma\ell_{\mathrm{Pl}}^2 \frac{\sqrt{3}}{2}. \tag{5.15}$$

The level spacing between consecutive eigenvalues is *not* uniform but decreases *exponentially* for large eigenvalues. This implies that, although the eigenvalues *are* fundamentally discrete, the continuum approximation becomes excellent *very rapidly*. On the full kinematic Hilbert space $\mathcal{H}$—as opposed to the gauge invariant subspace considered below—all these properties are insensitive to the topology of S.

5.1.3. The gauge invariant subspace. Let us now restrict ourselves to gauge invariant subspace $\mathcal{H}_{\mathrm{inv}}$ of $\mathcal{H}$. This is spanned by elements of Cyl which have zero eigenvalue for every vertex operator $\hat{J}_i^v$ (i.e., in the terminology of section 4.2, states in the subspaces $l_v = 0$ for all vertices v). Now, the spectrum of the area operator $\hat{A}_S$ *depends on some global properties of* S. If the closure of S is a manifold with a non-trivial boundary, then the spectrum is the same as in (5.14). However, if $\partial S = \emptyset$, then gauge invariance imposes certain additional conditions on the total spin 'coming in' S.

Let us focus on this case. Suppose first, that S divides M into two disjoint open sets (as would happen if M were R^3 and S a 2-sphere in it). Then the spins $j_I^{(u)}$, $j_I^{(d)}$ in (5.12) have to satisfy the following condition,

$$\sum_I j_I^{(u)} \in \mathbb{N}, \qquad \sum_I j_I^{(d)} \in \mathbb{N}, \tag{5.16}$$

where $\mathbb{N}$ is the set of natural numbers. In the case when S has no boundary but $M\backslash S$ is connected (as can happen if M is a 3-torus and S a 2-torus in it) the condition is milder,

$$\sum_I j_I^{(u)} + \sum_I j_I^{(d)} \in \mathbb{N}. \tag{5.17}$$

In particular, in these cases, the area gap increases. In the first case, it is given by $4\pi\gamma\ell_{\rm Pl}^2$ $(\sqrt{2})$ while in the second case, by $4\pi\gamma\ell_{\rm Pl}^2$. Thus, there is an interesting interplay between topology of (M, S) and the area gap.

If there are no fermionic fields, then all physically relevant states lie in the gauge invariant subspace $\mathcal{H}_{\rm inv}$ of $\mathcal{H}$ now under consideration. However, in the presence of fermions, the gravitational part of the state by itself will not be gauge invariant at vertices where fermions are located. In particular, then, if there are fermions in the interior of S (say when S is a 2-sphere) the area eigenvalues of S are less restricted and we can 'detect' the presence of these fermions from these eigenvalues!

Remarks

(i) Fix a surface S and consider only those states in Cyl for which the graph has no edge which lies within S and which are gauge invariant at each vertex where S intersects the graph. (This is, in particular, the case if all intersections of S with the graph are at simple bivalent vertices.) In this case, $j_I^{(u+d)} = 0$ and $j_I^{(u)} = j_I^{(d)}$, and the area spectrum simplifies considerably to

$$a_S = 8\pi\gamma\ell_{\rm Pl}^2 \sum_I \sqrt{j_I(j_I+1)}. \tag{5.18}$$

It was first believed, incorrectly, that these are all the area eigenvalues. However, in the case of an isolated horizon, only these eigenvalues are relevant and hence, even now, one often sees only this expression in the literature in place of the complete spectrum (5.14).

(ii) It follows from the definition (5.4) of area operators that $\hat{A}_S$ and $\hat{A}_{S'}$ fail to commute if the surfaces S and S' intersect. This is a striking property because it implies that the Riemannian geometry operators cannot all be diagonalized simultaneously[28]. At one level this is not surprising because, even in quantum mechanics, if the configuration space is a non-trivial manifold, in general the momentum representation does not exist. However, this result brings out a fundamental tension between connection-dynamics and geometrodynamics. As we saw, quantum connection-dynamics is very 'tight'; once we choose the holonomies $A(e)$ and the 'electric fluxes' $P(S, f)$ as basic variables, there is essentially no freedom in the background-independent quantization. Thus, under these seemingly mild assumptions, one is led to conclude that the metric representation does not exist (at least in the obvious sense). Although this non-commutativity is of considerable conceptual interest, in semi-classical states the expectation value of the commutator would be extremely small; the non-commutativity appears to have no observable effects except at the Planck scale [48].

5.2. Volume operators

Let R be an open subset of M. In the classical theory, its volume is a function on the phase space given by $V_R = (\sqrt{8\pi G\gamma})^3 \int_R {\rm d}^3x\sqrt{|\det P|}$ (see (2.25)). Our task is to construct the

[28] Thus, the assertion that 'the spin network basis diagonalizes all geometrical operators' that one sometimes finds in the literature is incorrect. As we saw in section 4.3.3, while there is a natural spin network decomposition (4.46) of $\mathcal{H}$, there is no natural spin network basis. Given a surface S, we *can* find a spin network basis which diagonalizes $\hat{A}_S$ but there is no basis which diagonalizes area operators associated with all surfaces.

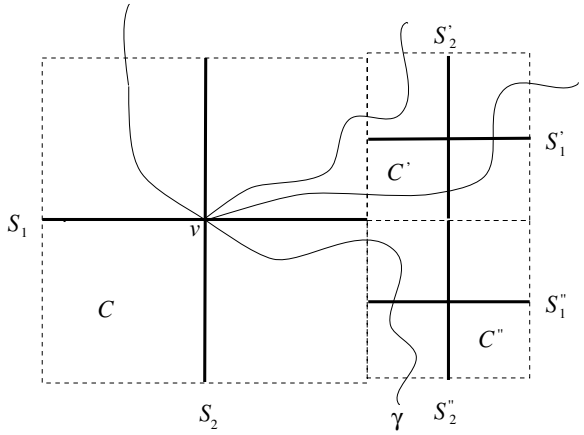

Figure 1. The figure illustrates a partition $\mathcal{P}_\epsilon$ with cells C, C', C'' (the dashed lines) and 2-surfaces S_a, S'_a, S''_a (the thick lines). v is a vertex of a graph γ. For simplicity, one dimension has been dropped.

quantum operator corresponding to this phase space function and analyse its properties. (For further details, see [70, 71, 73, 78]).

5.2.1. Regularization. As in the case of the area operator, we will first recast the classical expression of V_R in terms of the 'elementary' observables $P(S, f)$, and then replace each $P(S, f)$ by its unambiguous quantum analogue. This will provide the regularized volume operator. However the final step, in which the regulator is removed, turns out to be technically more subtle than that in section 5.1.1 and will require an additional construction.

Let us fix a coordinate system (x^a) in R and a positive number ϵ. We then define a partition $\mathcal{P}_\epsilon$ of R as follows. Divide R into a family $\mathcal{C}$ of cells C such that each C is a cube with volume less than ϵ in the given coordinate system and two different cells share only points on their boundaries. In each cell C, introduce three 2-surfaces $s = (S^1, S^2, S^3)$, such that each of the surfaces splits C into two disjoint parts, and $x^a|_{S^a} = \text{const}$ for $a = 1, 2, 3$. The family of pairs (C, s) defines $\mathcal{P}_\epsilon$ (see figure 1). Given a partition $\mathcal{P}_\epsilon$, we can introduce an approximate expression of the volume V_R:

$$V_{\mathcal{R}}^{\mathcal{P}_\epsilon} = \sum_{C\subset\mathcal{R}} \sqrt{|q_s|} \qquad \text{where} \quad q_{C,s} = \frac{(8\pi G\gamma)^3}{3!}\epsilon^{ijk}\eta_{abc}P^i(S^a)P^j(S^b)P^k(S^c). \tag{5.19}$$

It is easy to verify that

$$\lim_{\epsilon\to 0} V_R^{\mathcal{P}_\epsilon} = V_R,$$

the dependence on the coordinate system and the partition disappears in the limit.

To pass to the quantum theory, we need to define a consistent family of volume operators $\hat{V}_{R,\alpha}$, one for each graph α. Let us then fix a graph α and consider a partition $\mathcal{P}_\epsilon$ such that each vertex v of α is the intersection point of the triplet of 2-surfaces S^1, S^2, S^2 in some cell C_V. Then, we can easily promote the approximate volume function $V_R^{\mathcal{P}_\epsilon}$ to a quantum operator $\hat{V}_{R,\alpha}^{\mathcal{P}_\epsilon}$:

$$\hat{V}_{R,\alpha}^{\mathcal{P}_\epsilon} = \sum_{C\subset\mathcal{R}} \sqrt{|\hat{q}_{s,C}|}, \qquad \text{where} \quad \hat{q}_s = \frac{(8\pi G\gamma)^3}{3!}\epsilon_{ijk}\eta_{abc}\hat{P}^i(S^a)\hat{P}^j(S^b)\hat{P}^k(S^c). \tag{5.20}$$

This operator is well defined on $\mathrm{Cyl}^{(3)}_\alpha$, the space of thrice differentiable functions on $\bar{\mathcal{A}}_\alpha$. Furthermore, the limit $\epsilon \to 0$ of the operator exists. However, unlike in the classical theory, it carries a memory of the partition $\mathcal{P}_\epsilon$ used in the regularization process. This is a new complication which did not occur in the case of the simpler area operators. But one can handle it simply by averaging with respect to the essential background structures, used in the construction of the partition $\mathcal{P}_\epsilon$, prior to the removal of the regulator. This extra step can be carried out in detail [73]. The resulting operator is given by

$$\hat{V}_{R,\alpha} = \kappa_0 \sum_v \sqrt{|\hat{q}_{v,\alpha}|},$$

where

$$\hat{q}_{v,\alpha} = \left(8\pi\gamma\ell_{\mathrm{Pl}}^2\right)^3 \frac{1}{48} \sum_{e,e',e''} \epsilon_{ijk}\epsilon(e,e',e'')\hat{J}_i^{(v,e)}\hat{J}_j^{(v,e')}\hat{J}_k^{(v,e'')}. \tag{5.21}$$

Here κ_0 is an undetermined overall constant resulting from the averaging (usually set equal to 1); v runs over the set of vertices of α; each of e, e' and e'', over the set of edges of α meeting at v; and $\epsilon(e,e',e'')$ is the orientation factor. (Thus, $\epsilon(e,e',e'') = 0$ if the tangent vectors to the three edges are planar, i.e., lie in a 2-plane, at v, and ± 1 if the orientation they define is the same as or opposite to the fiducial orientation of M.) It is straightforward to verify that this family of operators is consistent and hence defines a single densely defined operator $\hat{V}_{\mathcal{R}}$ on $\mathcal{H}$ with domain $\mathrm{Cyl}^{(3)}$.

Again, it is also meaningful to introduce in each $\mathrm{Cyl}^{(3)}_\alpha$ a *volume element operator*

$$\widehat{\sqrt{q(x)}}_\alpha = \kappa_0 \sum_v \delta^{(3)}(x,v)\sqrt{|\hat{q}_v,\alpha|}. \tag{5.22}$$

The family of these operators is consistent and we thus have a densely defined operator $\widehat{\sqrt{q(x)}}$ on $\mathcal{H}$ satisfying $\hat{V}_R = \int_R \mathrm{d}^3x\,\widehat{\sqrt{q(x)}}$.

Finally, in the classical theory the Poisson bracket $\{A_a^i(x), V_R\}$ between the connection at a point x and the volume of a region containing that point is proportional to the co-triad $e_a^i(x)$. This fact has been exploited to introduce co-triad operators which, in turn, have led to a definition of the length operator [79] and features prominently in the discussion of quantum dynamics of section 6.3 [95–98].

5.2.2. Properties of volume operators. By inspection, $\hat{V}_{\mathcal{R}}$ is gauge invariant and covariant with respect to diffeomorphisms of M: under $\varphi : M \to M$, we have $\hat{V}_R \to \hat{V}_{\varphi\cdot R}$. The *total* volume operator $\hat{V}_M$ is diffeomorphism invariant. Hence, its action is well defined on the diffeomorphism invariant subspace of $\mathrm{Cyl}^\star$. This property plays an important role in the analysis of the Hamiltonian constraint.

Because of the presence of $\epsilon(e,e',e'')$ in (5.21), it is clear that $\hat{q}_{v,\alpha} = 0$ if all edges meeting at v are planar. In particular, then, $\hat{q}_{v,\alpha} = 0$ if v is a bivalent vertex. *More surprisingly, $\hat{q}_{v,\alpha}\Psi = 0$ also for every trivalent vertex v provided the state Ψ is invariant with respect to the gauge transformations at v* [67]. Indeed, let e, e', e'' be the edges of α which meet at v, and, for definiteness, suppose that $\epsilon(e,e',e'') = 1$. Then, using gauge invariance at v, i.e., $\hat{J}_i^{(v,e)} + \hat{J}_i^{(v,e')} + \hat{J}_i^{(v,e'')} = 0$, we obtain

$$\begin{aligned}
\frac{48}{\left(8\pi\gamma\ell_{\mathrm{Pl}}^2\right)^3}\hat{q}_{v,\alpha}\Psi &= \epsilon^{ijk}\hat{J}_i^{(v,e)}\hat{J}_j^{(v,e')}\hat{J}_k^{(v,e'')}\Psi \\
&= -\epsilon^{ijk}\hat{J}_i^{(v,e)}\left(\hat{J}_j^{(v,e)} + \hat{J}_j^{(v,e'')}\right)\hat{J}_k^{(v,e'')}\Psi \\
&= -2\mathrm{i}\left(\hat{J}_j^{(v,e)}\hat{J}_k^{(v,e'')}\eta^{jk} - \hat{J}_i^{(v,e)}\hat{J}_j^{(v,e'')}\eta^{ij}\right)\Psi \\
&= 0.
\end{aligned} \tag{5.23}$$

As with the area operator, it is easy to show that all eigenvalues of $\hat{q}_{v,\alpha}$, $\hat{V}_{R,\alpha}$ and $\hat{V}_R$ are real and discrete. The spectrum (i.e., the set of all eigenvalues) of $\hat{V}_R$ is the same irrespective of the specific open region R. Given a point $x \in M$, the spectrum (i.e., the complete set of eigenvalues) of $\widehat{\sqrt{q(x)}}$ is simply given by the union of the spectra of the restrictions of $\hat{q}_x$ to each of the (finite-dimensional) spin network spaces $\mathcal{S}_{\alpha,\mathbf{j},\mathbf{I}}$ in the orthogonal decomposition of $\mathcal{H}$. Because of this property, many eigenvalues and eigenvectors of the volume operators $\hat{V}_R$ have been calculated in a number of special cases [70, 74, 76, 77]. However the complete spectrum, or even an estimate of how the number of eigenvalues grows for large volumes, is not yet known.

On the space of gauge invariant states, the simplest eigenvectors arise from 4-valent vertices. Even in this case, the full set of eigenvalues is not known. However, a technical simplification enables one to calculate the matrix element of the volume operator, which have been useful in the analysis of the quantum Hamiltonian constraint. Let v be a 4-valent vertex of α at which edges e, e', e'', e''' meet and consider the action of $\hat{q}_{v,\alpha}$ on the subspace $S_{\alpha,\mathbf{j},\mathbf{I}}$ with $l_v = 0$. Then,

$$\hat{q}_{v,\alpha} = \left(8\pi\gamma\ell_{\mathrm{Pl}}^2\right)^3 \tfrac{1}{8}\kappa(e, e', e'', e''')\epsilon^{ijk}\hat{J}_i^{(v,e)}\hat{J}_j^{(v,e')}\hat{J}_k^{(v,e'')} \tag{5.24}$$

where $\kappa(e, e', e'', e''') \in \{-2, -1, 0, 2, 3, 4\}$ depending on the diffeomorphism class of the four tangent vectors at v. Using gauge invariance at v, the expression can be cast in the form

$$\hat{q}_{v,\alpha} = \left(8\pi\gamma\ell_{\mathrm{Pl}}^2\right)^3 \frac{1}{32\mathrm{i}}\kappa(e, e', e'', e''')[(\hat{J}^{(v,e')} + \hat{J}^{(v,e'')})^2, (\hat{J}^{(v,e)} + \hat{J}^{(v,e'')})^2], \tag{5.25}$$

which simplifies the task of calculating its matrix elements in the subspace $\mathcal{H}_{\alpha,\mathbf{j},\mathbf{I}}$ [78].

Finally, we note a property of the volume operator which plays an important role in quantum dynamics. Let $R(x, \epsilon)$ be a family of neighbourhoods of a point $x \in M$. Then, given any element Ψ of $\mathrm{Cyl}_\alpha^{(3)}$,

$$\lim_{\epsilon\to 0} \hat{V}_{R(x,\epsilon)}\Psi$$

exists but is not necessarily zero. This is a reflection of the 'distributional' nature of quantum geometry.

5.2.3. 'External' regularization. Since the basic momentum variables are smeared on 2-surfaces, in the regularization procedure for defining geometric operators one invariably begins by re-expressing the geometric functions on the classical phase space in terms of $P(S, f)$. However, there is considerable freedom in achieving this and, while different expressions may yield the same function on the classical phase space when the regulator is removed, their quantum analogues need not share this property. This is the standard 'factor ordering problem' of quantum theory. In particular, in the procedure summarized in section 5.2.1, we expressed the volume of each elementary cell C in terms of three 2-surfaces S^a $(a = 1, 2, 3)$ which lie *inside* that cell. This strategy goes under the name 'internal regularization'. A natural alternative is to use the six 2-surfaces $\tilde{S}^\alpha, \alpha = 1, \ldots, 6$ which *bound* the cell. This 'external regularization' strategy was first introduced by Rovelli and Smolin [65] for gauge invariant states on trivalent graphs. Although it was later realized that this volume operator is identically zero on these states [68], Rovelli and de Pietri [70] showed that the method extends also to non-trivial situations.

A detailed analysis [73] shows that this strategy is equally viable, once due attention is paid to the convergence issues (that arise while removing the regulator) by carefully constructing the partition of R. Then, the final volume operator is again of the form (5.21) but

given by

$$\hat{V}^{\rm Ext}_{R,\alpha} = \kappa_0 \sum_v \sqrt{|\hat{q}^{\rm Ext}_{v,\alpha}|}, \qquad \text{where} \quad \hat{q}^{\rm Ext}_{v,\alpha} = \left(8\pi\gamma\ell^2_{\rm Pl}\right)^3 \frac{1}{48} \sum_{e\neq e'\neq e''\neq e} \epsilon^{ijk}\, \hat{J}^{(v,e)}_i\, \hat{J}^{(v,e')}_j\, \hat{J}^{(v,e'')}_k. \tag{5.26}$$

A state Φ which is cylindrical with respect to a graph α and gauge invariant at a trivalent vertex v is again annihilated by the new $\hat{q}^{\rm Ext}_{v,\alpha}$. Furthermore, at gauge invariant 4-valent vertices, $\hat{q}^{\rm Ext}_{v,\alpha}$ agrees with $\hat{q}_{v,\alpha}$, *modulo* a multiplicative factor which depends on the diffeomorphism class of the tangent vectors at v. In spite of this close relation on simple states, the two operators are fundamentally different because of the absence of the orientation factor $\epsilon(e, e', e'')$ in (5.26). In particular, because of this factor, the operators constructed in section 5.2.1 know about the differential structure at vertices of graphs. By contrast, the action of (5.26) is 'topological'.

In the literature, internally regulated operators of section 5.2.1 are used more often. For example, Thiemann's analysis of properties volume operators in the continuum [78] and Loll's analysis on lattices [74, 75] refer to (5.21). The same is true of the volume operators used by Thiemann and others in the discussion of the Hamiltonian constraint.

6. Quantum dynamics

The quantum geometry framework provides the appropriate arena for a precise formulation of quantum Einstein equations. As indicated in section 2, because of the difficult problems of background-independent regularization of products of operator-valued distributions, quantum Einstein equations still remain formal in geometrodynamics. In connection-dynamics, by contrast, we have a well-defined Hilbert space $\mathcal{H}$ of kinematical states and it is natural to attempt to represent left-hand sides of quantum Einstein equations by well-defined operators on $\mathcal{H}$. Now, in interacting (low-dimensional) quantum field theories, there is a delicate relation between quantum kinematics and dynamics: unless the representation of the basic operator algebra is chosen appropriately, typically, the Hamiltonian fails to be well defined on the Hilbert space. For a complicated system such as general relativity, then, one would imagine that the problem of choosing the 'correct' kinematic representation would be extremely difficult (see, e.g., [85]). However, a major simplification arises from the striking uniqueness result discussed at the end of section 4: the requirement of general covariance picks out a unique representation of the algebra generated by holonomies and electric fluxes [55–58]. Therefore we have a *single arena* for background-independent theories of connections and a natural strategy for implementing dynamics provided, of course, this mathematically natural, kinematical algebra is also 'physically correct'. (This proviso exists also for the quantum field theories referred to above.) As we will summarize in this section, this strategy has led to well-defined candidates for quantum Einstein equations.

Recall from section 2 that because general relativity has no background fields, the theory is fully constrained in its phase space formulation. To pass to the quantum theory, one can use one of the two standard approaches: (i) find the reduced phase space of the theory representing 'true degrees of freedom' thereby eliminating the constraints classically and then construct a quantum version of the resulting unconstrained theory; or (ii) first construct quantum kinematics for the full phase space ignoring the constraints, then find quantum operators corresponding to constraints and finally solve quantum constraints to obtain the physical states. Loop quantum gravity follows the second avenue, which was initiated by Dirac[29]. This program has been carried out to completion in many simpler systems,

[29] Thus, in the canonical approach, the entire quantum dynamics is fully incorporated by solving quantum constraints. This may seem surprising because in the classical theory we have both the constraint and the evolution equations.

such as (2 + 1)-dimensional gravity [2, 5, 35] and a number of mini-superspaces in 3 + 1 dimensions [86], where one can explicitly see that the procedure incorporates all of quantum Einstein's equations. Readers who are not familiar with quantization of constrained systems should first familiarize themselves with the subject through simple examples (see, e.g., [87]). To adequately handle conceptual and technical intricacies encountered in general relativity, Dirac's original program has to be modified and extended suitably. We will use the resulting framework, called *refined algebraic quantization*. For further details, see, e.g., [88, 89, 141].

6.1. The Gauss constraint

Recall from section 2 that the Gauss constraint, $G_i \equiv \mathcal{D}_a P^a_i = 0$, generates internal $SU(2)$ rotations on the phase space of general relativity. More precisely, given an $su(2)$-valued function ξ on M, we can use it as a smearing field to obtain a phase space function

$$\mathcal{C}_{\rm G}(\xi) = -\int_M \mathrm{d}^3x\, P^a_i(x)\mathcal{D}_a\xi^i(x), \tag{6.1}$$

which generates infinitesimal canonical transformations $(A, P) \to (A - \mathcal{D}\xi, P + [\xi, P])$. Using the heuristic ansatz $P \to -\mathrm{i}\hbar\delta/\delta A$, it is straightforward to promote $\mathcal{C}_{\rm G}(\xi)$ to a well-defined operator on $\mathcal{H}$ [89]. For any $\Psi_\alpha \in \mathrm{Cyl}^{(1)}_\alpha$ we have

$$\hat{\mathcal{C}}_{\rm G}(\xi)\Psi_\alpha = \hbar \sum_v \sum_e \big(\xi^i(v) J_i^{(v,e)}\big)\Psi_\alpha \tag{6.2}$$

where the first sum extends over all vertices v of α and the second over all edges e meeting at v. Apart from the factor of $\hbar$, this action coincides with that of the generator of gauge transformations on $\mathcal{H}_\alpha$ discussed in section 4.2.2. This family of operators on $\mathcal{H}_\alpha$ is consistent and defines a self-adjoint operator on $\mathcal{H}$ which we will also denote by $\hat{\mathcal{C}}_{\rm G}(\xi)$. Finite gauge transformations are generated by the one-parameter unitary groups generated by these operators.

Physical states belong to the kernel $\mathcal{H}^{\rm G}_{\rm inv}$ of $\hat{\mathcal{C}}_{\rm G}(\xi)$ for all $\xi \in su(2)$. Because the action of $\hat{\mathcal{C}}_{\rm G}$ is familiar, the kernel is easy to find: in terms of the Hilbert space decompositions discussed in section 4.3.3,

$$\mathcal{H}^{\rm G}_{\rm inv} = \bigoplus_{\alpha,\mathbf{j}} \mathcal{H}'_{\alpha,\mathbf{j},\mathbf{l}=\mathbf{0}}.$$

Note that these states are automatically invariant under *generalized* gauge transformations in $\bar{\mathcal{G}}$ and can be regarded as functions on the reduced quantum configuration space $\bar{\mathcal{A}}/\bar{\mathcal{G}}$. $\mathcal{H}^{\rm G}_{\rm inv}$ is a subspace of $\mathcal{H}$ because zero is in the discrete part of the spectrum of the constraint operator $\hat{\mathcal{C}}_{\rm G}(\xi)$. In particular, $\mathcal{H}^{\rm G}_{\rm inv}$ inherits a Hilbert space structure from $\mathcal{H}$ and $\mathcal{H}^{\rm G}_{\rm inv} = L^2\big(\bar{\mathcal{A}}/\bar{\mathcal{G}}, \mathrm{d}\mu^{\rm G}_0\big)$, where $\mathrm{d}\mu^{\rm G}_0$ is the natural measure on $\bar{\mathcal{A}}/\bar{\mathcal{G}}$, the push-forward of $\mathrm{d}\mu_0$ under the natural projection map from $\bar{\mathcal{A}}$ to $\bar{\mathcal{A}}/\bar{\mathcal{G}}$. Every gauge invariant operator (such as areas $\hat{A}_S$ and volumes $\hat{V}_R$ in section 5) has a well-defined action on $\mathcal{H}^{\rm G}_{\rm inv}$.

The fact that the Gauss constraint could be imposed so easily and that the structure of $\mathcal{H}^{\rm G}_{\rm inv}$ is so simple hides the non-triviality of the procedure. For example if, in place of $\bar{\mathcal{A}}$, one uses one of the standard distribution spaces as the quantum configuration space, the imposition of the Gauss constraint and construction of the Hilbert space of physical states become complicated and it is not obvious that these difficulties can be surmounted.

However, because the evolution is generated by constraints in the Hamiltonian framework, in the quantum theory dynamics is encoded in the operator constraints. A simple example is provided by a free particle in Minkowski space, where the constraint $g^{ab}p_a p_b + m^2 = 0$ on the classical phase space becomes $\Box\phi - m^2\phi = 0$, governing dynamics in the quantum theory.

6.2. *The diffeomorphism constraint*

Let us now consider the diffeomorphism constraint. We will find that the imposition of this constraint is more complicated because of a key difference: while there is an infinite-dimensional subspace $\mathcal{H}^{\rm G}_{\rm inv}$ of $\mathcal{H}$ that is invariant under the $SU(2)$ gauge rotations, since diffeomorphisms move graphs, the only element of $\mathcal{H}$ left invariant by the action of all diffeomorphisms is the constant function on $\bar{\mathcal{A}}$! As a result, solutions to the quantum constraints lie not in the kinematical Hilbert space $\mathcal{H}$ but in a larger space, the dual Cyl* of Cyl. This is not unusual. Even in simple quantum mechanical systems, such as a particle in R^3 with a constraint $p_x = 0$, solutions to the constraint fail to have finite norm in the kinematic Hilbert space $L^2(R^3)$ and belong to a larger space, e.g., the space of distributions in R^3. In a similar fashion, we will be able to construct a systematic framework and obtain the general solutions to the diffeomorphism constraint.

6.2.1. Strategy. Recall from section 2.3.3 that each vector field N^a on M defines a constraint function $\mathcal{C}_{\rm Diff}(\vec{N})$ on the gravitational phase space:

$$\mathcal{C}_{\rm Diff}(\vec{N}) = \int_M {\rm d}^3x \big(N^a F^i_{ab} P^b_i - P^a \mathcal{D}_a\big(N^b A^i_b\big)\big). \tag{6.3}$$

Under infinitesimal canonical transformations generated by $\mathcal{C}_{\rm Diff}(\vec{N})$, we have $(A, P) \mapsto (A + \mathcal{L}_{\vec{N}} A, P + \mathcal{L}_{\vec{N}} P)$. In the mathematically precise literature on constrained systems, it is the *finite* gauge transformations generated by constraints that are of primary interest in the quantum theory. Therefore, in our case, it is appropriate to impose the diffeomorphism constraint by demanding that the physical states be left invariant under finite diffeomorphisms φ generated by $\vec{N}$. Since the measure ${\rm d}\mu_0$ on $\bar{\mathcal{A}}$ is diffeomorphism invariant, the induced action of φ on $\mathcal{H}$ is unitary. Thus, given the vector field N^a, we obtain a one-parameter family $\varphi(\lambda)$ of diffeomorphisms on M and a corresponding family $\hat{\varphi}(\lambda)$ of unitary operators on $\mathcal{H}$. But this family fails to be weakly continuous in λ because ${\rm Cyl}_\alpha$ is *orthogonal* to ${\rm Cyl}_{\varphi\cdot\alpha}$ if φ moves α. Hence, the infinitesimal generator of $\hat{\varphi}(\lambda)$ fails to exist. (For details, see appendix C in [89].) However, this creates no obstacle because, for the quantum implementation of the constraint, we can work directly with finite diffeomorphisms: physical states are to be invariant under the induced action $\hat{\varphi}$ of appropriate diffeomorphisms φ on M.

To solve the constraint, we will use the 'group averaging procedure', generally available for such constraints[30] (see, e.g., [88, 89]): physical states will be obtained by averaging elements of Cyl with respect to the induced action of the diffeomorphism group. It is intuitively obvious that the result of group averaging will be diffeomorphism invariant. However, although one begins with states in Cyl, the result naturally belongs to Cyl*, the algebraic dual of Cyl.[31] In finite-dimensional constrained systems, one generally uses a triplet, $\mathcal{S} \subset L^2(R^n) = H_{\rm kin} \subset \mathcal{S}^\star$, where $\mathcal{S}$ is typically the space of smooth functions with rapid decay at infinity, and $\mathcal{S}^\star$, the space of distributions. The solutions to constraints are obtained by averaging elements of $\mathcal{S}$ with respect to the group generated by constraints and they typically belong to $\mathcal{S}^\star$ rather than to the kinematical Hilbert space $H_{\rm kin}$ [88, 141]. In the present case, we have a completely analogous situation and now the triplet is Cyl $\subset \mathcal{H} \subset$ Cyl*.

[30] The quantum Gauss constraint can be rigorously implemented also through group averaging over $\bar{\mathcal{G}}$. The final result is the same as that obtained in section 6.1. For pedagogical purposes, in section 6.1 we adopted a procedure which is closer to that followed for the scalar constraint in section 6.3.

[31] In the end, one would have to introduce a suitable topology on Cyl which is finer than the Hilbert space topology and let Cyl* be the topological dual of Cyl. The program is yet to reach this degree of sophistication and, for the moment, the much bigger algebraic dual is used.

Finally, there is an important technical subtlety coming from the fact that graphs α are required to have closed-piecewise analytic edges. The classical phase space, on the other hand, consists of smooth (i.e., C^n) fields (A, P). Smooth diffeomorphisms φ correspond to finite canonical transformations generated by the constraint (6.3) and have a well-defined action on the phase space. It is just that the action does not extend to our full algebra of 'elementary variables' since their definition involves closed-piecewise analytic edges and surfaces. A natural strategy to impose the diffeomorphism constraint, therefore, is to enlarge the framework and allow smooth edges and surfaces. This is possible [47, 63, 64, 92] but then the technical discussion becomes much more complicated because, e.g., two smooth curves can intersect one another at an infinite number of points. Here we will adopt an 'in-between' approach and use the subgroup Diff of all C^n diffeomorphisms of M, introduced in section 4.3.5, which has a well-defined action on our elementary variables and the Hilbert space $\mathcal{H}$. From a physical perspective, this is more appropriate than averaging with respect to just the analytical diffeomorphisms and from a mathematical perspective it enables us to bypass the complications associated with non-analytical edges and surfaces.

6.2.2. Physical states. Our task now is to construct the general solution to the diffeomorphism constraint. For this, we will use the spin network decomposition (4.46): $\mathcal{H} = \oplus_\alpha \mathcal{H}'_\alpha$. Let us begin by introducing some notation. Given a graph α, denote by Diff_α the subgroup of Diff which maps α to itself and by TDiff_α its subgroup which has trivial action on α, i.e., which preserves every edge of α and its orientation. The induced action, $\widehat{\mathrm{TDiff}}_\alpha$, is *trivial* on Cyl_α. Next, let Diff_α be the group of all the diffeomorphisms that preserve α. Then, the quotient

$$\mathrm{GS}_\alpha = \mathrm{Diff}_\alpha / \mathrm{TDiff}_\alpha, \tag{6.4}$$

is the group of *graph symmetries* of α. It is a finite group and it has a non-trivial induced action $\widehat{\mathrm{GS}}_\alpha$ on Cyl_α. In the group averaging procedure, consistency requires that one must divide by the 'volume' of the orbits of these groups [89].

To construct the general solution to the diffeomorphism constraint, we proceed in two steps. First, given any $\Psi_\alpha \in \mathcal{H}'_\alpha$, we average it using *only the group of graph symmetries* and obtain a projection map $\hat{P}_{\mathrm{diff},\alpha}$ from $\mathcal{H}'_\alpha$ to its subspace which is invariant under $\widehat{\mathrm{GS}}_\alpha$,

$$\hat{P}_{\mathrm{diff},\alpha}\Psi_\alpha := \frac{1}{N_\alpha} \sum_{\varphi \in GS_\alpha} \varphi \star \Psi_\alpha, \tag{6.5}$$

where N_α is the number of the elements of GS_α (the volume of the orbit of GS) and $\varphi \star \Psi_a$ denotes the pullback of Ψ_α under φ. The map extends naturally to a projection $\hat{P}_{\mathrm{diff}}$ from $\mathcal{H} = L^2(\bar{\mathcal{A}}, \mathrm{d}\mu_0)$ to its subspace which is invariant under $\widehat{\mathrm{GS}}_\alpha$ for all α.

In the second step, we wish to average with respect to the remaining diffeomorphisms *which move the graph* α. This is a very large group and the result of averaging now belongs to $\mathrm{Cyl}^\star$ rather than $\mathcal{H}$. Thus, with each $\Psi_\alpha \in \mathcal{H}'_\alpha$, we now associate an element $(\eta(\Psi_\alpha)| \in \mathrm{Cyl}^\star$, defined by its (linear) action on arbitrary cylindrical functions $|\Phi_\beta\rangle \in \mathrm{Cyl}$,

$$(\eta(\Psi_\alpha)|\Phi_\beta\rangle = \sum_{\varphi \in \mathrm{Diff}/\mathrm{Diff}_\alpha} \langle \varphi \star \hat{P}_{\mathrm{diff},\alpha}\Psi_\alpha, \Phi_\beta\rangle, \tag{6.6}$$

where the bracket on the right-hand side denotes the inner product between elements of $\mathcal{H}$. Although $\varphi \in \mathrm{Diff}/\mathrm{Diff}_\alpha$ contains an infinite number of elements φ, for any given β only a finite number of terms are non-zero, whence $\eta(\Psi_\alpha)$ is well defined. However, there is no vector η in $\mathcal{H}$ such that $\langle \eta, \Phi_\beta\rangle$ equals the right-hand side of (6.6) for all $\Phi_\beta \in \mathrm{Cyl}$. Thus, $(\eta(\Psi_\alpha)|$ is a 'genuine distribution' on $\bar{\mathcal{A}}$ rather than a function. Because of the diffeomorphism invariance of the scalar product on $\mathcal{H}$, $\eta(\Psi_\alpha)$ is invariant under the action of $\mathrm{Diff}(M)$,

$$(\eta(\Psi_\alpha)|\varphi \star \Phi_\beta\rangle = (\eta(\Psi_\alpha)|\Phi_\beta\rangle \tag{6.7}$$

for all $\varphi \in \mathrm{Diff}(M)$. We will denote the space of these solutions to the diffeomorphism constraint by $\mathrm{Cyl}^\star_{\mathrm{diff}}$. Finally, since Ψ_α was an arbitrary element of Cyl, we have constructed a map:

$$\eta : \mathrm{Cyl} \to \mathrm{Cyl}^\star_{\mathrm{diff}}. \tag{6.8}$$

Thus, *every* element of Cyl gives rise, upon group averaging, to a solution to the diffeomorphism constraint. In this sense, we have obtained the general solution to the diffeomorphism constraint. The map η is the analogue of the projection from $\mathcal{H}$ to its gauge invariant subspace $\mathcal{H}^{\mathrm{G}}_{\mathrm{inv}}$ in the case of the Gauss constraint. However, because of the differences between the two constraints discussed above, η is *not* a projection since it maps Cyl on to a *different* space $\mathrm{Cyl}^\star_{\mathrm{diff}}$. Nonetheless, the group averaging procedure naturally endows the solution space with a Hermitian inner product,

$$(\eta(\Psi)|\eta(\Phi)) := (\eta(\Psi)|\Phi\rangle, \tag{6.9}$$

since one can show that the right-hand side is independent of the specific choice of Ψ and Φ made in the averaging [88, 141]; for subtleties, see [89, 141]. We will denote by $\mathcal{H}_{\mathrm{diff}}$ the Cauchy completion of $\mathrm{Cyl}^\star_{\mathrm{diff}}$. Finally, we can obtain the *general solution to both the Gauss and the diffeomorphism constraints* by simply restricting the initial $\Psi \in \mathrm{Cyl}$ to be gauge invariant, i.e., to belong to $\mathrm{Cyl} \cap \mathcal{H}^{\mathrm{G}}_{\mathrm{inv}}$. We will denote this space of solutions by $\mathrm{Cyl}^\star_{\mathrm{inv}}$:

$$\mathrm{Cyl}^\star_{\mathrm{inv}} = \eta\big(\mathrm{Cyl} \cap \mathcal{H}^{\mathrm{G}}_{\mathrm{inv}}\big). \tag{6.10}$$

What is the situation with respect to operators? Note first that there do exist non-trivial (gauge and) diffeomorphism invariant operators on Cyl; an example is the total volume operator $\hat{V}_M$. Let $\mathcal{O}$ be such an operator. Its dual, $\mathcal{O}^\star$, is well defined in $\mathrm{Cyl}^\star_{\mathrm{inv}}$:

$$(\mathcal{O}^\star\eta(\Psi)|\Phi\rangle := (\eta(\Psi)|\mathcal{O}\Phi\rangle. \tag{6.11}$$

(Furthermore, one can show that $\mathcal{O}^\star$ preserves the image of $\mathcal{H}'_\alpha$ for every α.) The operator $\mathcal{O}^\star$ is self-adjoint with respect to the natural scalar product (6.9) on $\mathcal{H}_{\mathrm{diff}}$ if and only if $\mathcal{O}$ is self-adjoint in $\mathcal{H}$. This property shows that the scalar product on $\mathcal{H}$ is not only mathematically natural but also 'physically correct'.

Let us summarize. The basic idea of the procedure used to solve the diffeomorphism constraint is rather simple: one averages the kinematical states with the action of the diffeomorphism group to obtain physical states. But the fact that this procedure can be implemented in detail is quite non-trivial. For example, a mathematically precise implementation still eludes the geometrodynamics program. Furthermore, even the final answer contains certain subtleties. We will conclude by pointing them out.

Remarks

(i) Note that while $(\eta(\Psi)|$ is a solution to the diffeomorphism constraint for any $\Psi \in \mathrm{Cyl}$, it is *not* true that there is a 1–1 correspondence between elements of Cyl and solutions to the diffeomorphism constraint. This is because the map η has a non-trivial kernel. In particular, the projection map $\hat{P}_{\mathrm{diff}}$ itself has a non-trivial kernel which, by (6.6) is also in the kernel of η. (In addition, elements of Cyl of the form $a_0\Psi_\alpha + a_1\varphi_1 \star \Psi_\alpha + \cdots + a_n\varphi_n \star \Psi_\alpha$, with $a_0 + \cdots + a_n = 0$ are also in the kernel of η.) Therefore statements such as—'solutions to the diffeomorphism constraint are diffeomorphism classes of spin network states'—that one often finds in the literature are only heuristic.

(ii) One also finds claims to the effect that the diffeomorphism constraint can be imposed simply by replacing spin networks embedded in a manifold M by abstract, non-embedded spin networks. Within the systematic approach summarized in this section, these claims

are simply incorrect (for a detailed discussion in the context of $2+1$ gravity on a lattice, see, e.g., [93]). In particular a graph, one of whose edges is knotted, cannot be mapped by a diffeomorphism to that in which all edges are unknotted, whence the mapping η sends spin network states associated with the two graphs to *distinct* solutions to the diffeomorphism constraint. As abstract, non-embedded graphs, on the other hand, they can be equivalent and define the same spin network functions. One can imagine a new approach in which one simply declares that the diffeomorphism constraint is to be incorporated by replacing embedded spin networks by abstract ones. But since the original diffeomorphism constraint acts on the basic canonical variables (A, E) on M and the action can be transferred to graphs only if they are embedded, it would be difficult to justify such an approach from first principles.

(iii) Note that there are continuous families of 4-valent or higher valent graphs which cannot be mapped to one another by C^n diffeomorphisms with $n > 0$. Consequently, states in $\mathcal{H}_{\rm diff}$ based on two of these graphs are mutually orthogonal. Thus, even though we have 'factored out' by a very large group Diff, the Hilbert space $\mathcal{H}_{\rm diff}$ is still non-separable. However, if we were to let $n = 0$, i.e., consider homeomorphisms of M which preserve the family of graphs under consideration, then these 'problematic' continuous families of graphs would all be identified in the group averaging procedure and the Hilbert space of solutions to the diffeomorphism constraint *would be* separable. However, since the classical constraints do not generate homeomorphisms, and furthermore homeomorphisms do not even have a well-defined action on the phase space, it is difficult to 'justify' this enlargement of Diff from direct physical considerations.

(iv) Note that $\mathrm{Cyl}^\star_{\rm diff}$ is a proper subset of the space $\mathrm{Cyl}^\star_{\rm Diff}$ of *all* elements of $\mathrm{Cyl}^\star$ invariant under Diff. However, every $(\Psi| \in \mathrm{Cyl}^\star_{\rm Diff}$ can be uniquely decomposed as

$$(\Psi| = \sum_{[\alpha]} (\Psi_{[\alpha]}|, \qquad \text{with} \quad (\Psi_{[\alpha]}| \in \eta(\mathrm{Cyl}_\alpha), \tag{6.12}$$

where $[\alpha]$ runs through the diffeomorphism classes of graphs. The sum on the right-hand side is uncountable but the result is a well-defined element of $\mathrm{Cyl}^\star$ because, in its action on any cylindrical function, only a finite number of the terms fails to vanish.

6.3. The scalar constraint

The canonical transformations generated by the Gauss and the diffeomorphism constraints are *kinematical* gauge symmetries of the classical theory in the sense that, in the spacetime picture, they operate at a 'fixed time'. The crux of quantum dynamics lies in the scalar constraint. One can imagine implementing it in the quantum theory also by a group averaging procedure. However, this strategy is difficult to adopt because the *finite* canonical transformations generated by this constraint are not well understood even at the classical level. Therefore, one follows the procedure used for the Gauss constraint: construct a quantum operator corresponding to the classical, smeared constraint function and then seek its kernel. Because the form of this constraint is so intricate, its implementation is still rather far from being as clean and complete as that of the other two constraints. In particular, genuine ambiguities exist in the regularization procedure and distinct avenues have been pursued [95–102]. What is non-trivial at this stage is the *existence* of well-defined strategies. Whether any of them is fully viable from a physical perspective is still an open issue. In this summary, we will essentially follow the most developed of these approaches, introduced by Thiemann [95–97]. However, to bring out quantization ambiguities we have generalized the method, emphasizing points at which there is freedom to modify the original procedure and still arrive

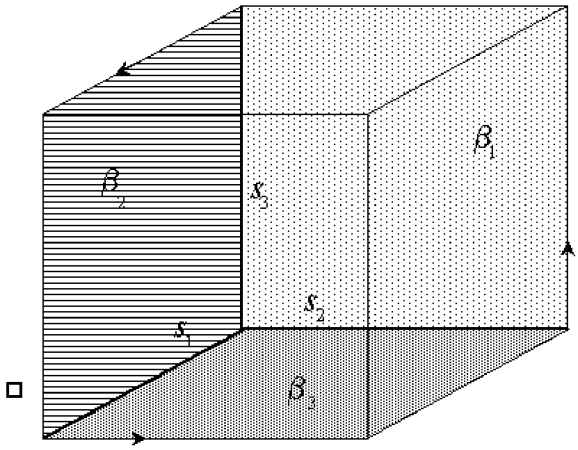

Figure 2. An elementary cell $\square$ in a cubic partition. s_1, s_2, s_3 are the edges of the cell and $\beta_1, \beta_2, \beta_3$ the three oriented loops which are boundaries of faces orthogonal to these edges.

at a well-defined constraint operator. Our emphasis is more on clarifying the underlying conceptual structure than on providing efficient calculational tools.

6.3.1. Regulated classical expression. As with area and volume operators, our first task is to re-express the classical expression of the scalar constraint as a Riemann sum involving only those phase space functions which have direct quantum analogues. Recall from section 2 that in terms of real connection variables, the scalar constraint (2.31) smeared with a lapse N can be written as a sum of two terms,

$$\mathcal{C}(N) = \Big(\frac{\gamma}{4k}\Big)^{\frac{1}{2}} \int_M \mathrm{d}^3x N \frac{P_i^a P_j^b}{\sqrt{\det P}} \big[\epsilon^{ij}{}_k F_{ab}^k + 2(\sigma - \gamma^2) K_{[a}^i K_{b]}^j\big] \tag{6.13}$$

where, as before, $k = 8\pi G$ and we have used the relation $\det q = (\kappa\gamma)^3 \det P$. Had we worked in the $+,+,+,+$ signature and in the half-flat sector $\gamma^2 = \sigma$, the second term would have been zero. Thus, the first term has the interpretation of the scalar constraint of Euclidean general relativity. Therefore, the full Lorentzian constraint can be written as

$$\mathcal{C}(N) = \sqrt{\gamma}\mathcal{C}^{\mathrm{Eucl}}(N) - 2(1+\gamma^2)\mathcal{T}(N) \tag{6.14}$$

where we have used $\sigma = -1$ corresponding to the Lorentzian signature.

Let us begin by exploring the first term. In comparison with geometric operators discussed in section 5, we now have three sets of complications. First, the expression of $\mathcal{C}^{\mathrm{Eucl}}(N)$ involves not only triads P_i^a but also curvature F_{ab}^i of the connection A_a^i. However, following the standard procedure in gauge theories, it is straightforward to express curvature in terms of holonomies which can be directly promoted to operators. The second complication arises from the fact that the expression of $\mathcal{T}(N)$ involves extrinsic curvature terms. Fortunately, we will see that these can be expressed using the Poisson bracket between $\mathcal{C}^{\mathrm{Eucl}}$ and the total volume, both of which have well-defined operator analogues. The final complication is the presence of the volume element $\sqrt{\det P}$ in the denominator. At first, this seems to be a fatal drawback. A key insight of Thiemann [95, 96] was that this is not the case (see also [100] for a further discussion). For, the combination

$$e_a^i := \frac{\sqrt{k\gamma}}{2} \eta_{abc} \epsilon^{ijk} \frac{P_j^b P_k^c}{\sqrt{\det P}}, \tag{6.15}$$

representing the co-triad e^i_a can be expressed as a manageable Poisson bracket:

$$e^i_a(x) = \frac{2}{k\gamma}\{A^i_a(x), V\}. \tag{6.16}$$

Using this fact, the Euclidean scalar constraint part $\mathcal{C}^{\mathrm{Eucl}}(N)$ is written as

$$\mathcal{C}^{\mathrm{Eucl}}(N) = -\frac{2}{k^2\gamma^{\frac{3}{2}}}\int_M \mathrm{d}^3x\, N(x)\eta^{abc}\,\mathrm{Tr}(F_{ab}(x)\{A_c(x), V\}). \tag{6.17}$$

We will see that this expression is well suited for quantization.

The second term $\mathcal{T}(N)$ in expression (6.14) of the constraint is given by

$$\mathcal{T}(N) = \frac{\sqrt{\gamma}}{2\sqrt{k}}\int_M \mathrm{d}^3x\, N\left(K^i_{[a}K^j_{b]}\frac{P^a_i P^b_j}{\sqrt{\det P}}\right). \tag{6.18}$$

To cast this term in the desired form, we first note that K^i_a can be expressed as a Poisson bracket,

$$K^i_a = \frac{1}{k\gamma}\{A^i_a, \bar{K}\} \tag{6.19}$$

where $\bar{K}$ is the integral of the trace of the extrinsic curvature,

$$\bar{K} = k\gamma\int_M \mathrm{d}^3x\, K^i_a P^a_i. \tag{6.20}$$

Now $\bar{K}$ itself can be expressed as a Poisson bracket,

$$\bar{K} = \gamma^{-\frac{3}{2}}\{\mathcal{C}^{\mathrm{Eucl}}(1), V\}. \tag{6.21}$$

Hence $\mathcal{T}(N)$ can be expressed as

$$\mathcal{T}(N) = -\frac{2}{k^4\gamma^3}\int_M \mathrm{d}^3x N(x)\eta^{abc}\,\mathrm{Tr}(\{A_a(x), \bar{K}\}\{A_b(x), \bar{K}\}\{A_c(x), V\}). \tag{6.22}$$

Thus, to express the constraint in terms of variables adapted to quantum theory, it only remains to re-express the curvature and connection terms appropriately. Now, if s is a line segment of coordinate length ε and if a loop β is the boundary of a coordinate plane P of area ε^2, we have

$$\begin{aligned}\left\{\int_s A, V\right\} &= -[\bar{A}(s)]^{-1}\{\bar{A}(s), V\} + o(\varepsilon)\\ \left\{\int_s A, \bar{K}\right\} &= -[\bar{A}(s)]^{-1}\{\bar{A}(s), \bar{K}\} + o(\varepsilon)\end{aligned} \tag{6.23}$$

and

$$\int_P F = \frac{1}{2}(\bar{A}(\beta^{-1}) - \bar{A}(\beta)) + o(\varepsilon^2). \tag{6.24}$$

These formulae provide a concrete strategy to replace the connection and curvature terms in terms of holonomies. For example, if M is topologically R^3, it is simplest to introduce a cubic partition where the coordinate length of edges of elementary cells is ε. Denote by s_1, s_2, s_3, the edges of an elementary cell $\square$ based at a vertex $v_\square$ and by β_1, β_2 and β_3 the three oriented loops based at $v_\square$ which are boundaries of faces orthogonal to these edges (see figure 2). Then, the $\sum_\square \mathcal{C}^{\mathrm{Eucl}}_\square(N)$, where

$$\mathcal{C}^{\mathrm{Eucl}}_\square(N) = -\frac{2N(v_\square)}{k^2\gamma^{\frac{3}{2}}}\sum_I \mathrm{Tr}\big(\big(\bar{A}(\beta_I) - \bar{A}\big(\beta_I^{-1}\big)\big)\bar{A}(s_I)^{-1}\{\bar{A}(s_I), V\}\big), \tag{6.25}$$

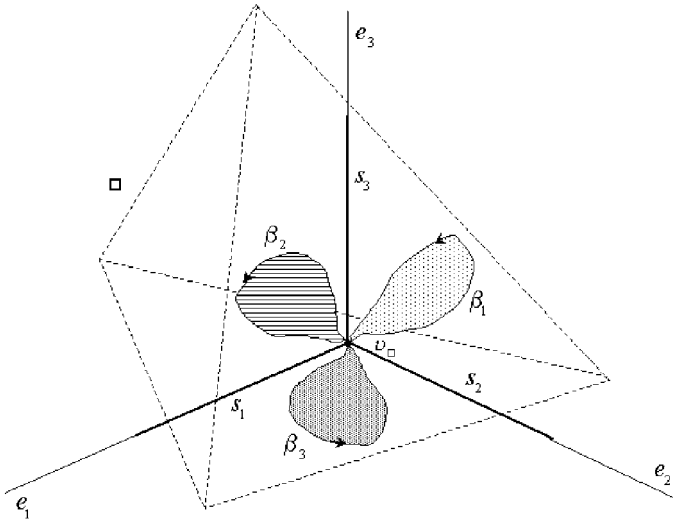

Figure 3. An elementary cell □ in a general partition. Segments s_i now lie along the edges of the given graph which has a vertex $v_\square$ in the interior of □. Each of the loops β_i originates and ends at $v_\square$ and lies in a coordinate plane spanned by two edges.

is a Riemann sum which converges to $\mathcal{C}^{\mathrm{Eucl}}(N)$ as the cell size tends to zero (and the number of cells tends to infinity). Similarly, the sum $\sum_\square \mathcal{T}_\square(N)$

$$\mathcal{T}_\square(N) = \frac{2N(v_\square)}{k^4\gamma^3}\epsilon^{IJL}\,\mathrm{Tr}\big(\bar{A}\big(s_I^{-1}\big)\{\bar{A}(s_I),\bar{K}\}\bar{A}\big(s_J{}^{-1}\big)\{\bar{A}(s_J),\bar{K}\}\bar{A}\big(s_L{}^{-1}\big)\{\bar{A}(s_L),V\}\big) \tag{6.26}$$

is a Riemann sum which converges to $\mathcal{T}(N)$ as the cell size tends to zero. These Riemann sums can therefore be regarded as providing a 'regularization' of the classical constraint. As in the discussion of the geometric operators of section 5, the idea is to first replace classical quantities in the 'regularized expression' by their quantum counterparts and *then* remove the regulator. *A remarkable feature of this regularization, first pointed out by Rovelli and Smolin* [94], *is that the regulating parameter* ε *has disappeared from the expression*. Hence it is not necessary to multiply the constraint by a suitable power of ε before removing the regulator; no renormalization is involved.

The cubic partition is the simplest example of a more general classical regularization. The available freedom can be summarized as follows. To every value $\varepsilon \in [0, \varepsilon_0]$, assign a partition of Σ into cells □ of possibly arbitrary shape (see figure 3). In every cell □ of the partition we define edges s_J, $J = 1, \ldots, n_s$ and loops $\beta_i, i = 1, \ldots, n_\beta$, where n_s, n_β may be different for different cells. Finally, fix an arbitrarily chosen representation ρ of $SU(2)$. This entire structure will be denoted by R_ε and called a *permissible classical regulator* if the following property holds,

$$\lim_{\varepsilon\to 0}\mathcal{C}^{\mathrm{Eucl}}_{R_\varepsilon}(A,E) = \mathcal{C}^{\mathrm{Eucl}}(A,E) \quad \text{and} \quad \lim_{\varepsilon\to 0}\mathcal{T}_{R_\varepsilon}(A,E) = \mathcal{T}(A,E), \tag{6.27}$$

where

$$\mathcal{C}^{\mathrm{Eucl}}_{R_\varepsilon} = \sum_\square \mathcal{C}^{\mathrm{Eucl}}_{R_\varepsilon\square}, \tag{6.28}$$

$$\mathcal{C}^{\mathrm{Eucl}}_{R_\varepsilon} = \frac{N(v_\square)}{k^2\gamma^{\frac{3}{2}}}\sum_{iJ} C^{iJ}\mathrm{Tr}\big(\big(\rho[\bar{A}(\beta_i)] - \rho\big[\bar{A}\big(\beta_i^{-1}\big)\big]\big)\rho\big[\bar{A}\big(s_J{}^{-1}\big)\big]\{\rho[\bar{A}(s_J)],V\}\big), \tag{6.29}$$

$$T_{R_\varepsilon} = \sum_{\square} T_{R_\varepsilon}\square \tag{6.30}$$

$$T_{R_\varepsilon}\square = \frac{N(v_\square)}{k^4\gamma^3}\sum_{I,J,K} T^{IJK}\mathrm{Tr}(\rho[\bar{A}(s_I^{-1})]\{\rho[\bar{A}(s_I)],\bar{K}\}\rho[\bar{A}(s_J{}^{-1})]\{\rho[\bar{A}(s_J)],\bar{K}\} \times \rho[\bar{A}(s_K{}^{-1})]\{\rho[\bar{A}(s_K)],V\}), \tag{6.31}$$

and C^{iJ}, T^{IJK} are fixed constants, *independent* of the scale parameter ε. A large family of the classical regulators can be constructed by modifications of the cubic example, changing the shape of the cells, loops and edges, and their relative positions suitably.

6.3.2. The quantum scalar constraint. Our task is to first promote the regulated classical constraint to a quantum operator and then remove the regulator. In the detailed implementation of this procedure, one encounters three non-trivial issues.

As in the case of geometric operators, the first step is rather straightforward because the regulated expressions involve only those phase functions which have direct quantum analogues. However, while the 'obvious' quantum operator would be well defined on states which are cylindrical with respect to any one graph, in the end we have to ensure that the resulting family of operators is consistent. This is the first non-trivial issue. The simplest way to address it is to use the decomposition $\mathcal{H} = \oplus_\alpha \mathcal{H}'_\alpha$ of the Hilbert space, introduced in section 4.3.3, and define the quantum constraint on each $\mathcal{H}'_\alpha$ separately. Because of the orthogonality of any two $\mathcal{H}'_\alpha$, the resulting family of operators would then be automatically consistent.

Let us begin with $\mathcal{C}^{\mathrm{Eucl}}_{R_\varepsilon}$. Fix a subspace $\mathcal{H}'_\alpha$ of $\mathcal{H}$. The quantum operator can be obtained simply by promoting the holonomies and volume functions to operators and replacing the Poisson brackets by $1/i\hbar$-times commutators. Thus, for any given graph α and $\varepsilon > 0$,

$$\hat{\mathcal{C}}^{\mathrm{Eucl}}_{R_{\varepsilon,\alpha}}(N) := \sum_{\square} \hat{\mathcal{C}}^{E}_{\square}(N) \tag{6.32}$$

with

$$\hat{\mathcal{C}}^{\mathrm{Eucl}}_{\square}(N) := -\frac{\mathrm{i}N(v_\square)}{k^2\gamma^{\frac{3}{2}}\hbar}\sum_{iJ} C^{iJ}\,\mathrm{Tr}\big(\big(\rho[\bar{A}(\beta_i)] - \rho\big[\bar{A}\big(\beta_i^{-1}\big)\big]\big)\rho\big[\bar{A}\big(s_J{}^{-1}\big)\big][\rho[\bar{A}(s_J)],\hat{V}]\big), \tag{6.33}$$

is a densely defined operator on $\mathcal{H}'_\alpha$ with domain $\mathcal{D}_\alpha = \mathcal{H}'_\alpha \cap \mathrm{Cyl}_\alpha$ for any classical regulator R_ε. We now encounter the second non-trivial issue: we have to ensure that the final operator is diffeomorphism covariant. To address it, we need to use regulators which are not fixed but transform covariantly as we move from a graph α to any of its images under diffeomorphisms. Therefore, we will restrict our regulators appropriately.

A *diffeomorphism covariant quantum regulator* $R_{\varepsilon,\alpha}$ is a family of permissible classical regulators, one for each choice of the graph α, satisfying the following properties:

(a) the partition is sufficiently refined in the sense that every vertex v of α is contained in exactly one cell of $R_{\varepsilon,\alpha}$; and
(b) if (α, v) is diffeomorphic to (α', v') then, for every ε and ε', the quintuple $(\alpha, v, \square, (s_I), (\beta_J))$ is diffeomorphic to the quintuple $(\alpha', v', \square', (s'_I), (\beta'_J))$ where $\square$ and $\square'$ are the cells of R_ε and R'_ε respectively, containing v and v' respectively[32].

[32] We need the restriction only on cells which contain vertices because the properties of the volume operator imply that the action of $\hat{\mathcal{C}}^{\mathrm{Eucl}}_{\square}$ is non-trivial only if one of the segments s_I of the regulator intersects a vertex of α.

Such diffeomorphism compatible quantum regulators exist; an explicit example is given in [96]. Given such an $R_{\varepsilon,\alpha}$, for every value of ε, the operators $\hat{\mathcal{C}}^{\rm Eucl}_{R_{\varepsilon,\alpha}}(N)$ are densely defined on $\mathcal{H}'_\alpha$ with a common domain $\mathcal{D}_\alpha$. This family of operators determines a densely defined operator $\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}$ on the full Hilbert space $\mathcal{H}$ with domain Cyl, independent of the value of ε. Furthermore, for any value of ε, this domain is mapped to itself by the operator $\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}$.

Thus, it only remains to remove the regulator. Here we encounter the third non-trivial issue. Typically $\hat{\mathcal{C}}^{\rm Eucl}_{R_{\varepsilon,\alpha}}|\Psi_\alpha\rangle$ is orthogonal to $\hat{\mathcal{C}}^{\rm Eucl}_{R_{\varepsilon',\alpha}}|\Psi_\alpha\rangle$ if $\varepsilon \neq \varepsilon'$, whence the operator does not converge (even in the weak topology) on $\mathcal{H}$. This is a rather general problem associated with the topology of $\mathcal{H}$; we also encountered it while defining the operator analogue of the diffeomorphism constraint $\mathcal{C}_{\rm Diff}(\vec{N})$. Recall, however, that solutions to the diffeomorphism constraint also fail to lie in the kinematical Hilbert space $\mathcal{H}$; they belong to Cyl*, the algebraic dual of Cyl. Therefore, for the consistency of the whole picture, what we need is the action of the scalar constraint only on a sufficiently large subspace of Cyl* and not on $\mathcal{H}$. *And this action is well defined and non-trivial.* More precisely, for each $(\Psi| \in {\rm Cyl}^\star$, the action of the regulated constraint operator is naturally given by

$$\big[(\Psi|\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}\big]|\Phi\rangle := (\Psi|\big[\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}|\Phi\rangle\big] \tag{6.34}$$

for all $|\Phi\rangle \in {\rm Cyl}^\star$. We can now remove the regulator in the obvious fashion. Define

$$\hat{\mathcal{C}}^{\rm Eucl}(N) = \lim_{\varepsilon\to 0} \hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}(N) \tag{6.35}$$

via

$$[(\Psi|\hat{\mathcal{C}}^{\rm Eucl}(N)]|\Phi\rangle = \lim_{\varepsilon\to 0}(\Psi|\big[\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}(N)|\Phi\rangle\big]. \tag{6.36}$$

Note that the limit has to exist only pointwise, i.e., for each $|\Phi\rangle \in {\rm Cyl}$ separately. As a consequence the domain of the operator, the set of $(\Psi|$ in Cyl* for which the limit exists, is quite large. In particular, as discussed below, it includes a large class of solutions to the diffeomorphism constraints. The term $\mathcal{T}(N)$ in the scalar constraint can be handled in a completely parallel fashion. Specifically, we can first define the operator $\hat{\bar{K}}$ through

$$\hat{\bar{K}} := \frac{\rm i}{\hbar\gamma^{\frac{3}{2}}}[\hat{V}, \hat{\mathcal{C}}^E(1)] \tag{6.37}$$

and use $\hat{\bar{K}}$ and the quantum regulator $R_{\varepsilon,\alpha}$ to define the regulated operator $\hat{\mathcal{T}}_\square(N)$,

$$(\Psi|\hat{\mathcal{T}}_\square(N) = \frac{{\rm i}N(v_\square)}{k^4\gamma^3\hbar^3} T^{IJK}(\Psi|\,{\rm Tr}\big(\bar{A}\big(s_I^{-1}\big)[\bar{A}_{s_I}, \hat{\bar{K}}]\bar{A}\big(s_J{}^{-1}\big)[\bar{A}(s_J), \hat{\bar{K}}]\bar{A}\big(s_K{}^{-1}\big)[\bar{A}(s_K), \hat{V}]\big), \tag{6.38}$$

on the domain Cyl*. Collecting these definitions, we now have a regulated scalar constraint operator,

$$(\Psi|\hat{\mathcal{C}}_{R_\varepsilon}(N) := (\Psi|\left(\sqrt{\gamma}\hat{\mathcal{C}}^{\rm Eucl}_{R_\varepsilon}(N) - 2(1+\gamma^2)\sum_\square \hat{T}_\square(N)\right) \tag{6.39}$$

for all $(\Psi| \in {\rm Cyl}^\star$. Again, we can remove the regulator by taking the limit as $\varepsilon \to 0$ as in (6.36). By construction, the action of this operator is diffeomorphism covariant. Thus, each diffeomorphism covariant quantum regulator defines a scalar constraint operator. Since there is a great deal of freedom in choosing these regulators, there is considerable quantization ambiguity. Nonetheless, all these constructions exhibit some very non-trivial properties. We will conclude this section by providing two illustrations.

First, as mentioned in the beginning of this section, it is significant that well-defined prescriptions exist to give precise meaning to quantum Einstein equations in a background-independent setting. In geometrodynamics, for example, the Wheeler–DeWitt equation still

remains only formal. Secondly, these constructions match surprisingly well with the solutions of the diffeomorphism constraint. To see this note first that, irrespective of the choice of the diffeomorphism covariant quantum regulator, *up to diffeomorphisms*, the operator $\hat{\mathcal{C}}_{R_\varepsilon}(N)$ is independent of ε: for every ε, ε' and $|\Psi\rangle \in \mathrm{Cyl}$, there is a diffeomorphism φ such that

$$\hat{\mathcal{C}}_{R_{\varepsilon'}}(N)|\Psi\rangle = U_\varphi \hat{\mathcal{C}}_{R_{\varepsilon'}}(N)|\Psi\rangle, \tag{6.40}$$

for every N. Next, suppose that $(\Psi| \in \mathrm{Cyl}^\star$ is diffeomorphism invariant. Then, for every lapse function N, the result $(\Psi|\big(\hat{\mathcal{C}}_{R_\varepsilon}(N)\big)$ is in fact *independent* of ε, and so is the expression under the limit on the right-hand side of (6.39). Hence the regulator can be removed trivially [94]. Thus because of the form of the regulated operators, diffeomorphism invariant states in $\mathrm{Cyl}^\star$ (constructed by group averaging elements of $\mathcal{H}'_\alpha \cap \mathrm{Cyl}_\alpha$) are automatically in the domain of the scalar constraint operator. This tight matching between the way in which the two constraints are handled is quite non-trivial.

Remark. In the original construction by Thiemann [96], the Hamiltonian constraint operator was defined on $\mathrm{Cyl}^\star_{\mathrm{diff}}$. Now, as we saw in section 2, the Poisson bracket of any two scalar constraints is given by a diffeomorphism constraint in the classical theory. Therefore, on diffeomorphism invariant states, one would expect the quantum scalar constraint operators to commute. Irrespective of the choice of the regulator R_ε, they do. To obtain a more stringent test, the domain of the Thiemann operator was extended slightly in [100, 101]. The extended domain, called the 'habitat' [101], also includes certain elements of $\mathrm{Cyl}^\star$ which are not diffeomorphism invariant. Nonetheless, it turned out that the commutator between scalar constraints continues to vanish on the habitat. This may seem alarming at first. However, it turns out that the quantum operator corresponding to the classical Poisson bracket also annihilates every state in the habitat [100]. Thus, there is no inconsistency; the habitat just turned out to be too small to provide a non-trivial viability criterion of this quantization procedure. The domain of the operator introduced in this section includes the habitat and the same result continues to hold. More importantly, it is likely that this domain is significantly larger and may contain semi-classical states. If this turns out to be the case, stronger viability criteria to test this quantization procedure will become available. In particular, in addition to the relation to the classical Poisson algebra of constraints, one may be able to analyse the relation between the classical evolution and the action of the constraint on semi-classical states.

6.3.3. Solutions to the scalar constraint. In this section, we will illustrate how the difficult problem of finding solutions to the quantum constraints can be systematically reduced to a series of simpler problems. For this, we will need to make the quantum regulator $R_{\varepsilon,\alpha}$ more specific.

The most convenient class of regulators requires some modifications in the original construction due to Thiemann. Restrictions defining this class can be summarized as follows. Fix a graph α and consider a cell $\square$ containing a vertex v in the partition of M defined by the regulator. The first restriction is that every edge s_k assigned to $\square$ must be a proper segment of an edge incident at v, oriented to be outgoing at v. The next restriction is on the closed loops β. To every pair of edges e_I, e_J assign a triangular closed loop β_{IJ} such that (i) the loop contains v but no other point in the graph α; (ii) it lies 'between' the 2-edges, in a 2-plane containing the edges, where the plane is defined up to diffeomorphisms preserving α; and (iii) it is oriented clockwise with respect to the orientation defined in the plane by the ordered pair of segments s_I, s_J. (The double index IJ labelling the loop corresponds to the single index i before, e.g., in (6.29).) Finally, the constants C^{IJK} and T^{IJK} of the regulator are,

respectively, $\pm\kappa_1, 0$ and $\pm\kappa_2, 0$ depending on the orientation of a triad of vectors tangent to the segments s_I, s_J, s_K at v relative to the background orientation on M, where κ_1 and κ_2 are fixed constants.

Given such a regulator, the action of the resulting operators $\hat{\mathcal{C}}^{\rm Eucl}(N)$ and $\hat{\mathcal{C}}(N)$ on diffeomorphism invariant elements of Cyl* has a rather simple geometric structure which can be roughly summarized as follows. Suppose $(\Psi_\alpha| \in {\rm Cyl}^\star$ is obtained by group averaging a state in $\mathcal{H}'_\alpha$. Then, if α contains no closed loops of the type introduced by the regulator at any of its vertices, it is annihilated by both the operators. If α does contain such closed loops, $\hat{\mathcal{C}}^{\rm Eucl}(N)$ removes one loop, $\hat{\mathcal{C}}(N)$ removes two loops, and in each case there is also a possible change in the intertwiners at the vertex. Following a terminology introduced by Thiemann in his regularization, closed loops of the type introduced by the regulator will be called *extraordinary*.

More precisely, constraint operators act as follows. Consider a labelled graph (α_0, j_0) such that no labelled graph belonging to the same diffeomorphism class contains an extraordinary loop labelled by $j(\rho)$ where ρ is the representation used in the regularization. Call a labelled graph with this property *simple*. Given a simple graph α, all states $(\Psi_\alpha|$ in ${\rm Cyl}^\star_{\rm diff}$ obtained by group averaging elements of $\mathcal{H}'_a$ are in the kernel of $\hat{\mathcal{C}}(N)$. This is a large class of solutions. However, these states are annihilated by each of the two terms, $\hat{\mathcal{C}}^{\rm Eucl}(N)$ and $\hat{h}\mathcal{T}(N)$, of $\hat{\mathcal{C}}(N)$ separately whence they solve both the Euclidean and the Lorentzian scalar constraint. In this sense, they are the analogues of time-symmetric solutions to the classical Hamiltonian constraint and will at best capture very special physical situations.

More interesting solutions can be obtained starting from graphs which do admit extraordinary edges. We begin by introducing some notation. Consider the set of all the labelled graphs (α', j') that can be obtained from a given (α_0, j_0) by creation of n extraordinary loops labelled by $j(\rho)$, and by the diffeomorphisms. Denote this set by $\Gamma^{(n)}_{[(\alpha_0, j_0])]}$ and denote by $\mathcal{D}^{(n)}_{[(\alpha_0, j_0)]}$ the linear span of the corresponding diffeomorphism averaged spin-network states. The resulting spaces are *finite* dimensional and have trivial intersection with one another:

$$([(\alpha_0, j_0)], n) \neq ([(\alpha'_0, j'_0)], n') \Rightarrow \mathcal{D}^{(n)}_{[(\alpha_0, j_0)]} \cap \mathcal{D}^{(m)}_{[(\alpha'_0, j'_0)]} = \{0\}. \tag{6.41}$$

As a consequence, one can show that they have the following very useful property: every $(\Psi| \in {\rm Cyl}^\star_{\rm diff}$ can be uniquely decomposed as

$$(\Psi| = \sum_{\alpha, j, n} (\Psi|_{(n)[(\alpha, j)]}, \qquad \text{where} \qquad (\Psi|_{(n)[(\alpha, j)]} \in \mathcal{D}^{(n)}_{[(\alpha, j)]}. \tag{6.42}$$

The availability of this decomposition systematizes the task of finding solutions to the scalar constraint.

Let us begin with $\hat{\mathcal{C}}^{\rm Eucl}(N)$. For the Euclidean theory, one can obtain the following surprising result:

$$(\Psi|\hat{\mathcal{C}}^{\rm Eucl}(N) = 0 \quad \Leftrightarrow \quad (\Psi|_{(n)[(\alpha, j)]}\hat{\mathcal{C}}^{\rm Eucl} = 0, \qquad \text{for every } [(\alpha, j)], n. \tag{6.43}$$

Thus, $(\Psi|$ *is a solution of the Euclidean part of the constraint if and only if each of its components with respect to the decomposition* (6.42) *is also a solution*. This is a very useful property because the problem of finding a general diffeomorphism invariant solution to the Euclidean constraint is reduced to that of finding solutions in *finite*-dimensional subspaces. On each of these subspaces, one just has to find the kernel of certain matrices, a problem that can be readily put on a computer. Reciprocally, given any diffeomorphism invariant solution to the constraint (e.g., the state supported just on flat connections) the decomposition provides a family of new solutions. In $2+1$ dimensions this property implies that any semi-classical state can be obtained by a superposition of these 'elementary' solutions.

Finally, let us turn to the full (Lorentzian) scalar constraint operator, $\hat{\mathcal{C}}(N)$. In the present scheme, the problem of obtaining diffeomorphism invariant solutions is reduced to a hierarchy of steps. More precisely, the equation

$$(\Psi|\hat{\mathcal{C}}(N) = 0, \tag{6.44}$$

is equivalent to the following hierarchy of equations

$$\begin{aligned}
&(\Psi_{(1)[\alpha,j]}|\hat{\mathcal{T}}(N) = 0,\\
&(\Psi_{(2)[\alpha,j]}|\hat{\mathcal{T}}(N) = (\Psi_{(1)[\alpha,j]}|\hat{\mathcal{C}}^{\mathrm{Eucl}}(N),\\
&\ldots
\end{aligned} \tag{6.45}$$

$$\begin{aligned}
&(\Psi_{(n+1)[\alpha,j]}|\hat{\mathcal{T}}(N) = \sum_n (\Psi_{(n)[\alpha,j]}|\hat{\mathcal{C}}^{\mathrm{Eucl}}(N),\\
&\ldots
\end{aligned} \tag{6.46}$$

In general, the procedure involves infinitely many steps. However, it gives a partial control on the solutions and suggests new ansatze, e.g., requiring that the series terminate after a finite number of steps.

Remark. Since the procedure outlined above is a variation on Thiemann's original strategy, it is worth comparing the relative merits. Thiemann's procedure is simpler in that two of the three edges of triangles β_i (holonomy around which captures the curvature term in $\mathcal{C}^{\mathrm{Eucl}}(N)$) are along edges of the graph under consideration. However, now the analogues of our spaces $\mathcal{D}^{(n)}_{(\alpha,j)}$ overlap, making the procedure for solving the constraint more complicated.

To summarize, in this section we have presented a general framework for defining the Hamiltonian constraint and for finding its solutions. This procedure provides a good handle on the problem and also brings out the ambiguities involved. Specifically, each choice of a diffeomorphism covariant quantum regulator $R_{\varepsilon,\alpha}$ gives rise to a quantum constraint operator $\hat{\mathcal{C}}$ on $\mathcal{D}$. For each choice of the regulator, there is also a certain factor ordering freedom which was ignored for brevity. In general these operators will differ from each other, defining distinct quantum dynamics and one has to invoke *physical* criteria to test their viability. Quantum cosmology results discussed in section 7 favour the factor ordering used here. There have also been attempts at restricting the freedom in the choice of the quantum regulator by imposing heuristically motivated conditions. However, a canonical choice has not emerged. Thus, there is still a great deal of ambiguity and it is not clear if any of the candidates are fully viable. A key criterion is that the solution set to the constraints should be rich enough to admit a large number of semi-classical states. This issue will not be systematically resolved until one has a greater control on the semi-classical sector of the theory. As discussed above there is now a general strategy to find solutions, whence one can hope to address this issue. Partial support for this strategy comes from $(2+1)$-dimensional Euclidean general relativity. As mentioned above, in this theory, all semi-classical states can be recovered by superposing the 'elementary solutions' onto quantum constraints, obtained via our systematic procedure. This result is encouraging because the $2+1$ theory has all the conceptual problems associated with the absence of a background geometry. However, it cannot be taken as a strong indication because the $(2+1)$-dimensional theory has only a finite number of degrees of freedom.

This concludes the general framework for quantum kinematics and dynamics. In the next three sections we will discuss various applications.

7. Applications of quantum geometry: quantum cosmology

In cosmology, one generally freezes all but a finite number of degrees of freedom by imposing spatial homogeneity (and sometimes also isotropy). Because of the resulting mathematical simplifications, the framework provides a simple arena to test ideas and constructions introduced in the full theory both at the classical and quantum levels. Moreover, in the classical regime, the symmetry reduction captures the large scale dynamics of the universe as a whole quite well. Therefore, in the quantum theory, it provides a fertile test bed for analysing the important issues related to the fate of the initial singularity (highlighted in section 1). Over the last three years, Bojowald and his collaborators have made striking advances in this area by exploiting the quantum nature of geometry [104–117]. In this section, we will provide a self-contained summary of the core developments using constructions which mimic those introduced in sections 4–6. (For subtleties and details, see especially [115].)

Loop quantum cosmology also provides a number of lessons for the full theory. However, to fully understand their implications, it is important to keep track of the differences between the symmetry reduced and the full theories. The most obvious difference is the tremendous simplification resulting from the reduction of a field theory to a mechanical system. However, there are also two other differences which make it conceptually and technically *more* complicated, at least when one tries to directly apply the methods developed for the full theory in section 6. First, the reduced theory is usually treated by gauge fixing and therefore fails to be diffeomorphism invariant. As a result, key simplifications that occur in the treatment of full quantum dynamics do not carry over and, in a certain sense, dynamics now acquires *new* ambiguities in the reduced theory! The second complication arises from the fact that spatial homogeneity introduces distant correlations. Consequently, in contrast to section 4, quantum states associated with distinct edges and electric flux operators associated with distinct 2-surfaces are no longer independent. Both these features give rise to certain complications which are *not* shared by the full theory. Once these differences are taken into account, loop quantum cosmology can be used to gain valuable insights about certain qualitative features of the methods introduced in section 6.3 to formulate and solve the Hamiltonian constraint in the full theory.

7.1. Phase space

For simplicity, we will restrict ourselves to spatially homogeneous, isotropic cosmologies. (For non-isotropic models, see [113, 116].) Specifically, we will focus only on the case where the isometry group $\mathcal{S}$ is the Euclidean group. Then the three-dimensional group T of translations (ensuring homogeneity) acts simply and transitively on the 3-manifold M. Therefore, M is topologically $\mathbb{R}^3$. It is convenient to fix on M a fiducial flat metric ${}^0q_{ab}$, an associated constant orthonormal triad ${}^0e^a_i$ and the dual co-triad ${}^0\omega^i_a$.

Let us now turn to the gravitational phase space in the connection variables. As we saw in section 2, in the full theory, the phase space consists of pairs $\left(A^i_a, P^a_i\right)$ of fields on a 3-manifold M, where A^i_a is an $SU(2)$ connection and P^a_i a triplet of vector fields with density weight 1. A pair $\left(A'^{\,i}_a, P'^a_i\right)$ on M is said to be spatially homogeneous and isotropic or, for brevity, *symmetric* if for every $s \in \mathcal{S}$ there exists a local gauge transformation $g : M \to SU(2)$, such that

$$(s^\star A', s^\star P') = (g^{-1}A'g + g^{-1}\,\mathrm{d}g, g^{-1}P'g). \tag{7.1}$$

As is usual in cosmology, we will fix the local diffeomorphism and gauge freedom. To do so, note first that for every symmetric (A', P') (satisfying the Gauss and diffeomorphism

constraints) there exists an equivalent pair (A, P) (under (7.1)) such that

$$A = \underline{c}\,{}^0\omega^i\tau_i, \qquad P = \underline{p}\sqrt{{}^0q}\,{}^0e_i\tau^i \tag{7.2}$$

where $\underline{c}$ and $\underline{p}$ are constants, carrying the only non-trivial information contained in the pair (A', E'), and the density weight of P has been absorbed in the determinant of the fiducial metric. Denote by $\mathcal{A}_S$ and $\mathbf{\Gamma}^S_{\rm grav}$ the subspace of the gravitational configuration space $\mathcal{A}$ and of the gravitational phase space $\mathbf{\Gamma}$ defined by (7.2). Tangent vectors δ to $\mathbf{\Gamma}^S_{\rm grav}$ are of the form

$$\delta = (\delta A, \delta P), \qquad \text{with} \quad \delta A \equiv (\delta\underline{c})\,{}^0\omega^i_a, \quad \delta P \equiv (\delta\underline{p})\,{}^0e^a_i. \tag{7.3}$$

Thus, $\mathcal{A}_S$ is one dimensional and $\mathbf{\Gamma}^S_{\rm grav}$ is two dimensional: we made a restriction to *symmetric* fields and solved and gauge-fixed the gauge and the diffeomorphism constraints, thereby reducing the infinite, local, gravitational degrees of freedom to just 1.

Because M is non-compact and fields are spatially homogeneous, various integrals featuring in the Hamiltonian framework of section 2 diverge. This is, in particular, the case for the symplectic structure of the full theory. However, one can bypass this problem in a natural fashion: fix a 'cell' $\mathcal{V}$ adapted to the fiducial triad and restrict all integrations to this cell. The volume V_0 of this cell (with respect to the fiducial metric ${}^0q_{ab}$) can also be used to absorb the dependence of the basic variables $\underline{c}$, $\underline{p}$ on the fiducial ${}^0q_{ab}$. Let us rescale the basic variables to remove this dependence,

$$c := V_0^{\frac{1}{3}}\underline{c} \quad \text{and} \quad p := 8\pi G\gamma V_0^{\frac{2}{3}}\underline{p}, \tag{7.4}$$

and express the gravitational symplectic structure $\mathbf{\Omega}$ on $\mathbf{\Gamma}$ in terms of them:

$$\begin{aligned}\mathbf{\Omega}(\delta_1, \delta_2) &= \int_{\mathcal{V}} \mathrm{d}^3x \left(\delta_1 A^i_a(x)\delta_2 P^a_i(x) - \delta_2 A^i_a(x)\delta_1 P^a_i(x)\right) \\ &= 3\,\mathrm{d}c \wedge \mathrm{d}p.\end{aligned} \tag{7.5}$$

This expression also makes no reference to the fiducial metric (or the volume V_0 of the cell $\mathcal{V}$). We will work with this phase space description. Note that now the configuration variable c is dimensionless while the momentum variable p has dimensions (length)^2. (While comparing results in the full theory, it is important to bear in mind that these dimensions are different from those of the gravitational connection and the triad there.) In terms of p, the physical triad and co-triad are given by

$$e^a_i = (\mathrm{sgn}\,p)|p|^{-\frac{1}{2}}\left(V_0^{\frac{1}{3}}\,{}^0e^a_i\right) \qquad \text{and} \qquad e^i_a = (\mathrm{sgn}\,p)|p|^{\frac{1}{2}}\left(V_0^{-\frac{1}{3}}\,{}^0\omega^0_a\right). \tag{7.6}$$

We have specified the gravitational part of the reduced phase space. We do not need to specify matter fields explicitly but only note that, upon similar restriction to symmetric fields and fixing of gauge and diffeomorphism freedom, one is led to a finite-dimensional phase space also for matter fields. Finally, let us turn to constraints. Since the Gauss and the diffeomorphism constraints are already satisfied, there is a single non-trivial scalar/Hamiltonian constraint (corresponding to a constant lapse):

$$-\frac{6}{\gamma^2}c^2\,\mathrm{sgn}\,p\sqrt{|p|} + C_{\rm matter} = 0. \tag{7.7}$$

7.2. Quantization: kinematics

We will now adapt the general procedure of sections 4 and 5 to the symmetry reduced phase space and emphasize how it leads to interesting departures from the 'standard' quantum cosmology in geometrodynamic variables.

7.2.1. Elementary variables. Let us begin by singling out 'elementary functions' on the classical phase space which are to have unambiguous quantum analogues. In the full theory, the configuration variables were constructed from holonomies $A(e)$ associated with edges e and momentum variables, from $E(S, f)$, triads E smeared with test fields f on 2-surfaces. But now, because of homogeneity and isotropy, we do not need all edges e and surfaces S. Symmetric connections A in $\mathcal{A}_S$ can be recovered knowing holonomies $h(e)$ along straight lines in M. Similarly, it is now appropriate to smear triads only by constant fields, $f_i = \tau_i$, and across squares to which the fiducial triads ${}^0e^a_i$ are tangent[33].

The $SU(2)$ holonomy along an edge e is given by

$$A(e) = \cos\frac{\ell c}{2} + 2\left[\sin\frac{\ell c}{2}\right]\left(\dot{e}^{a\,0}\omega^i_a\right)\tau^i \tag{7.8}$$

where $\ell V_0^{1/3}$ is the oriented length of the edge. Therefore, a typical element of the algebra generated by sums of products of matrix elements of these holonomies can be written as

$$F(A) = \sum_j \xi_j\, \mathrm{e}^{\mathrm{i}\ell_j c} \tag{7.9}$$

where j runs over a finite number of integers (labelling edges), $\ell_j \in \mathbb{R}$ and $\xi \in \mathbb{C}$. These are precisely the *almost periodic functions* which have been studied in the mathematical literature in detail. One can regard a finite number of edges as providing us with a graph (since, because of homogeneity, the edges need not actually meet in vertices now) and the function $F(A)$ as a cylindrical function with respect to that graph. The vector space of these almost periodic functions is the space of cylindrical functions of symmetric connections and will be denoted by $\mathrm{Cyl_S}$.

To define the momentum functions, we are now led to smear the triads with constant test functions and integrate them on a square (with respect to the fiducial metric). The resulting phase space function is then just p multiplied by a kinematic factor. We will therefore regard p itself as the momentum function. In terms of classical geometry, p is related to the physical volume of the elementary cell $\mathcal{V}$ via $V = V_0|p|^{3/2}$. Finally, the only non-vanishing Poisson bracket between elementary functions is

$$\{F(A), p\} = \frac{8\pi\gamma G}{6}\sum_j (\mathrm{i}\ell_j\xi_j)\,\mathrm{e}^{\mathrm{i}\ell_j c}. \tag{7.10}$$

Since the right-hand side is again in $\mathrm{Cyl_S}$, the space of elementary variables is closed under the Poisson bracket. Note that, in contrast with the full theory, now there is only one momentum variable whence non-commutativity of triads is no longer an issue. Therefore, the triad representation also exists in quantum theory. In fact, it turns out to be convenient in making the quantum dynamics explicit.

7.2.2. Representation of the algebra of elementary variables. To construct quantum kinematics, let us seek a representation of this algebra of elementary variables. We will find that the quantum theory is quite different from the 'standard' geometrodynamical quantum cosmology. This difference arises from the fact that the configuration variables are not smooth functions of compact support on $\mathcal{A}_s$, but rather, almost periodic functions. As we saw in section 7.2.1, this choice can be directly traced back to the full theory where holonomies play a primary role. Because we are repeating the procedure used in the full theory as closely as possible, the fundamental discreteness underlying polymer geometry will trickle down to

[33] Indeed, we could just consider edges lying in a single straight line and a single rectangle bounded by ${}^0e^a_i$. We chose not to break the symmetry artificially and consider instead all lines and all rectangles.

quantum cosmology and lead to results which are *qualitatively different* from those in standard quantum cosmology.

Recall from section 4.3.4 that one can construct the representation of the algebra of elementary variables using Gel'fand theory. In the reduced model under consideration, the theory implies that the Hilbert space must be the space of square integrable functions on a suitable completion $\bar{\mathcal{A}}_S$ of the classical configuration space $\mathcal{A}_S$. Now, $\mathcal{A}_S = \mathbb{R}$ and $\bar{\mathcal{A}}_S$ is the Gel'fand spectrum of the $C^\star$-algebra of almost periodic functions (see (7.9)) on $\mathcal{A}_S$. This is a well-understood space, called the *Bohr compactification* of the real line (discovered by the mathematician Harold Bohr, Neils' brother). This is now the quantum configuration space. It is an Abelian group and carries a canonical, normalized Haar measure μ_0^S. Following the procedure used in the full theory, it is natural to set $\mathcal{H}^S_{\rm grav} = L^2\big(\bar{\mathcal{A}}_S, \mathrm{d}\mu_0^S\big)$ and use it as the kinematical Hilbert space for the gravitational sector of the theory. The general theory can also be used to represent the algebra of elementary variables by operators on $\mathcal{H}^S_{\rm grav}$.

As in section 4, one can make this representation concrete. Set

$$\mathcal{N}_\ell(\bar{A}) = \mathrm{e}^{\mathrm{i}\ell c} \tag{7.11}$$

and introduce on Cyl the following Hermitian inner product,

$$\big\langle \mathcal{N}_{\ell_1} \big| \mathcal{N}_{\ell_2} \big\rangle = \delta_{\ell_1,\ell_2} \tag{7.12}$$

where the right-hand side is the Kronecker delta, rather than Dirac. Then $\mathcal{H}^S_{\rm grav}$ is the Cauchy completion of this space. Thus the almost periodic functions $\mathcal{N}_\ell$ constitute an orthonormal basis in $\mathcal{H}^S_{\rm grav}$; they play the role of spin networks in the reduced theory. $\mathrm{Cyl}_{\rm S}$ is dense in $\mathcal{H}^S_{\rm grav}$, and serves as a common domain for all elementary operators. The configuration and momentum operators have expected actions:

$$(\hat{F}\mathcal{N}_\ell)(\bar{A}) = F(\bar{A})\mathcal{N}_\ell(\bar{A}), \qquad (\hat{p}\mathcal{N}_\ell)(\bar{A}) = \frac{8\pi\ell\gamma\ell_{\rm Pl}^2}{6}\mathcal{N}_\ell(A). \tag{7.13}$$

As in the full theory, the configuration operators are bounded, whence their action can be extended to the full Hilbert space $\mathcal{H}^S_{\rm grav}$, while the momentum operators are unbounded but essentially self-adjoint. The basis vectors $\mathcal{N}_\ell$ are normalized eigenstates of $\hat{p}$. As in quantum mechanics, let us use the bra–ket notation and write $\mathcal{N}_\ell(c) = \langle c|\ell\rangle$. Then,

$$\hat{p}|\ell\rangle = \frac{8\pi\ell\gamma\ell_{\rm Pl}^2}{6}|\ell\rangle \equiv p_\ell|\ell\rangle. \tag{7.14}$$

Using the relation $V = |p|^{3/2}$ between p and the physical volume of the cell $\mathcal{V}$ we have

$$\hat{V}|\ell\rangle = \left(\frac{8\pi\gamma|\ell|}{6}\right)^{\frac{3}{2}}\ell_{\rm Pl}^3|\ell\rangle \equiv V_\ell|\ell\rangle. \tag{7.15}$$

This provides us with a physical meaning of ℓ: when the universe is in the quantum state $|\ell\rangle$, (modulo a fixed constant) $|\ell|^{3/2}$ is the *physical* volume of the cell $\mathcal{V}$ *in Planck units*. Thus, in particular, while the volume V_0 of the cell $\mathcal{V}$ with respect to the fiducial metric ${}^0q_{ab}$ may be 'large', its physical volume in the quantum state $|\ell = 1\rangle$ is just $(\gamma/6)^{3/2}\ell_{\rm Pl}^3$.

As our notation makes clear, the construction of the Hilbert space and the representation of the algebra are entirely parallel to that in the full theory. In the full theory, holonomy operators are well defined but there is no operator representing the connection itself. Similarly, $\hat{\mathcal{N}}_\ell$ are well-defined unitary operators on $\mathcal{H}^S_{\rm grav}$ but they fail to be continuous with respect to ℓ, whence there is no operator corresponding to c on $\mathcal{H}^S_{\rm grav}$. Thus, as in section 4.3, to obtain physically interesting operators, one has to *first express them in terms of the elementary variables $\mathcal{N}_\ell$ and p and then promote those expressions to the quantum theory*.

There is, however, one important difference between the full and the reduced theories: while eigenvalues of the momentum (and other geometric) operators in the full theory span only a discrete subset of the real line, now every real number is a permissible eigenvalue of $\hat{p}$. This difference can be directly attributed to the high degree of symmetry. In the full theory, eigenvectors are labelled by a pair (e, j) consisting of continuous label e (denoting an edge) and a discrete label j (denoting the 'spin' on that edge), and the eigenvalue is dictated by j. Because of homogeneity and isotropy, the pair (e, j) has now collapsed to a single continuous label ℓ. Note, however, that there *is* a weaker sense in which the spectrum is discrete: all eigenvectors are *normalizable*. Hence the Hilbert space can be expanded out as a direct *sum*—rather than a direct integral—of the one-dimensional eigenspaces of $\hat{p}$; i.e., the decomposition of identity on $\mathcal{H}_S$ is given by a (continuous) sum

$$I = \sum_{\ell} |\ell\rangle\langle\ell| \tag{7.16}$$

rather than an integral. Consequently, the natural topology on the spectrum of $\hat{p}$ is *discrete*. Although weaker, this discreteness plays a critical role both technically and conceptually.

7.3. Triad operator

In the reduced classical theory, curvature is simply a multiple of the *inverse* of the square of the scale factor $a = \sqrt{|p|}$. Similarly, the matter Hamiltonian invariably involves a term corresponding to an *inverse* power of a. Therefore, one needs to obtain an operator corresponding to the inverse scale factor, or the triad (with density weight zero) of (7.6). In the classical theory, the triad coefficient diverges at the big-bang and a key question is whether quantum effects 'tame' the big-bang sufficiently to make the triad operator (and hence the curvature and the matter Hamiltonian) well behaved there.

Now, given a self-adjoint operator $\hat{A}$ on a Hilbert space, the function $f(\hat{A})$ is well defined if and only if f is a measurable function on the spectrum of A. Thus, for example, in non-relativistic quantum mechanics, the spectrum of the operator $\hat{r}$ is the positive half of the real line, equipped with the standard Lesbegue measure, whence the operator $1/\hat{r}$ is a well-defined, self-adjoint operator. By contrast, the spectrum of $\hat{p}$ has discrete topology and $[\hat{p}]^{-1}$ is not a measurable function of $\hat{p}$. More explicitly, since $\hat{p}$ admits a *normalized* eigenvector $|\ell = 0\rangle$ with zero eigenvalue, the naive expression of the triad operator fails to be densely defined on $\mathcal{H}^S_{\rm grav}$. One could circumvent this problem in the reduced model in an *ad hoc* manner by just making up a definition for the action of the triad operator on $|\ell = 0\rangle$. But then the result would have to be considered as an artefact of a procedure expressly invented for the model and one would not have any confidence in its implications for the big-bang. However, we saw in section 6.3 that a similar problem arises in the full theory and can be resolved using a strategy due to Thiemann. It is natural to use the same procedure also in quantum cosmology. As in the general theory, therefore, we will proceed in two steps. In the first, we note that, on the reduced phase space $\mathbf{\Gamma}^S_{\rm grav}$, the triad coefficient $\mathrm{sgn}\, p|p|^{-\frac{1}{2}}$ can be expressed as the Poisson bracket $\{c, V^{1/3}\}$ which can be replaced by $i\hbar$ times the commutator in quantum theory. However, as in the full theory, a second step is necessary because there is no operator $\hat{c}$ on $\mathcal{H}^S_{\rm grav}$ corresponding to c: one has to re-express the Poisson bracket in terms of *holonomies* which do have unambiguous quantum analogues. The resulting triad (coefficient) operator is given by

$$\widehat{\left[\frac{\mathrm{sgn}(p)}{\sqrt{|p|}}\right]} = -\frac{12\mathrm{i}}{8\pi\gamma\ell^2_{\rm Pl}}\left(\sin\frac{c}{2}\hat{V}^{\frac{1}{3}}\cos\frac{c}{2} - \cos\frac{c}{2}\hat{V}^{\frac{1}{3}}\sin\frac{c}{2}\right) \tag{7.17}$$

where V_ℓ is the eigenvalue of the volume operator given in (7.15). Although the triad involves both configuration and momentum operators, it commutes with $\hat{p}$, whence its eigenvectors are again $|\ell\rangle$. The eigenvalues are given by

$$\left[\widehat{\frac{\mathrm{sgn}(p)}{\sqrt{|p|}}}\right]|\ell\rangle = \frac{6}{8\pi\gamma\ell_{\mathrm{Pl}}^2}\left(V_{\ell+1}^{1/3} - V_{\ell-1}^{1/3}\right)|\ell\rangle. \tag{7.18}$$

where V_ℓ is the eigenvalue of the volume operator (see (7.15)).

A key property of the triad operator follows immediately: it is bounded above! The upper bound is obtained at the value $\ell = 1$:

$$|p|_{\mathrm{max}}^{-\frac{1}{2}} = \sqrt{\frac{12}{8\pi\gamma}}\ell_{\mathrm{Pl}}^{-1}. \tag{7.19}$$

Since in the classical theory the curvature is proportional to p^{-1}, in quantum theory, it is bounded above by $(12/\gamma)\ell_{\mathrm{Pl}}^{-2}$. Note that $\hbar$ is essential for the existence of this upper bound; as $\hbar$ tends to zero, the bound goes to infinity just as one would expect from classical considerations. This is rather reminiscent of the situation with the ground state energy of the hydrogen atom in non-relativistic quantum mechanics, $E_0 = -(m_{\mathrm{e}}e^4/2)(1/\hbar)$, which is bounded from below because $\hbar$ is non-zero.

While this boundedness is physically appealing, at first it also seems puzzling because the triad coefficient and the momentum are algebraically related in the classical theory via $p \cdot (\mathrm{sgn}\, p/|p|^{1/2})^2 = 1$ and $\hat{p}$ admits a *normalized* eigenvector with zero eigenvalue. A key criterion of viability of the triad operator is that the classical relation should be respected in an appropriate sense. More precisely, one can tolerate violations of this condition on states only in the Planck regime; the equality must be satisfied to an excellent approximation on states with large ℓ (i.e., with large volume). Is this the case? We have

$$\begin{aligned}\frac{6}{\gamma\ell_{\mathrm{Pl}}^2}\left(V_{\ell+1}^{1/3} - V_{\ell-1}^{1/3}\right) &= \sqrt{\frac{6|\ell|}{8\pi\gamma\ell_{\mathrm{Pl}}^2}}\left(\sqrt{1+1/|\ell|} - \sqrt{1-1/|\ell|}\right)\\ &= \sqrt{\frac{6}{8\pi\gamma|\ell|\ell_{\mathrm{Pl}}^2}}\left(1 + O(\ell^{-2})\right).\end{aligned} \tag{7.20}$$

Thus, up to order $O(\ell^{-2})$, the eigenvalue of the triad operator is precisely $1/\sqrt{|p_\ell|}$, where p_ℓ is the eigenvalue of $\hat{p}$ (see (7.15)). On states representing a large universe ($\ell \gg 1$), the classical algebraic relation between the triad coefficient and p is indeed preserved to an excellent approximation. Violations are significant only on the eigen-subspace of the volume operator with eigenvalues of the order of ℓ_{Pl}^3 or less, i.e., in the fully quantum regime.

Since the classical triad diverges at the big-bang, it is perhaps not surprising that quantum effects usher in the Planck scale. However, the mechanism by which this came about is new and conceptually important. For, the procedure did not call for a cut-off or a regulator. The classical expression of the triad coefficient we began with is *exact* whence the issue of removing the regulator does not arise. It *is* true that the quantization procedure is somewhat 'indirect'. However, this was necessary because *the spectrum of the momentum operator* $\hat{p}$ (or of the 'scale factor operator' corresponding to a) *is discrete* in the sense detailed in section 7.2. Had the Hilbert space $\mathcal{H}_{\mathrm{grav}}^S$ been a direct integral of the eigenspaces of $\hat{p}$ (rather than a direct sum) the triad operator could then have been defined directly using the spectral decomposition of $\hat{p}$ and would have been unbounded above. Indeed, this is precisely what happens in geometrodynamics. Thus, the key differences between the mathematical structures of the present quantum theory and quantum geometrodynamics are responsible for the boundedness of the triad coefficient on the entire Hilbert space [115].

A natural question then is as follows: how can there be such inequivalent quantizations? After all, here we are dealing with a system with a finite number of degrees of freedom. Does the von Neumann uniqueness theorem not ensure that there is a unique representation of the exponentiated Heisenberg relations? The answer is no: the von Neumann theorem requires that the unitary operators $U(\ell)$, $V(\lambda)$ corresponding respectively to the classical functions $\exp(\mathrm{i}\ell c)$ and $\exp(\mathrm{i}\lambda p)$ be *weakly continuous* in the parameters ℓ and λ. As we noted in section 7.2.2, this assumption is violated by $U(\ell)$ in our representation—this in fact is the reason why there is no operator corresponding to c and only the holonomies are well defined. *A priori* it may seem that by dropping the continuity requirement, one opens up a Pandora's box. Which of the possible representations is one to use? The most important aspect of this construction is that it came directly from the full theory where, as discussed at the end of section 4.3.4, diffeomorphism invariance severely constrains the choice representation.

7.4. Quantum dynamics: the Hamiltonian constraint

Since the curvature is bounded above on the entire kinematical Hilbert space $\mathcal{H}^S_{\rm grav}$, one might expect that the classical singularity at the big-bang would be naturally resolved in the quantum theory. This turns out to be the case.

7.4.1. The constraint operator. Rather than starting from the reduced Hamiltonian constraint (7.7), to bring out the relation to the full theory, we will return to the full constraint and use the procedure spelled out in section 6.3. Because of spatial homogeneity and flatness, two simplifications arise: (i) the two terms in expression (6.14) of the full Hamiltonian constraint are now proportional; and (ii) without loss of generality we can restrict ourselves to a constant lapse function N, and we will just set it to 1. Then, the gravitational part of the constraint can be written as

$$\mathcal{C}_{\rm grav} = -\frac{V_0}{\sqrt{8\pi G\gamma^2}}\epsilon_{ijk}F^i_{ab}\frac{P^{aj}P^{bk}}{\sqrt{\det P}}. \tag{7.21}$$

As in section 6.3, we have to 'regulate' this classical expression by writing it in terms of phase space functions which can be directly promoted to quantum operators. As in the full theory, we can express the curvature components F^i_{ab} in terms of holonomies. Consider a square α_{ij} in the i–j plane spanned by two of the triad vectors ${}^0e^a_i$, each of whose sides has length $\ell_0 V_0^{1/3}$ with respect to the fiducial metric ${}^0q_{ab}$. Then, 'the ab component' of the curvature is given by

$$F^i_{ab}\tau_i = {}^0\omega^i_a\,{}^0\omega^j_b\left(\frac{A(\alpha_{ij})-1}{\ell_0^2 V_0^{2/3}} + O(c^3\ell_0)\right). \tag{7.22}$$

The holonomy $A(\alpha_{ij})$ around the square α_{ij} can be expressed as a product

$$A(\alpha_{ij}) = A(e_i)A(e_j)A\big(e_i^{-1}\big)A\big(e_j^{-1}\big) \tag{7.23}$$

where holonomies along individual edges are given by

$$A(e_i) := \cos\frac{\ell_0 c}{2} + 2\sin\frac{\ell_0 c}{2}\tau_i. \tag{7.24}$$

Next, let us consider the triad term $\epsilon_{ijk}E^{aj}E^{bk}/\sqrt{\det E}$. As in the full theory, this can be handled through the Thiemann regularization. Thus, let us begin with the identity on the symmetry reduced phase space $\mathbf{\Gamma}^S_{\rm grav}$,

$$\epsilon_{ijk}\tau^i\frac{P^{aj}P^{bk}}{\sqrt{\det P}} = -2\big(\gamma\ell_0(8\pi G)^{1/2}V_0^{1/3}\big)^{-1}\epsilon^{abc}\,{}^0\omega^k_c A(e_k)\big\{A\big(e_k^{-1}\big), V\big\} \tag{7.25}$$

where $A(e_k)$ is the holonomy along the edge e_k parallel to the kth basis vector of length $\ell_0 V_0^{1/3}$ with respect to ${}^0q_{ab}$. Note that, unlike expression (7.22) for F^i_{ab}, (7.25) is exact, i.e., does not depend on the choice of ℓ_0.

Collecting terms, we can now express the gravitational part of the 'regulated' constraint as

$$C^{\ell_0}_{\rm grav} = -4\big(8\pi G\gamma^3\ell_0^3\big)^{-1}\sum_{ijk}\epsilon^{ijk}\,{\rm Tr}\big(A(e_i)A(e_j)A\big(e_i^{-1}\big)A\big(e_j^{-1}\big)A(e_k)\big\{A\big(e_k^{-1}\big),V\big\}\big) \qquad (7.26)$$

where the term proportional to identity in the leading contribution to F^i_{ab} in (7.22) drops out because of the trace operation and where we used $\epsilon^{abc}{}^0\omega_a^i{}^0\omega_b^j{}^0\omega_c^k = \sqrt{{}^0q}\epsilon^{ijk}$. In the limit $\ell_0 \to 0$, the right-hand side of $C^{\ell_0}_{\rm grav}$ reproduces the classical expression (7.7) of the constraint. Thus, ℓ_0 (or the length of the edge used while expressing F_{ab} in terms of the holonomy around the square α_{ij}) plays the role of a regulator in (7.26). Because of the presence of the curvature term, there is no natural way to express the constraint *exactly* in terms of our elementary variables; a limiting procedure is essential. This faithfully mirrors the situation in the full theory: there, again, the curvature term is recovered by introducing small loops at vertices of graphs and the classical expression of the constraint is recovered only in the limit in which the loop shrinks to zero.

It is now straightforward to pass to quantum theory. The regulated quantum constraint is

$$\begin{aligned}\hat{C}^{(\ell_0)}_{\rm grav} &= 4{\rm i}\big(8\pi\gamma^3\ell_0^3\ell_{\rm Pl}^2\big)^{-1}\sum_{ijk}\epsilon^{ijk}\,{\rm Tr}\big(\bar{A}(e_i)\bar{A}(e_j)\bar{A}\big(e_i^{-1}\big)\bar{A}\big(e_j^{-1}\big)\bar{A}(e_k)\big[\bar{A}\big(e_k^{-1}\big),\hat{V}\big]\big)\\ &= 96{\rm i}\big(8\pi\gamma^3\ell_0^3\ell_{\rm Pl}^2\big)^{-1}\sin^2\frac{\ell_0 c}{2}\cos^2\frac{\ell_0 c}{2}\left(\sin\frac{\ell_0 c}{2}\hat{V}\cos\frac{\ell_0 c}{2}-\cos\frac{\ell_0 c}{2}\hat{V}\sin\frac{\ell_0 c}{2}\right).\end{aligned} \qquad (7.27)$$

Its action on the eigenstates of $\hat{p}$ is

$$\hat{C}^{(\ell_0)}_{\rm grav}|\ell\rangle = 3\big(8\pi\gamma^3\ell_0^3\ell_{\rm Pl}^2\big)^{-1}\big(V_{\ell+\ell_0}-V_{\ell-\ell_0}\big)(|\ell+4\ell_0\rangle-2|\ell\rangle+|\ell-4\ell_0\rangle). \qquad (7.28)$$

On physical states, this action must equal that of the matter Hamiltonian $-8\pi G\hat{C}_{\rm matter}$.

In the full theory, one could remove the regulator to obtain a well-defined operator on (suitable) diffeomorphism invariant states in Cyl*. The reduced model, on the other hand, does not have diffeomorphism invariance. Therefore, one would expect that the obvious $\ell_0 \to 0$ limit would run into problems. This is indeed what happens. In this limit, the classical regulated expression (7.26) equals the Hamiltonian constraint (7.7) which, however, contains c^2. Consequently, the naive limit of the operator $\hat{C}^{(\ell_0)}_{\rm grav}$ also contains $\hat{c}^2$. However, since $\hat{c}^2$ is not well defined on $\mathcal{H}^S_{\rm grav}$, now the limit as $\ell_0 \to 0$ fails to exist. Thus, one cannot remove the regulator in the quantum theory of the reduced model. This feature can be traced back directly to the symmetry reduction [115].

A detailed analysis shows that the presence of ℓ_0 in the quantum Hamiltonian constraint should be regarded as a quantization ambiguity. Indeed, as we discussed in section 6.3, even in the full theory, there is a similar ambiguity associated with the choice of the j label used on the new edges introduced to define the operator corresponding to F_{ab} [102]. More precisely, in the full theory, the quantization procedure involves the introduction of a pair of labels (e, j) where e is a continuous label denoting the new edge and j is a discrete label denoting the spin on that edge. Diffeomorphism invariance ensures that the quantum constraint is insensitive to the choice of e but the dependence on j remains as a quantization ambiguity. In the reduced model, diffeomorphism invariance is lost and the pair (e, j) of the full theory collapses into a single continuous label ℓ_0 denoting the length of the edge introduced to define F_{ab}. The dependence on ℓ_0 persists—there is again a quantization ambiguity but it is now labelled by a *continuous* label ℓ_0.

If one works in the strict confines of the reduced model, there does not appear to exist a natural way to remove this ambiguity. In the full theory, on the other hand, one can fix the ambiguity by assigning the *lowest* non-trivial j value, $j = 1/2$, to each extra loop introduced to determine the operator analogue of F_{ab}. This procedure can be motivated by the following heuristics. In the classical theory, one can use a loop enclosing an arbitrarily small area in the a–b plane to determine F_{ab} locally. In quantum geometry, on the other hand, the area operator (of an open surface) has a lowest eigenvalue $a_0 = (\sqrt{3}\pi\gamma)\ell_{\rm Pl}^2$, suggesting that it is *physically* inappropriate to try to localize F_{ab} on arbitrarily small surfaces. The best one could do is to consider a loop spanning an area a_0, consider the holonomy around the loop to determine the integral of F_{ab} on a surface of area a_0, and then extract an effective, local F_{ab} by setting the integral equal to $a_0 F_{ab}$. It appears natural to use the same physical considerations to remove the quantization ambiguity also in the reduced model. Then, we are led to set the area of the smallest square spanned by α_{ij} to a_0, i.e., to set $(\gamma\ell_0)\ell_{\rm Pl}^2 = a_0$ or $\ell_0 = \sqrt{3}\pi$. Thus, while in the reduced model itself, area eigenvalues can assume arbitrarily small values, if one 'imports' from the full theory the value of the smallest non-zero area eigenvalue, one is naturally led to set $\ell_0 = \sqrt{3}\pi$.

To summarize, in loop quantum cosmology, one adopts the viewpoint that (7.27), with $\ell_0 = \sqrt{3}\pi$, is the 'fundamental' Hamiltonian constraint operator which 'correctly' incorporates the underlying discreteness of quantum geometry and the classical expression (7.7) is an approximation which is valid only in regimes where this discreteness can be ignored and the continuum picture is valid. This viewpoint is borne out by detailed calculations [115]: the expectation values of $\hat{C}^{\ell_0}_{\rm grav}$ in semi-classical states reproduce the classical constraint. Furthermore, one can calculate corrections to the classical expression arising from the fundamental discreteness and quantum fluctuations inherent in the semi-classical quantum states [117].

7.4.2. Physical states. Let us now solve the quantum constraint and obtain physical states. For simplicity, we assume that the matter is only minimally coupled to gravity (i.e., there are no curvature couplings). As in non-trivially constrained systems, one expects that the physical states would fail to be normalizable in the kinematical Hilbert space $\mathcal{H}^S = \mathcal{H}^S_{\rm grav} \otimes \mathcal{H}^S_{\rm matter}$ (see, e.g., [88, 89]). However, as in the full theory, they do have a natural 'home'. We again have a triplet

$$\mathrm{Cyl}_{\rm S} \subset \mathcal{H}^S \subset \mathrm{Cyl}^\star_S$$

of spaces and physical states will belong to $\mathrm{Cyl}^\star_{\rm S}$, the algebraic dual of $\mathrm{Cyl}_{\rm S}$. Since elements of $\mathrm{Cyl}^\star_{\rm S}$ need not be normalizable, as in section 6.3, we will denote them by $(\Psi|$. (The usual, normalizable bras will be denoted by $\langle\Psi|$.)

It is convenient to exploit the existence of a triad representation. Then, every element $(\Psi|$ of $\mathrm{Cyl}^\star_S$ can be expanded as

$$(\Psi| = \sum_\ell \psi(\phi, \ell)\langle\ell| \tag{7.29}$$

where ϕ denotes the matter field and $\langle\ell|$ are the (normalized) eigenbras of $\hat{p}$. Note that the sum is over a continuous variable ℓ whence $(\Psi|$ need not be normalizable. Now, the constraint equation

$$(\Psi|\big(\hat{C}^{(\ell_0)}_{\rm grav} + 8\pi G\hat{C}^{(\ell_0)}_{\rm matter}\big)^\dagger = 0 \tag{7.30}$$

turns into the equation

$$\begin{aligned}\big(V_{\ell+5\ell_0} - V_{\ell+3\ell_0}\big)\psi(\phi, \ell+4\ell_0) - 2\big(V_{\ell+\ell_0} - V_{\ell-\ell_0}\big)\psi(\phi, \ell) + \big(V_{\ell-3\ell_0} - V_{\ell-5\ell_0}\big)\psi(\phi, \ell-4\ell_0)\\ = -\tfrac{1}{3}8\pi G\gamma^3\ell_0^3\ell_{\rm Pl}^2\hat{C}^{(\ell_0)}_{\rm matter}(\ell)\psi(\phi, \ell)\end{aligned} \tag{7.31}$$

for the coefficients $\psi(\phi, \ell)$, where $\hat{C}^{(\ell_0)}_{\rm matter}(\ell)$ only acts on the matter fields (and depends on ℓ via metric components in the matter Hamiltonian). Note that, even though ℓ is a continuous variable, the quantum constraint is a *difference* equation rather than a differential equation. Strictly, (7.31) just constrains the coefficients $\psi(\phi, \ell)$ to ensure that $(\Psi|$ is a physical state. As in the full quantum theory, we do not have a background spacetime, hence no natural notion of 'time' or 'evolution'. However, since each $\langle\ell|$ is an eigenbra of the volume operator, it tells us how the matter wavefunction is correlated with volume, i.e., geometry. Now, if one wishes, one can regard p as providing a heuristic 'notion of time', and then think of (7.31) as an evolution equation for the quantum state of matter with respect to this time. (Note that p goes from $-\infty$ to ∞, negative values corresponding to triads which are oppositely oriented to the fiducial one. The classical big-bang corresponds to $p = 0$.) This heuristic interpretation often provides physical intuition for (7.31); one can regard it as a discrete 'evolution' equation. However, it is *not* essential for what follows; one can forego this interpretation entirely and regard (7.31) only as a constraint equation.

What is the fate of the classical singularity? At the big-bang, the scale factor goes to zero. Hence it corresponds to the state $|\ell = 0\rangle$ in $\mathcal{H}^S_{\rm grav}$. So, the key question is whether the quantum 'evolution' breaks down at $\ell = 0$. Let us examine this issue. Starting at $\ell = -4N\ell_0$ for some large positive N, and fixing values of $\psi(\phi, -4N\ell_0)$ and $\psi(\phi, (-4N+4)\ell_0)$, one can use the equation to determine the coefficients $\psi(\phi, (-4N+4n)\ell_0)$ for all $n > 1$, *provided* the coefficient of the highest order term in (7.31) continues to remain non-zero. Now, it is easy to verify that the coefficient vanishes if and only if $n = N$. Thus, the coefficient $\psi(\phi, \ell = 0)$ remains undetermined. In its place, one just obtains a consistency condition constraining the coefficients $\psi(\phi, \ell = -4)$ and $\psi(\phi, \ell = -8)$. Now, since $\psi(\phi, \ell = 0)$ remains undetermined, at first sight, it may appear that one cannot 'evolve' past the singularity, i.e., the quantum evolution also breaks down at the big-bang. However, the main point is that this is *not* the case. For the structure of the quantum scalar constraint is such that *the coefficient* $\psi(\phi, \ell = 0)$ *just decouples from the rest*. This comes about because of two facts: (i) the minimally coupled matter Hamiltonians annihilate $\psi(\phi, \ell = 0) = 0$ [104, 110]; and (ii) $V_{\ell_0} = V_{-\ell_0}$. Thus, unlike in the classical theory, evolution does not stop at the singularity; the difference equation (7.31) lets us 'evolve' right through it. In this analysis, we started at $\ell = -4N\ell_0$ because we wanted to test what happens if one encounters the singularity 'head on'. If one begins at a generic ℓ, the 'discrete evolution' determined by (7.31) just 'jumps' over the classical singularity without encountering any subtleties.

Next, let us consider the space of solutions. An examination of the classical degrees of freedom suggests that the freedom in physical quantum states should correspond to two functions *just* of matter fields ϕ. The space of solutions to the Hamiltonian constraint, on the other hand, is much larger: there are as many solutions as there are functions $\psi(\phi, \ell)$ on an interval $[\ell' - 4\ell_0, \ell' + 4\ell_0)$, where ℓ' is any fixed number. This suggests that a large number of these solutions may be redundant. Indeed, to complete the quantization procedure, one needs to introduce an appropriate inner product on the space of solutions to the Hamiltonian constraint. The physical Hilbert space is then spanned by just those solutions to the quantum constraint which have finite norm. In simple examples one generally finds that, while the space of solutions to all constraints can be very large, the requirement of finiteness of norm suffices to produce a Hilbert space of the physically expected size.

For the reduced system considered here, one has a quantum mechanical system and with a single constraint in quantum cosmology. Hence it should be possible to extract physical states using the group averaging technique of the 'refined algebraic framework' [88, 89]. However, this analysis is yet to be carried out explicitly and therefore one does not yet have a good control on how large the physical Hilbert space really is. This issue is being investigated.

To summarize, two factors were key to the resolution of the big-bang singularity: (i) as a direct consequence of quantum geometry, the Hamiltonian constraint is now a difference equation rather than a differential equation as in geometrodynamics; and (ii) the coefficients in the difference equation are such that one can evolve unambiguously 'through' the singularity even though the coefficient $\psi(\phi, \ell = 0)$ is undetermined. Both these features are robust: they are largely insensitive to factor ordering ambiguities and persist in more complicated cosmological models [111, 113, 116]. The qualitative changes introduced by quantum geometry in kinematics and dynamics are significant only in the Planck regime. A careful analysis shows that the discrete evolution is extremely well approximated by the Wheeler–DeWitt differential equation at scales larger than ℓ_0 in a precise sense [115]. Thus the fundamental discreteness, characteristic of loop quantum gravity, intervenes in a subtle way precisely in the Planck regime where geometrodynamics fails to resolve singularities. Since loop quantum cosmology mimics the full theory as closely as possible, within the limitations discussed at the very beginning of section 7, these results provide support for the approach to quantum dynamics in the full theory.

8. Applications: quantum geometry of isolated horizons and black-hole entropy

Loop quantum cosmology illuminates dynamical ramifications of quantum geometry but within the limited context of mini-superspaces where all but a finite number of degrees of freedom are frozen. In this section, we will discuss a complementary application where one considers the full theory but probes consequences of quantum geometry which are not sensitive to the details of how the Hamiltonian constraint is imposed. (For further details, see [127–129, 133–137, 139]. For early work, see [122–124].)

As was explained in section 1, since mid-1970s, a key question in the subject has been as follows: what is the statistical mechanical origin of the black-hole entropy $S_{\rm BH} = \big(a_{\rm hor}/4\ell_{\rm Pl}^2\big)$? What are the microscopic degrees of freedom that account for this entropy? This relation implies that a solar mass black hole must have $(\exp 10^{77})$ quantum states, a number that is *huge* even by the standards of statistical mechanics. Where do all these states reside? To answer these questions, in the early 1990s Wheeler [140] suggested the following heuristic picture, which he christened '*It* from *Bit*'. Divide the black-hole horizon into elementary cells, each with one Planck unit of area $\ell_{\rm Pl}^2$, and assign to each cell two micro-states. Then the total number of states $\mathcal{N}$ is given by $\mathcal{N} = 2^n$ where $n = \big(a_{\rm hor}/\ell_{\rm Pl}^2\big)$ is the number of elementary cells, whence entropy is given by $S = \ln \mathcal{N} \sim a_{\rm hor}$. Thus, apart from a numerical coefficient, the entropy (*It*) is accounted for by assigning two states (*Bit*) to each elementary cell. This qualitative picture is simple and attractive. Thus, one is led to ask the following: can these heuristic ideas be supported by a systematic analysis from first principles? What is the origin of the 'elementary cells'? Why is each cell endowed with precisely two states? And most importantly: *what has all this to do with a black hole?* The 'It from Bit' considerations seem to apply to any 2-surface! Quantum geometry enables one to address all these issues in detail.

The precise picture is much more involved: because the area spectrum is quite complicated in quantum geometry, 'elementary cells' need not all carry the same area and the number of 'internal states' of each cell is also not restricted to 2. Nonetheless, it does turn out that the dominant contribution in the entropy calculation comes from states envisaged by Wheeler. The purpose of this section is to summarize the overall situation. Our discussion is divided into three parts. In the first, we recall the 'isolated horizon framework' in classical general relativity which serves as a point of departure for quantum theory. In the second we discuss the quantum geometry of the simplest (undistorted and non-rotating) horizons and present the

entropy calculation allowing for the presence of minimally coupled matter. In the third, we discuss extensions that include non-minimal couplings, distortions and rotation.

8.1. Isolated horizons

A systematic approach to the problem of entropy requires that we first specify the class of horizons of interest. Since the entropy formula is expected to hold unambiguously for black holes in equilibrium, most analyses were confined to *stationary*, eternal black holes (i.e., to the Kerr–Newman family in four-dimensional general relativity). From a physical viewpoint however, this assumption seems overly restrictive. After all, in statistical mechanical calculations of entropy of ordinary systems, one only has to assume that the given system is in equilibrium, not the whole world. Therefore, it should suffice for us to assume that the black hole itself is in equilibrium; the exterior geometry should not be forced to be time independent. Furthermore, the analysis should also account for entropy of black holes which may be distorted or carry (Yang–Mills and other) hair. Finally, it has been known since the mid-1970s that the thermodynamical considerations apply not only to black holes but also to cosmological horizons [121]. A natural question arises: can these diverse situations be treated in a single stroke? In classical general relativity, the isolated horizon framework provides a natural avenue by encompassing all these situations. It also provides a Hamiltonian framework which serves as a natural point of departure for quantization [128, 133, 134]

Let us begin with the basic definitions [133]. In this discussion, we need to begin with a four-dimensional spacetime manifold $\mathcal{M}$ although we will return to 3-manifolds M once we have a Hamiltonian framework.

A *non-expanding horizon* Δ is a null, three-dimensional sub-manifold of the four-dimensional spacetime $(\mathcal{M}, g_{\alpha\beta})$, with topology $S^2 \times R$, such that

(i) the expansion θ_ℓ of its null normal ℓ vanishes; and
(ii) field equations hold on Δ with stress energy, $T_{\alpha\beta}$, satisfying the very weak requirement that $-T^\alpha{}_\beta \ell^\beta$ is a future-directed, causal vector. (Throughout, ℓ^α will be assumed to be future pointing.)

Note that (i) if the expansion vanishes for one null normal, it vanishes for all; and (ii) the condition on stress energy is satisfied by all the standard matter fields *provided they are minimally coupled to gravity*.

The definition ensures that the area of any 2-sphere cross-section of the horizon is constant and matter flux across Δ vanishes. It also implies that the spacetime derivative operator ∇ naturally induces a unique derivative operator $\mathcal{D}$ on Δ. Since Δ is a null 3-surface, it has a degenerate intrinsic 'metric' q_{ab} of signature 0, +, +. The pair $(q_{ab}, \mathcal{D})$ is referred to as *the geometry* of Δ. The notion that the black hole itself is in equilibrium is captured by requiring that this geometry is time independent:

An *isolated horizon* (Δ, ℓ) is a non-expanding horizon Δ equipped with a null normal ℓ such that ℓ is a symmetry of the geometry, i.e., $\mathcal{L}_\ell q_{ab} = 0$ and $[\mathcal{L}_\ell, D] = 0$ on Δ.

One can show that, generically, the null normal ℓ satisfying these conditions is unique up to a constant rescaling. For simplicity, we will assume that we are in the generic case although the main results go through in all cases.

The isolated horizon definition extracts from the notion of the Killing horizon just that 'tiny' part which turns out to be essential for black-hole mechanics and, more generally, to capture the notion that the horizon is in equilibrium, allowing for dynamical processes and radiation in the exterior region. Indeed, Einstein's equations admit solutions with isolated horizons in which there is radiation arbitrarily close to the horizons [131, 132]. Note that the

definition uses conditions which are *local* to Δ. Thus, unlike event horizons, this notion is local, not 'teleological'. For our purposes, the two important considerations are as follows:

(i) The definition is satisfied not only by event horizons of stationary black holes but also by the standard cosmological horizons. Thus, all situations in which thermodynamical considerations apply are treated in one stroke.
(ii) If one restricts oneself to spacetimes which admit an internal boundary which is an isolated horizon, *the action principle and the Hamiltonian description are well defined* and the resulting phase space has an infinite number of degrees of freedom. This would not be the case if one used general event horizons or Killing horizons instead.

Next, let us examine symmetry groups of isolated horizons. A *symmetry* of $(\Delta, \ell, q, \mathcal{D})$ is a spacetime diffeomorphism which maps Δ to itself; at most rescales ℓ by a constant, and preserves q and $\mathcal{D}$. It is clear the diffeomorphisms generated by any smooth extension of ℓ^α are symmetries. So, the symmetry group G_Δ is at least one dimensional. The question is: are there any other symmetries? At infinity, we generally have a universal symmetry group (such as the Poincaré or the anti-de Sitter) because all metrics under consideration approach a fixed metric (Minkowskian or anti-de Sitter) there. In the case of the isolated horizons, generically we are in the strong field regime and spacetime metrics do not approach a universal metric. Therefore, the symmetry group is not universal. However, there are only three universality classes:

(i) Type I: the isolated horizon geometry is spherical; in this case, G_Δ is four dimensional.
(ii) Type II: the isolated horizon geometry is axi-symmetric; in this case, G_Δ is two dimensional.
(iii) Type III: the diffeomorphisms generated by ℓ^α are the only symmetries; G_Δ is one dimensional.

Note that these symmetries refer *only to* the horizon geometry. The full spacetime metric need not admit any isometries even in a neighbourhood of the horizon. Physically, type II horizons are the most interesting ones. They include the Kerr–Newman horizons as well as their generalizations incorporating distortions (due to exterior matter of other black holes) and hair. The zeroth and the first laws of black-hole mechanics can be naturally extended to type II isolated horizons [133, 134]. In particular, for the Einstein–Maxwell theory, one can define the mass M_Δ and angular momentum J_Δ of the horizon using only the intrinsic geometry of the isolated horizon and show that the first law holds:

$$\mathbf{d}M_\Delta = \frac{\kappa}{8\pi G}\,\mathbf{d}a_\Delta + \Omega\,\mathbf{d}J_\Delta + \Phi\,\mathbf{d}Q_\Delta \tag{8.1}$$

where κ, Ω, Φ are, respectively, the surface gravity, angular velocity and the electric potential at the horizon and $\mathbf{d}$ denotes the exterior derivative on the (infinite-dimensional) phase space. This law of isolated horizon mechanics encompasses all black holes and cosmological horizons in equilibrium, including those with arbitrary distortion and rotation.

8.2. Type I isolated horizons: quantum theory

Let us first discuss type I isolated horizons in detail and then generalize the results to include non-minimally coupled matter and type II horizons. We will divide the type I discussion into three parts. In the first, we introduce the Hamiltonian formulation; in the second, we describe the quantum horizon geometry; and in the third, we summarize the entropy calculation.

8.2.1. Hamiltonian framework. Consider the sector of general relativity consisting of gravitational and matter fields for which the underlying spacetime admits an internal boundary which is a type I isolated horizon Δ *with a fixed area* a_0. We will focus on geometrical structures near Δ and on the modifications of the Hamiltonian framework of section 2 caused by the presence of an internal boundary.

Denote by S the 2-sphere intersection of the (partial) Cauchy surface M with the isolated horizon Δ. Introduce on Δ an internal vector field r^i, i.e., any isomorphism from the unit 2-sphere in the Lie algebra of $SU(2)$ to S, and partially gauge fix the internal $SU(2)$ freedom to $U(1)$ by requiring that $r^i P_i^a = \sqrt{|\det q|}\, r^a$, where r^a is the unit normal to S. Then it turns out that *the intrinsic geometry of* Δ *is completely determined by the pullback* $\underline{A}^i r_i =: 2W$ *to* S *of the connection* A^i *on* M [135]. Furthermore, W is in fact a spin connection intrinsic to the 2-sphere S: $W = \frac{1}{2}\underline{A}^i r_i = \frac{1}{2}\underline{\Gamma}^i r_i$ on S (see (2.20)). Thus, if we consider orthonormal dyads $(m, \bar{m})$ on S with internal rotation freedom in $SO(2)$, W is a connection on the corresponding $U(1)$ bundle. Now, this $U(1)$ bundle on S is non-trivial and $\oint_S \mathrm{d}W$ equals -2π, rather than zero. (But since the Chern class of any spin connection is the same, $\oint_S \delta W = 0$; tangent vectors δW to the phase space are genuine 1-forms, globally defined on S. This fact will be useful in the discussion of the symplectic structure.) Finally, the fact that S is (the intersection of M with) a type I isolated horizon is captured in a relation between the two canonically conjugate fields,

$$F := \mathrm{d}W = -\frac{2\pi}{a_0} 8\pi G\gamma \underline{\Sigma}^i r_i, \tag{8.2}$$

where $\underline{\Sigma}^i$ is the pullback to S of the 2-forms $\Sigma^i_{ab} = \eta_{abc} P^a_j \eta^{ij}$ on M, dual to the momentum P^a_i. Thus, because of the isolated horizon boundary conditions, fields which would otherwise be independent are now related. As one would expect, the boundary conditions *reduce* the number of independent fields; in particular, the pullbacks to S of the canonically conjugate fields A^i_a, Σ^i_{ab} are completely determined by the $U(1)$ connection W.

The main modification in the Hamiltonian framework of section 2 is that the gravitational symplectic structure now acquires a surface term,

$$\mathbf{\Omega}(\delta_1, \delta_2) = -\int_M \mathrm{Tr}(\delta_1 A \wedge \delta_2 \Sigma - \delta_2 A \wedge \delta_1 \Sigma) + \frac{1}{2\pi} \frac{a_0}{4\pi G\gamma} \oint_S \delta_1 W \wedge \delta_2 W, \tag{8.3}$$

where, as in section 2, $\delta \equiv (\delta A, \delta\Sigma)$ denote tangent vectors to the phase space $\mathbf{\Gamma}$. Since W is essentially the only 'free data' on the horizon, it is not surprising that the surface term of the symplectic structure is expressible entirely in terms of W. However, it is interesting and somewhat surprising that the new surface term is precisely the symplectic structure of the $U(1)$-Chern–Simons theory. *The symplectic structures of the Maxwell, Yang–Mills, scalar and dilatonic fields do not acquire surface terms.* Conceptually, this is an important point: this, in essence, is the reason why (for minimally coupled matter) the black-hole entropy depends just on the area and not, in addition, on the matter charges.

8.2.2. Quantum horizon geometry. In the classical theory, the bulk fields determine the surface fields just by continuity; there are no *independent* degrees of freedom on the surface in the classical phase space. In the quantum theory, on the other hand, the fields are distributional and arbitrarily discontinuous whence the surface and the bulk fields effectively decouple. It is this phenomenon that is responsible for creating 'independent surface states' in the quantum theory.

The main task is to extend the 'bulk' quantum geometry of sections 4 and 5 to allow for the presence of an internal boundary S. Now, the space of generalized connections $\bar{\mathcal{A}}$ is a product

$\bar{\mathcal{A}} = \bar{\mathcal{A}}_V \times \bar{\mathcal{A}}_S$, where a volume generalized connection $\bar{A}_V$ assigns an $SU(2)$ element to any (closed-piecewise analytic) edge lying in the bulk while $\bar{A}_S$ assigns a $U(1)$ element to each (closed-piecewise analytic) edge lying in the surface S. Therefore, it is natural to begin with a total Hilbert space $\mathcal{H} = \mathcal{H}_V \otimes \mathcal{H}_S$ where $\mathcal{H}_V$ is built from suitable functions of generalized connections in the bulk and $\mathcal{H}_S$ from suitable functions of generalized surface connections. The volume Hilbert space $\mathcal{H}_V$ is that which comes from bulk quantum geometry of section 4: $\mathcal{H}_V = L^2(\bar{\mathcal{A}}, \mu_0)$. The question is: what would be the surface Hilbert space? The answer is suggested by the structure of the surface term in the symplectic structure: it should be the Hilbert space of Chern–Simons theory on the horizon. Furthermore, the coefficient in front of this surface term tells us that the quantum Chern–Simons theory must have a (dimensionless) coupling constant/level k given by

$$k = \frac{a_0}{4\pi\gamma\ell_{\rm Pl}^2}. \tag{8.4}$$

But we also have to incorporate the boundary condition (8.2) *which ensures that S is not any old 2-surface but an isolated horizon*. The *key* idea is to impose it quantum mechanically, *as an operator equation* on $\mathcal{H}$, i.e., via

$$(1 \otimes \hat{F})\Psi = -\frac{2\pi}{a_0} 8\pi G\gamma (\hat{\underline{\Sigma}} \cdot r) \otimes 1)\Psi, \tag{8.5}$$

where the notation emphasizes that $\hat{F}$ is a *surface* operator while $\hat{\underline{\Sigma}}$ is an operator on the *volume* Hilbert space. It is easy to show that a basis of solutions is given by states of the type $\Psi = \Psi_V \otimes \Psi_S$ where Ψ_V is an eigenstate of the volume operator, Ψ_S an eigenstate of the surface operator with the *same* eigenvalues. Now, all the eigenvalues of the bulk operator on the right-hand side of (8.5) are known from bulk quantum geometry of section 4.3.4 (see (4.49)). They are given by

$$-\left(\frac{2\pi}{a_0}\right)\left(8\pi\ell_{\rm Pl}^2 \sum_I m_I \delta^3(x, p_I)\eta_{ab}\right), \tag{8.6}$$

where m_I are half-integers, the sum ranges over a finite set of points—called punctures—on S and where η_{ab} is the metric-independent Levi-Civita density on S. Therefore, the quantum boundary condition (8.5) tells us that $\mathcal{H}_S$ should be the Hilbert space of $U(1)$-Chern–Simons theory on a punctured 2-sphere S where the curvature F has the form of a δ-distribution concentrated at a finite number of punctures.

Let us then begin by fixing a set $\mathcal{P}$ of punctures on S and consider $U(1)$-Chern–Simons theory on this punctured sphere. The phase space of this theory is $\mathbf{\Gamma}_S^{\mathcal{P}} = \left(\bar{\mathcal{A}}_S^0\right)/(\bar{\mathcal{G}}^{\mathcal{P}} \rtimes \mathcal{D}^{\mathcal{P}})$ where $\bar{\mathcal{A}}_S^0$ is the space of connections which is flat everywhere except at the punctures; $\bar{\mathcal{G}}^{\mathcal{P}}$ is the space of local $U(1)$ gauge transformations which are identity except at the punctures; $\mathcal{D}^{\mathcal{P}}$ is the space of diffeomorphisms of S which fix the punctures and certain structure at the punctures; and $\rtimes$ stands for semi-direct product[34]. This phase space is isomorphic with the torus $\mathbb{T}^{2(n-1)}$ if there are n punctures in the set $\mathcal{P}$, equipped with the natural symplectic structure on $\mathbb{T}^{2(n-1)}$. A convenient set of canonically conjugate coordinates can be introduced as follows. Let us fix the nth puncture as the 'origin' and, as in figure 4, denote by γ_I, a family of curves joining the Ith puncture to the nth (with $I = 1, 2, \ldots, n-1$) and by η_I,

[34] The extra structure one needs to fix at the punctures is listed at the end of section 4.3.1 of [129]. From the physics perspective, this is the most delicate of the technical subtleties in the subject, although this procedure is 'standard practice' in the mathematics literature. It plays an important role in the imposition of the diffeomorphism constraint and state counting.

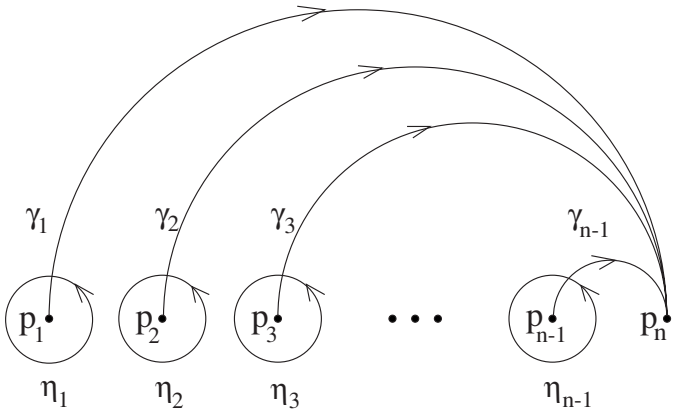

Figure 4. Coordinatization of the surface phase space

'small' closed loops surrounding each of the first $n-1$ punctures. Then, for each I, two holonomies around γ_I and η_I are canonically conjugate. This phase space is often referred to as the *non-commutative torus*.

The Hilbert space $\mathcal{H}_S^{\mathcal{P}}$ of surface states results from geometric quantization of this torus [129]. This is the space of quantum states of the $U(1)$-Chern–Simons theory on $(S, \mathcal{P})$. The total surface Hilbert space $\mathcal{H}_S$ is the inductive limit of these Hilbert spaces as the set $\mathcal{P}$ becomes larger and larger. As discussed in section 4, the volume Hilbert space $\mathcal{H}_V = L^2(\bar{\mathcal{A}}_v, \mathrm{d}\mu_0)$ can also be obtained as the inductive limit of the Hilbert spaces $(\mathcal{H}_V)_\alpha$ associated with graphs α [44, 45, 66].

Next one has to impose the quantum boundary condition (8.5). *This introduces a highly non-trivial test of the whole framework.* Construction of the surface Hilbert space was strongly motivated by (8.5). However, now that it is complete, there is no more freedom. In the Chern–Simons Hilbert space, one can compute the eigenvalues of the surface operator $\hat{F}$. This calculation is *completely independent* of the volume Hilbert space; it has never heard of the quantum geometry in the bulk. The *key* question on which everything hinges is: are the eigenvalues of $\hat{F}$ the same as the eigenvalues (8.6) of the bulk operator in (8.5)? If not, there will be no solutions to the quantum boundary conditions! The remarkable fact is that *the infinite set of eigenvalues of the two operators match*, even though the two calculations are completely distinct. This comes about because the level of the Chern–Simons theory is related to the Barbero–Immirzi parameter γ and the area a_0 in a very specific way, which in turn is a consequence of the isolated horizon boundary conditions. Thus there is a seamless matching between three completely independent theories: the isolated horizon framework in classical general relativity; the bulk quantum geometry; and the Chern–Simons theory on the punctured horizon. And this matching provides a coherent mathematical description of the quantum geometry of the horizon.

Finally, one has to impose quantum Einstein's equations. The Gauss constraint asks that the total state $\Psi_V \otimes \Psi_S$ should be gauge invariant. The diffeomorphism constraint asks that diffeomorphisms on S should be regarded as gauge. Again there are important mathematical subtleties. But the final picture is simple. While each of the bulk and the surface states transforms non-trivially under the remaining gauge freedom $(\bar{\mathcal{G}}/\bar{\mathcal{G}}^{\mathcal{P}})$, the total state is gauge invariant as needed. Implementation of the remaining diffeomorphism

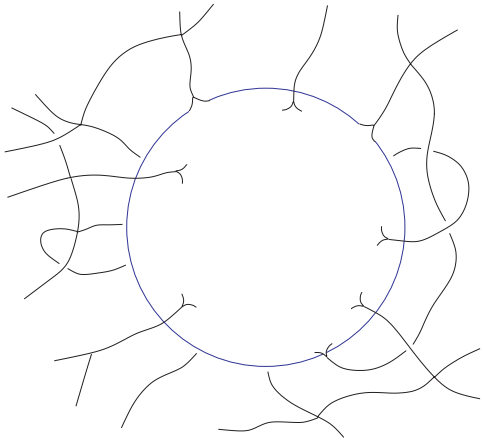

Figure 5. Quantum horizon. Polymer excitations in the bulk puncture the horizon, endowing it with quantized area. Intrinsically, the horizon is flat except at punctures where it acquires a quantized deficit angle. These angles add up to endow the horizon with a 2-sphere topology.

constraint (corresponding to $\mathcal{D}/\mathcal{D}^{\mathcal{P}}$) requires only the number of punctures; their location is irrelevant[35].

The Hamiltonian constraint, by contrast, does not restrict the surface states, i.e., the quantum geometry of the horizon. This is because in the classical theory, the constraint is functionally differentiable (i.e., generates gauge) only when the smearing function (the lapse) goes to zero on the isolated horizon boundary (and, as usual, at infinity). The time evolution along the isolated horizon is generated by a *true Hamiltonian*, not just the constraint.

Let us summarize. The physical surface Hilbert space $\mathcal{H}_S^{\rm Phys}$ is given by $\mathcal{H}_S^{\rm Phys} = \oplus_n \mathcal{H}_s^n$, with $\mathcal{H}_S^n$, the Hilbert space of the $U(1)$ Chern–Simons theory on the sphere S with n punctures, where the polymer excitations of the bulk geometry intersect S (see figure 5). Let us focus on $\mathcal{H}_S^n$. Since W is the intrinsic spin connection on S and since F vanishes except at the punctures, the intrinsic geometry of the quantum horizon is flat except at the n punctures. The η_i-holonomies around these punctures are non-trivial, whence the punctures carry deficit angles which, furthermore, are quantized. They add up to 4π, providing a quantum analogue of the Gauss–Bonnet theorem.

Remark. (i) Note that the above analysis makes a crucial use of the horizon boundary condition; it is not applicable to a general 2-surface. Thus the most important limitation of Wheeler's 'It from Bit' considerations is overcome. The strategy of incorporating the boundary condition through an *operator equation* (8.5) allows both the connection W and the triad $\underline{\Sigma}^i r_i$ to fluctuate and requires only that they do so in tandem. This equation provides the first step in the answer to the question: what is a quantum black hole?

[35] The subtleties are as follows: (i) in the Chern–Simons theory, only the exponentiated operator $\exp(\mathrm{i}\hat{F})$ is well defined; $\hat{F}$ itself is not. Therefore, the mathematically meaningful quantum boundary condition is the exponentiated version of (8.5). (ii) The $U(1)$ gauge group at the punctures is replaced by the quantum $U(1)$ group. The deformation parameter is supplied by the level k of the Chern–Simons theory which is required to be an integer because of the pre-quantization requirements. This also implies that the deficit angles at each puncture are quantized. (iii) Recall that one has to fix certain structure at the punctures in the construction of the surface Hilbert space. Under $\mathcal{D}/\mathcal{D}^{\mathcal{P}}$, this structure changes. Therefore, strictly, to begin with one has infinitely many copies of the surface Hilbert spaces, one for each choice of the extra structure, and the fact that $\mathcal{D}/\mathcal{D}^{\mathcal{P}}$ is gauge implies that only one of these copies is physically relevant. That is, the fact that diffeomorphisms in $\mathcal{D}/\mathcal{D}^{\mathcal{P}}$ are gauge is incorporated by 'gauge fixing'.

(ii) For extensions of this framework to include non-minimally coupled fields and type II horizons (discussed in section 8.3), let us note that there are just three essential mathematical ingredients which serve as the input for this construction of the surface Hilbert space: (a) the form of the surface term in (8.3) which shows that the surface symplectic structure is that of the $U(1)$-Chern–Simons theory with level k of (8.4); (b) the horizon boundary condition (8.2); and (c) the spectrum (8.6) of the triad operator $\underline{\Sigma}^i r_i$.

8.2.3. Entropy: counting surface states. In the classical theory, a_0 in the expression of the surface term of the symplectic structure (8.3) and in the boundary condition (8.2) is the horizon area. However in the *quantum theory*, a_0 has simply been a parameter so far; we have not tied it to the *physical area of the horizon*. To calculate entropy, one has to construct a suitable 'micro-canonical' ensemble by relating a_0 to the physical area.

It follows from the definition of volume connections that, as depicted in figure 5, the polymer excitations of the bulk geometry puncture the horizon *transversely* from the 'exterior'. Hence, the relevant area eigenvalues are those given in (5.18):

$$8\pi\gamma\ell_{\rm Pl}^2 \sum_I \sqrt{j_I(j_I+1)}.$$

Therefore, one is led to construct the micro-canonical ensemble by considering only that subspace of the bulk theory which, at the horizon, satisfies

$$a_0 - \epsilon \leqslant 8\pi\gamma\ell_{\rm Pl}^2 \sum_I \sqrt{j_I(j_I+1)} \leqslant a_0 + \epsilon \tag{8.7}$$

where I ranges over the number of punctures, j_I is the spin label (the eigenvalue of the vertex operator $\hat{J}_{p_I}$ associated with the puncture p_I) and $\ell_{\rm Pl}^2 < \epsilon \ll a_0$. In the presence of matter fields carrying charges, one fixes values of horizon charges $Q_0^{(\alpha)}$ (labelled by α) and restrict the matter configurations so that

$$Q_0^{(\alpha)} - \epsilon^{(\alpha)} \leqslant Q_0^{(\alpha)} \leqslant Q_0^{(\alpha)} + \epsilon^{(\alpha)} \tag{8.8}$$

for suitably chosen $\epsilon^{(\alpha)}$. (As is usual in statistical mechanics, the leading contribution to the entropy is independent of the precise choice of these small intervals.) Now, the physical states belonging to this ensemble contain information also about gravitational and electromagnetic radiation far away from the horizon which is obviously irrelevant to the calculation of black-hole entropy. What is relevant are the states directly associated with the horizon of a given area a_0, and charges $Q_0^{(\alpha)}$. One is therefore led to trace over the volume degrees of freedom and construct a density matrix ρ_Δ describing a maximum entropy mixture of surface satisfying (8.7) and (8.8). The statistical mechanical entropy is then given by $S_\Delta = -{\rm Tr}(\rho_\Delta \ln \rho_\Delta)$. As usual, this number can be calculated simply by counting states,

$$S_\Delta = \ln \mathcal{N}_\Delta \tag{8.9}$$

where $\mathcal{N}_\Delta$ is the number of Chern–Simons surface states consistent with the area and charge constraints. A detailed analysis [129] estimates this number and leads to the expression of entropy of large black holes:

$$S_\Delta := \ln \mathcal{N}_\Delta = \frac{\gamma_0}{\gamma}\frac{a_0}{4\ell_{\rm Pl}^2} + o\left(\frac{\ell_{\rm Pl}^2}{a_0}\right), \qquad \text{where} \quad \gamma_0 = \frac{\ln 2}{\sqrt{3}\pi}. \tag{8.10}$$

Thus, ignoring terms $o\big(\frac{\ell_{\rm Pl}^2}{a_0}\big)$, entropy is indeed proportional to the horizon area. However, even for large black holes, one obtains agreement with the Hawking–Bekenstein formula *only* in the sector of quantum geometry in which the Barbero–Immirzi parameter γ takes the value

$\gamma = \gamma_0$. Thus, while all γ sectors are equivalent classically, the standard quantum field theory in curved spacetimes is recovered in the semi-classical theory only in the γ_0 sector of quantum geometry. It is quite remarkable that thermodynamical considerations involving *large* black holes can be used to fix the quantization ambiguity which dictates such Planck scale properties as eigenvalues of geometric operators.

Now, the value of γ can be fixed by demanding agreement with the semi-classical result just in one case; e.g., a spherical horizon with zero charge or a cosmological horizon in the de Sitter spacetime, etc. Once the value of γ is fixed, the theory is completely fixed and one can ask: does this theory yield the Hawking–Bekenstein value of entropy of *all* isolated horizons, irrespective of the values of charges, angular momentum and cosmological constant, the amount of distortion or hair. The non-trivial fact is that the answer is in the affirmative. Thus, the agreement with quantum field theory in curved spacetimes holds in *all* these diverse cases. The physical interpretation of S_Δ is that it represents the entropy that observers in the 'external region' (used in the construction of the phase space) associate with the horizon.

Why does γ_0 not depend on other quantities such as charges? As noted in section 8.2.2, only the gravitational part of the symplectic structure develops a surface term at the horizon; the matter symplectic structures have only volume terms. (Furthermore, the gravitational surface term is insensitive to the value of the cosmological constant.) Consequently, there are no independent surface quantum states associated with matter. This provides a natural 'explanation' of the fact that the Hawking–Bekenstein entropy depends only on the horizon geometry and is independent of electromagnetic (or other) charges (of minimally coupled matter fields).

Finally, let us return to Wheeler's 'It from Bit'. One can ask: what are the states that dominate the counting? Perhaps not surprisingly, they turn out to be those which assign to each puncture the smallest quantum of area (i.e., spin value $j = \frac{1}{2}$), thereby maximizing the number of punctures. In these states, each puncture defines Wheeler's 'elementary cell' and his two states correspond to whether the deficit angle is positive or negative.

To summarize, quantum geometry naturally provides the micro-states responsible for the huge entropy associated with horizons. In this analysis, all black holes and cosmological horizons are treated in an unified fashion; there is no restriction, e.g., to near-extremal black holes. The sub-leading term has also been calculated and shown to be proportional to $\ln a_{\rm hor}$ [130, 141]. Finally, in this analysis quantum Einstein equations *are* used. In particular, had the quantum Gauss and co-vector/diffeomorphism constraints not been imposed on surface states, the spurious gauge degrees of freedom would have given an infinite entropy. However, because of the isolated horizon boundary conditions, the scalar/Hamiltonian constraint has to be imposed just in the bulk. Since in the entropy calculation one traces over bulk states, the final result is insensitive to the details of how this (or any other bulk) equation is imposed. Thus, as in other approaches to black-hole entropy, the calculation does not require complete knowledge of quantum dynamics.

8.3. Non-minimal couplings and type II horizons

We will now show that, while the introduction of non-minimal couplings [136, 137] and distortion and rotation [138, 139] does introduce interesting modifications, the qualitative picture of section 8.2 remains unaltered.

8.3.1. Non-minimal couplings. Consider a scalar field ϕ non-minimally coupled to gravity through the action

$$\mathbf{S}[g_{ab}, \phi] = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{16\pi G} f(\phi)R - \frac{1}{2}g^{ab}\nabla_a\phi\nabla_b\phi - V(\phi)\right] \qquad (8.11)$$

where f is a nowhere vanishing function (minimal coupling results if $f(\phi) = 1$) and $V(\phi)$ is the potential. Now the stress–energy tensor does not satisfy even the weak energy requirement (ii) in the definition of a non-expanding horizon. However, one can replace it by:

(ii)$'$ field equations hold on Δ and the scalar field ϕ satisfies $\mathcal{L}_\ell \phi = 0$ on Δ

to incorporate the idea that the scalar field is time independent on Δ, reflecting the fact that the horizon is in equilibrium. (For minimal couplings, time independence of matter fields on Δ is ensured by (ii).) The isolated horizon framework then leads to the zeroth and first laws [136]. However, now the form of the first law is modified,

$$\mathbf{d}M = \frac{\kappa}{8\pi G}\mathbf{d}\left[\oint_S f(\phi)\,\mathrm{d}^2V_S\right] + \Omega\,\mathbf{d}J_\Delta \tag{8.12}$$

suggesting that the entropy should be given by

$$S_\Delta = \frac{1}{4\ell_{\mathrm{Pl}}^2}\oint_S f(\phi)\,\mathrm{d}^2V. \tag{8.13}$$

(The same conclusion is reached using the general framework of [125, 126] which deals with a broad class of theories but which requires a globally defined Killing field with a bifurcate Killing horizon.) The question now is: can the statistical mechanical derivation of entropy of section 8.2, based on quantum geometry, go through also in this case? This is a non-trivial check on the robustness of that framework, first because the seamless matching between Chern–Simons theory and bulk geometry required for a coherent description of the quantum horizon geometry is very delicate, and second, because the entropy now depends not only on geometry but also on the scalar field. In spite of these non-trivialities, the framework does turn out to be robust.

For type I horizons, as one would expect, ϕ is constant on S. Let us consider the sector of the phase space consisting of fields for which Δ is a type I horizon with fixed area a_0 and scalar field ϕ_0. Then, the main modifications of the discussion of section 8.2, caused by the non-minimal coupling, are the following:

(i) Denote by Π_i^a the momentum conjugate to the gravitational connection A_a^i. If we only have minimally coupled matter, $\Pi_i^a = P_i^a$ and the geometrical triad $\tilde{E}_i^a$ is given by $\tilde{E}_i^a = 8\pi G\gamma P_i^a$ (see (2.21)). With non-minimal coupling, the geometrical triad involves *both* Π_i^a *and* the scalar field: $\tilde{E}_i^a = (8\pi G\gamma/f(\phi))\Pi_i^a$. Conceptually, this is an important change because quantum geometry is now dictated not just by the gravitational variables $\left(A_a^i, \Pi_i^a\right)$ but it also involves the matter variable ϕ.

(ii) The geometrical relation between the pulled back triad and the gravitational connection on the horizon remains unaltered. Therefore, in terms of the phase space variables, the boundary condition (8.2) is now replaced by

$$F := \mathrm{d}W = -\frac{2\pi}{f(\phi_0)a_0}8\pi G\gamma\,\underline{\Sigma}^i r_i \tag{8.14}$$

where $\underline{\Sigma}_{ab}^i$ is now the pullback to S of the dual $\eta_{abc}\Pi_j^c\eta^{ij}$ of the gravitational momentum.

(iii) The surface term in the symplectic structure is again given by that of the $U(1)$ Chern–Simons theory but the level is modified:

$$k = \frac{f(\phi_0)a_0}{4\pi\gamma\ell_{\mathrm{Pl}}^2}. \tag{8.15}$$

(iv) Since the description of the bulk Hilbert space in terms of the gravitational momentum variables remains unaltered, the eigenvalues of the gravitational momentum operator $\underline{\Sigma}^i_{ab} r_i$ continue to be given by (8.6):

$$-\left(\frac{2\pi}{a_0}\right)\left(8\pi\ell_{\rm Pl}^2 \sum_I m_I \delta^3(x, p_I)\eta_{ab}\right). \tag{8.16}$$

(v) Finally, for the 2-surface S, the area eigenvalues are now given by

$$\frac{8\pi\gamma\ell_{\rm Pl}^2}{f(\phi_0)} \sum_I \sqrt{j_I(j_I+1)}. \tag{8.17}$$

in place of (5.18).

Thus, in the key equations, a_0 is just replaced by $f(\phi_0)a_0$ everywhere. One can now repeat the analysis of section 8.2 (using the 'polymer representation' also of the scalar field [136].) Equations (8.14), (8.15) and (8.17) now imply that the quantum boundary condition does have 'enough' solutions: although the level of the Chern–Simons theory and the boundary condition are both modified, the delicate interplay between the surface and the volume sectors required for a coherent theory of the geometry of quantum horizons survives intact. The state counting procedure can be repeated and (8.17) now implies that the entropy is given by

$$S_\Delta := \ln\mathcal{N} = \frac{\gamma_0}{\gamma}\frac{f(\phi_0)a_0}{4\ell_{\rm Pl}^2} + o\left(\frac{\ell_{\rm Pl}^2}{a_0}\right), \qquad \text{where} \quad \gamma_0 = \frac{\ln 2}{\sqrt{3}\pi}. \tag{8.18}$$

Thus, if $\gamma = \gamma_0$, one obtains the answer suggested by the classical analysis. Note that the value γ_0 of the Barbero–Immirzi parameter is the *same* as it was for minimally coupled matter fields.

8.3.2. Inclusion of distortion and rotation. Let us now extend the framework of section 8.1 to type II horizons [138, 139]. It turns out that the *type II problem can be mapped on to the type I problem* already at the classical level. Thus, in the quantum theory, the underlying mathematics will be the same as in the type I case. However, the physical meaning of the Chern–Simons connection on the boundary will be different. As is usual when one maps a given, complicated problem to a mathematically simpler one, the physical non-trivialities are contained in the map. In the present case it is the map that carries all the information about distortion and angular momentum.

For brevity, let us focus just on the gravitational sector and ignore other fields on Δ. The 'free data' on type II isolated horizon are again determined by a $U(1)$ connection $V := \frac{1}{2}\underline{\rm A}^i r_i$ on S. However, now the connection also has the information about distortion and angular momentum : $\underline{\rm A}^i r_i = \underline{\Gamma}^i r_i + \gamma \underline{\rm K}^i r_i$, where $\underline{\Gamma}^i$ is a real $U(1)$ connection on the spin-bundle over S which now carries information about distortion and $\underline{\rm K}^i r_i$ is a globally defined, real-valued 1-form on S which carries information about angular momentum (see (2.20)). The gauge and diffeomorphism invariant characterization of the free data V can be coded in a pair, $M_n, J_n, n = 0, 1, 2, \ldots, \infty$ of mass and angular momentum multipoles [138]. In the type I case, M_0 is the mass and all other horizon multipoles vanish. In the type II case, M_0 is again the mass, J_0 continues to vanish, J_1 is the angular momentum and higher multipoles represent departures from sphericity. The M_n and J_n have the interpretation of 'source multipoles' of the black hole.

Recall from section 8.1 that two ingredients from the classical theory play the key role in quantization: the isolated horizon boundary condition (8.2) and the surface term in the symplectic structure (8.3). Now, (8.2) is replaced by

$$F := {\rm d}V = f 8\pi G\gamma \underline{\Sigma}^i r_i. \tag{8.19}$$

The major difference from the type I case is that f is *not a constant but a genuine function on S* (determined by the Newman–Penrose component Ψ_2 of) the Weyl curvature. Similarly, while the symplectic structure does have a surface term which is fully determined by the surface connection V, it is *not* the Chern–Simons symplectic structure for V. So, at first, the situation appears to be quite different from that in the type I case.

However, one can in fact map the present problem to the type I problem which was already solved. To see this, note first that if one is interested in any one macroscopic black hole, one has to fix its macroscopic parameters. In the globally vacuum context of classical general relativity, for example, one would fix the mass (or the horizon area) and the angular momentum. In the present, very general discussion of isolated horizons, we have allowed arbitrary distortions. Therefore, to fix the macroscopic black hole, one has to fix all multipole moments M_n, J_n. Type I phase space can also be constructed by fixing all multipoles (now $M_0 \neq 0$ and all other multipoles zero). Hence, one would intuitively expect that this sector is 'of the same size' as the type I sector. However, is this really the case? More importantly, do the arguments that one should be considering a *Chern–Simons theory on a punctured 2-sphere* go through?

The answer to both questions turns out to be in the affirmative: one can explicitly coordinatize the type II surface phase space $\mathbf{\Gamma}_S$ with a new $U(1)$ connection W on the spin-bundle over S such that the surface symplectic structure $\mathbf{\Omega}_S$ is given by

$$\mathbf{\Omega}_S(\delta_1, \delta_2) = \frac{1}{8\pi G}\frac{a_0}{\gamma\pi}\oint_S \delta_1 W \wedge \delta_2 W \tag{8.20}$$

and the curvature of W is given by

$$\mathrm{d}W = -\frac{2\pi\gamma}{a_0} 8\pi G\gamma \underline{\Sigma}^i r_i. \tag{8.21}$$

Thus, the sector of the surface phase space corresponding to any fixed set of multipoles on a type II horizon is isomorphic with the phase space of a type I horizon. Moreover, the horizon boundary condition in the type II case (when expressed in terms of W) is identical to those in the type I case. Therefore in terms of the surface connection W, one can proceed with quantization as we did before in the type I case. All the mathematics underlying the quantum horizon geometry and the state counting is the same! However, the physical meaning of symbols and constructions is different. In particular, in the type I case, W was the natural spin connection which directly described the horizon geometry and therefore the punctures where its curvature is concentrated could be directly associated with deficit angles. In the present case, it is the connection V that determines the physical horizon geometry and not W and the relation between the two involves distortion and rotation. Classically, this non-trivial information is coded in multipole moments of V. On the quantum Hilbert space, one can introduce the multipole moment operators and their eigenvalues distinguish the physical situation of interest, coded in V, from the physics of the fiducial connection W. Thus, there are non-trivial differences on issues related to interpretation. However, the counting argument of section 8.3 is unaffected by these.

To summarize, one can treat generic type II isolated horizons via the following steps: (i) construct parameters (the multipoles) which characterize these horizons in an invariant fashion macroscopically (i.e., in the classical theory); and (ii) introduce an isomorphism from the phase space of horizons of interest to that of type I horizons, which maps the physical isolated horizon condition to the horizon condition in the type I case. Together, these properties enable us to construct the quantum theory of horizon geometry and count the horizon states. *The procedure guarantees that the value of the Barbero–Immirzi parameter that reproduces the Hawking–Bekenstein formula for large black holes is the same as that used in the type I case, independent of the values of the mass and angular momentum multipoles.* The value

is thus robust. Finally, note that because this analysis incorporates arbitrary distortions, we are going well beyond the Kerr–Newman family. The method encompasses a vast class of astrophysically realistic black holes.

9. Current directions

In the last six sections, we presented a self-contained summary of the quantum geometry framework and its physical applications which have been worked out in detail. In this section, we turn to current research. There are two major thrusts: (i) recovery of low energy physics through semi-classical quantum geometries; and (ii) spin foam models, which provide a 'sum over histories' approach based on quantum geometry. Since these are frontier areas, a finished physical picture is yet to emerge. Therefore, our discussion will be briefer.

9.1. Low energy physics

A basic premise of loop quantum gravity is that there should be no background fields; everything, including spacetime geometry is dynamical and treated quantum mechanically from the start. However, of necessity, a background-independent description must use physical concepts and mathematical tools that are quite different from those normally used in low energy quantum physics which is rooted in classical, Minkowskian geometry. A major challenge, then, is to show that this low energy description does arise from the pristine, Planckian world in an appropriate sense. This challenge is now being met step by step, although one is still far from reaching the final goal.

Let us begin by listing some of the main issues and questions. Loop quantum gravity is based on *quantum geometry*, the essential discreteness of which permeates all constructions and results. The fundamental excitations are one dimensional and polymer-like. A convenient basis of states is provided by spin networks. Low energy physics, on the other hand, is based on quantum field theories which are rooted in a flat space continuum. The fundamental excitations of these fields are three dimensional, typically representing wavy undulations on the background Minkowskian geometry. The convenient Fock basis is given by specifying the occupation number in one particle states labelled by momenta and helicities. At first sight, the two frameworks seem disparate. *What then is the precise sense in which the Fock states are to arise in the low energy limit of the full theory?*

From a mathematical physics perspective, the basic variables of quantum geometry are holonomies (or Wilson loops) of the gravitational connection A along one-dimensional curves and fluxes of the conjugate momenta (the triads) E across 2-surfaces. In the final quantum theory, the connection A fails to be a well defined operator(-valued distribution); only the holonomies are well defined. In perturbative quantum field theories, by contrast, the vector potential operators are distributions, whence, *a priori*, their holonomies fail to be well-defined operators. Similarly, fluxes of electric field operators across 2-surfaces fail to be well defined on the Fock space of photons. Heuristically, then, it would appear that, even at a kinematic level, loop quantum gravity describes a 'phase' of gauge theories which is distinct from that used in electrodynamics. Since it is generally believed that distinct phases carry distinct physics, it is natural to ask: *Is the well-tested, macroscopic 'Coulomb phase' of low energy gravity compatible at all with the Planck scale discreteness of quantum geometry?* If so, in what sense? How does it emerge from loop quantum gravity?

So far these issues have been analysed through simple examples, where the focus is on constructing mathematical and conceptual tools that will be ultimately necessary for the systematic analysis of quantum fields on semi-classical states of quantum geometry [148] .

9.1.1. Quantum mechanics of particles. In non-relativistic quantum mechanics, one generally begins with the Weyl algebra generated by operators $U(\lambda) = \exp(\mathrm{i}\lambda X)$ and $V(\mu) = \exp(\mathrm{i}(\mu/\hbar)P)$ and seeks representations in which $U(\lambda)$ and $V(\mu)$ are represented by a one-parameter family of unitary operators which are weakly continuous in λ and μ. The von Neumann uniqueness theorem tells us that every irreducible representation of this algebra is isomorphic with the Schrödinger representation. Therefore, one typically develops the theory using just this representation. However, if one drops the requirement of weak continuity, say in μ, new representations become available.

Specifically, there is one in which states $\Psi(k)$ are almost periodic functions of $k = p/\hbar$. In this representation, operators $U(\lambda)$, $V(\mu)$ are unitary as desired but the self-adjoint generator of $V(\mu)$, which provides the momentum operator in the Schrödinger representation, fails to exist. The position operator X, on the other hand, does exist and is self-adjoint. Furthermore, its spectrum is discrete in the sense that all its eigenvectors are normalizable. This representation is referred to as 'polymer particle' because of its close mathematical similarities with the 'polymer' representation of the algebra generated by holonomies and electric fluxes introduced in section 4. (Indeed, the underlying mathematical framework is the same as that used in quantum cosmology in section 7, but the physical interpretations are very different.) X is analogous to the electric flux operators and its eigenstates provide us with analogues of spin network states. $V(\mu)$ is analogous to the holonomies. Just as the connection operator does not exist in quantum geometry, the generator of space translations—-the momentum operator of the Schrödinger theory—does not exist on the Hilbert space $\mathcal{H}_{\mathrm{poly}}$ of the polymer particle representation. While the absence of the standard momentum operator is alarming from the perspective of non-relativistic quantum mechanics, *heuristically* it can be thought of as arising from a fundamental discreteness of spatial geometry. However, this motivation cannot be taken too literally: non-relativistic quantum mechanics has limitations which become manifest *much* before quantum gravity discreteness can become significant.

Rather, the primary motivation in this study is mathematical: we have a simple toy model to probe the questions raised in the beginning of this section. In this analogy, Schrödinger quantum mechanics plays the role of quantum theories used in low energy physics and the main question is: Can the polymer particle framework reproduce the results of Schrödinger quantum mechanics, in spite of the fact that the two descriptions are fundamentally so different? The answer is in the affirmative and the analysis has provided some conceptual and technical insight to recover low energy physics from the Planck scale framework based on polymer geometry.

The main results [155] can be summarized as follows:

- Although the standard creation and annihilation operators fail to be well defined in $\mathcal{H}_{\mathrm{poly}}$, their exponentials *are* well defined and can be used to construct coherent states purely in the polymer framework. As one might expect of semi-classical states, they belong to $\mathrm{Cyl}^{\star}_{\mathrm{poly}}$, the analogue of $\mathrm{Cyl}^{\star}$ of quantum geometry. A key question is whether they can be regarded as semi-classical states. At first this appears to be difficult because $\mathrm{Cyl}^{\star}_{\mathrm{poly}}$ does not carry a Hermitian inner product. However, one *can* provide a meaningful criterion of semi-classicality through the notion of *shadow states* (explained in section 9.1.2) and verify that these coherent states satisfy the criterion.
- As in quantum cosmology, by introducing a length scale μ_0 which is thought of as arising from the fundamental discreteness of spatial geometry, one can define a momentum operator and the kinetic energy term in the Hamiltonian using $V(\mu)$,

$$\frac{P^2}{2m} = \frac{\hbar^2}{2m}\frac{1}{\mu_0^2}[2 - V(\mu_0) - V(-\mu_0)]$$

with $\mu_0 \ll d$, where d is the smallest length scale in the problem[36]. The Schrödinger equation then reduces to a difference equation. For the case of the harmonic oscillator, one can transform it to the well-known Mathieu equation and, using the rather large body of results on this equation, show that all energy eigenstates are non-degenerate and eigenvalues discrete, given by

$$E_n \sim (2n+1)\frac{\hbar\omega}{2} - \frac{2n^2+2+1}{16}\left(\frac{\mu_0}{d}\right)^2 \frac{\hbar\omega}{2} + O\left(\frac{\mu_0}{d}\right)^4.$$

Thus the 'polymer corrections' to the Schrödinger eigenvalues become significant only when $n \sim 10^7$! Using the notion of shadow states, one can also show that there is a precise sense in which the eigenvectors are 'close' to the Schrödinger eigenvectors. Since in the final picture one is in effect using a discrete approximation to the Schrödinger equation, it may seem 'obvious' that a close agreement with Schrödinger quantum mechanics must occur. However, the detailed analysis contains a number of subtleties and the agreement emerges only when these subtleties are handled appropriately [155]. More importantly, one does not simply begin with the Schrödinger equation and discretize it 'by hand'. Rather, one follows procedures that are natural from the 'polymer' perspective and arrive at the discrete substitute of the Schrödinger equation.

Thus, the polymer particle has turned out to be a simple toy model to illustrate how the gap between inequivalent mathematical frameworks can be bridged and how they can lead to physically equivalent results in the 'low energy regime' in spite of the deep conceptual and structural differences at a fundamental level.

9.1.2. The Maxwell field and linearized gravity. The next two models that have been studied in detail are the Maxwell field and linearized gravity in Minkowski spacetime [152–154, 156–158]. Since our goal is only to provide a bird's eye view, we will focus the main discussion on the Maxwell case and return to linearized gravity at the end.

Following the general procedures outlined in section 4 to the case when the gauge group G is $U(1)$, one can construct a 'polymer representation' of the Maxwell field (see appendix A). Polymer representations were first introduced by Buchholz and Fredenhagen in a different context. The goal is to understand its relation to the standard Fock representation of photons. Again, the viewpoint is *not* that the polymer representation provides a better physical description of photons in Minkowski space. Rather, the primary goal is to develop mathematical and conceptual tools to compare the disparate descriptions, tools which will be finally useful in understanding the relation between quantum field theories on semi-classical quantum geometries representing classical spacetimes and continuum quantum field theories in these spacetimes.

The first major difference between the polymer and the Fock representations lies in their algebras of elementary observables. In the polymer representation, these are given by holonomies $\mathbb{A}(e)$ of the Maxwell connection $\mathbb{A}$ along edges e in $M = \mathbb{R}^3$, and electric fields $\mathbb{P}(g)$ smeared by smooth 1-forms g of compact support in M. In the Fock representation, by contrast, the configuration variables $\mathbb{A}(f)$ are vector potentials smeared with smooth vector densities f of compact support; $\mathbb{A}(e)$ fail to be well defined [151]. To resolve this tension, one can proceed as follows [152, 153]. Introduce a test bi-tensor field $r_a^{a'}(x, x')$ which is a 1-form in its x dependence and a vector density in its x' dependence:

$$\mathbb{A}_a^{(r)}(x) := \int_M \mathrm{d}^3x'\, r_a^{a'}(x, x')\mathbb{A}_{a'}(x') \qquad \text{and} \qquad \mathbb{P}^{a'}_{(r)}(x') := \int_M \mathrm{d}^3x\, r_a^{a'}(x, x')\mathbb{P}^a(x). \quad (9.1)$$

[36] For a harmonic oscillator, $d = \sqrt{\hbar/m\omega}$, which , for the vibrational modes of a CO molecule is 10^{-10} cm. Since laboratory experiments show no signature of discreteness at the 10^{-17} cm scale, it is safe to take $\mu_0 \leqslant 10^{-19}$ cm.

Then, the map

$$\mathcal{I}^{\rm Fock}_{\rm poly}(r) : (\mathbb{A}(e), \mathbb{P}_{(r)}(g)) \mapsto (\mathbb{A}^{(r)}(e), \mathbb{P}(g)) \tag{9.2}$$

is an isomorphism from the Poisson–Lie algebra of the elementary observables used in the polymer representation to that used in the Fock representation. A prototype example of $r_a^{a'}(x, x')$ is given in Cartesian coordinates by

$$r_a^{a'}(x, x') = \frac{1}{r^3} \exp \frac{|x - x'|^2}{2r^2} \delta_a^{a'} \equiv f_r(x, x') \delta_a^{a'} \tag{9.3}$$

with $r > 0$, for which $\mathbb{A}^{(r)}(e)$ is simply the holonomy around a 'thickening' of the edge e. For simplicity, we will restrict ourselves to this specific choice in what follows.

Using isomorphisms $\mathcal{I}^{\rm Fock}_{\rm poly}$, one can pass back and forth between the polymer and the Fock descriptions. Specifically, the image of the Fock vacuum can be shown to be the following element of $\mathrm{Cyl}^\star_{\rm Max}$ [152, 153],

$$(V| = \sum_{\alpha, \vec{n}} \exp \left[-\frac{\hbar}{2} \sum_{IJ} G_{IJ} n_I n_J \right] (F_{\alpha, \vec{n}}| \, . \tag{9.4}$$

where $(F_{\alpha,\vec{n}}|$, called flux network states, constitute a basis in $\mathrm{Cyl}^\star_{\rm Max}$ and are analogous to the spin network states in $\mathrm{Cyl}^\star$ (see appendix A). These states do not have any knowledge of the underlying Minkowskian geometry. This information is neatly coded in the matrix G_{IJ} associated with the edges of the graph α, given by

$$G_{IJ} = \int_{e_I} \mathrm{d}t \, \dot{e}^a_J(t) \int_{e_J} \mathrm{d}t' \dot{e_I}^b(t') \int \mathrm{d}^3x \, q_{ab}(x) [f_r(x, e_I(t)) |\Delta|^{-\frac{1}{2}} f(x, e_J(t'))] \tag{9.5}$$

where q_{ab} is the flat Euclidean metric and Δ its Laplacian. A key insight of Varadarajan [152] was to note that, as in the case of the polymer particle, *one can single out this state directly in the polymer representation by invoking Poincaré invariance*, without any reference to the Fock space.

Similarly, one can directly locate in $\mathrm{Cyl}^\star_{\rm Max}$ all coherent states, i.e., all eigenstates of the (exponentiated) annihilation operators. Let us denote by $\big(C_{(\mathbb{A}^0, \mathbb{P}_0)}\big|$ the state peaked at classical fields $(\mathbb{A}^0, \mathbb{P}_0)$. Given a graph α, one can show that the restriction of the action of $\big(C_{(\mathbb{A}^0, \mathbb{P}_0)}\big|$ to cylindrical functions associated with α is fully encoded in a state $C_\alpha^{(\mathbb{A}^0, \mathbb{P}_0)}$ in the Hilbert space $(\mathcal{H}_{\rm Max})_\alpha$,

$$(C_{(\mathbb{A}^0, \mathbb{P}_0)}|\Psi_\alpha\rangle = \int_{\bar{\mathcal{A}}_{\rm Max}} \mathrm{d}\mu_0 \big[C_\alpha^{(\mathbb{A}^0, \mathbb{P}_0)}(\bar{\mathbb{A}})\big]^\star \Psi_\alpha(\bar{\mathbb{A}}) \tag{9.6}$$

for all cylindrical functions Ψ_α associated with the graph α . The states $C_\alpha^{(\mathbb{A}^0, \mathbb{P}_0)}(\bar{\mathbb{A}})$ in $(\mathcal{H}_{\rm Max})_\alpha$ are referred to as *shadows* of the element $\big(C_{(\mathbb{A}^0, \mathbb{P}_0)}\big| \in \mathrm{Cyl}^\star$ on graphs α. Note that the set of all shadows captures the full information in $\big(C_{(\mathbb{A}^0, \mathbb{P}_0)}\big|$. By analysing shadows on *sufficiently refined graphs*, one can introduce criteria to test if a given element of $\mathrm{Cyl}^\star_{\rm Max}$ represents a semi-classical state [155, 156]. The states $\big(C_{(\mathbb{A}^0, \mathbb{P}_0)}\big|$ do satisfy this criterion and can therefore be regarded as semi-classical in the polymer framework. Finally, using the isomorphism $\mathcal{I}^{\rm Fock}_{\rm poly}$ one can check that these states are the images of the Fock coherent states. To summarize, although the polymer representation is inequivalent to the Fock, it is possible to single out and analyse the 'correct' semi-classical states of the quantum Maxwell field directly in the polymer framework [156].

For Maxwell fields, the Fock representation is compatible with only the Coulomb phase: the vacuum expectation value of (regularized) Wilson loops goes as the exponential of the perimeter and one can read off the Coulomb potential from the sub-leading term in the

exponent. It turns out that one can translate Wilson's criterion as a condition on the overlap of certain coherent states defined by the type of loops used in the original criterion. All these considerations go through also for the linearized gravitational field in Minkowski spacetime [157, 158]. Moreover, the reformulation of the Wilson criterion provides a means of testing whether candidate semi-classical states of the *full* theory, approximating Minkowski spacetime and fluctuations thereon, are compatible with the Coulomb phase; i.e., if, in a suitable limit, the gravitational force between two particles will be given by the Coulomb law [158]. Physically, this is a key constraint on the viability of proposed semi-classical states.

9.1.3. Quantum geometry. The experience gained from simpler models is currently being used to construct semi-classical states of quantum geometry peaked at initial data corresponding to physically interesting spacetimes. In particular, are there 'preferred' semi-classical states peaked at such classical spacetimes, analogous to the coherent states of photons and gravitons in Minkowski spacetime?

The early work [142, 143] focused on constructing states which are peaked at a given spatial triad E^a_i. However, the mathematical precision was low and, moreover, the analysis ignored connections altogether. The challenge of constructing states which are peaked at given values of a set of observables constructed from both the triads and the connection was taken up in [144, 146, 147]. In particular, a detailed mathematical framework developed in the series of papers [146, 147] focused on observables associated with a given graph: holonomies of edges of the graph and fluxes of triads across certain surfaces 'dual' to the edges. This work led to states in Cyl which are sharply peaked at given values of these observables. However, this set of observables is too small from physical considerations and these states do not have the 'non-local' correlations which are the hallmark of semi-classical states in Minkowskian physics. Nonetheless, this analysis introduced a number of mathematical techniques which continue to be useful in the current investigations.

As a prelude to current research directions, let us begin by recasting the construction of the familiar coherent states in a form that is suitable for generalization. For a harmonic oscillator (or for free fields in Minkowski spacetimes) coherent states can be constructed using heat kernel methods on the configuration space. In this procedure, one starts by selecting a suitable, positive function F on the phase space which is quadratic in momenta. For the harmonic oscillator, this can be taken to be simply the kinetic energy, $F = \vec{P} \cdot \vec{P}$. By rescaling the quantum analogue of this function with suitable constants, one obtains the (negative definite) *Laplacian* Δ. The associated heat kernel provides a smoothing operator which maps the generalized eigenstates of the configuration operator to coherent states. For the oscillator, the coherent state $C_{\vec{x}_0,0}$, peaked at $\vec{x} = \vec{x}_0$ and $\vec{p} = 0$ is given by

$$C_{(\vec{x}_0,\vec{p}=0)}(x) = [\exp(t\Delta)]\delta(\vec{x}, \vec{x}_0) \tag{9.7}$$

where t determines the width of the Gaussian. (t has physical dimensions (length)^2. The value $t = \hbar/m\omega$ yields the standard coherent states.) A general coherent state $C_{\vec{x}_0,\vec{p}_0}$ is obtained simply by taking the analytical continuation of this state with respect to $\vec{x}_0$,

$$C_{\vec{x}_0,\vec{p}_0}(x) = ([\exp(t\Delta)]\delta(\vec{x}, \vec{x}_0))_{\vec{x}_0 \mapsto \vec{z}_0} \tag{9.8}$$

where $\vec{z}_0 = \vec{x}_0 + (\mathrm{i}/\hbar)p_0$. Hall [145] generalized this construction for the case when the configuration space is a compact Lie group. Let us consider the example of a free particle moving on the group manifold $SU(2)$ (see section 4.1). We can again use for F the kinetic energy term: $F = \eta^{ij} p_i p_j$ where as before η^{ij} is the Cartan–Killing metric on $SU(2)$. One can use the natural isomorphism between the complexification of and the cotangent bundle over

$SU(2)$ to label the points in the phase space by elements $g^{\mathbb{C}}$ of $\mathbb{C}SU(2)$. Then, the coherent states $C_{g_0^{\mathbb{C}}}$ peaked at the point $g_0^{\mathbb{C}}$ of the phase space is given by

$$C_{g_0^{\mathbb{C}}}(g) = ([\exp(t\Delta)]\delta(g, g_0))_{g_0 \mapsto g_0^{\mathbb{C}}}. \tag{9.9}$$

These states are sharply peaked at the phase space point $g_0^{\mathbb{C}}$ [146, 147]. The generalization of Hall's procedure to quantum theories of connection on a graph (discussed in section 4.2) is straightforward since now the configuration space is isomorphic to $[SU(2)]^n$, where n is the number of edges of the graph.

For theories of connections in the continuum, one can again follow the same procedure. For the Maxwell theory in Minkowski spacetime discussed in section 9.1.2, in the polymer picture one can proceed as follows. First, given a graph α, one can set $F = G^{IK} p_I p_K$ where p_I denotes the momentum vector in the cotangent bundle over $U(1)$ associated with the Ith edge. Then, the Laplacian on the Hilbert space $(\mathcal{H}_{\rm Max})_\alpha$ is given simply by $\Delta_\alpha^{\rm Max} = -(\hbar/2)\sum_{I,K}[G^{IK} J_I J_K]$. Interestingly, this family $\Delta_\alpha^{\rm Max}$ of operators is *consistent* and leads to a negative definite, self-adjoint operator $\Delta^{\rm Max}$ on the full Hilbert space $\mathcal{H}_{\rm Max}$. This Laplacian can now be used to define coherent states. The result is precisely the coherent states in $\mathrm{Cyl}^\star_{\rm Max}$ discussed in section 9.1.2:

$$(C_{\mathbb{A}^0, \mathbb{P}_0}|\Psi\rangle = \int_{\bar{\mathcal{A}}} \mathrm{d}\mu_0 \big(\mathrm{e}^{\Delta^{\rm Max}} \delta(\bar{\mathbb{A}}, \mathbb{A}_0)|_{\mathbb{A}_0 \mapsto \mathbb{A}_{\mathbb{C}}^0}\big)^\star \Psi(\bar{\mathbb{A}}) \tag{9.10}$$

for all cylindrical functions Ψ, where $\mathbb{A}_{\mathbb{C}}^0 = \mathbb{A}^0 - \mathrm{i}|\Delta|^{-1/2}\mathbb{P}_0$. In particular $(V|$, the image of the Fock vacuum in $\mathrm{Cyl}^\star_{\rm Max}$, is obtained by this procedure simply by setting $\mathbb{A}_0 = \mathbb{P}_0 = 0$. These Laplacians and the corresponding coherent states belong to a general framework discussed in [46].

This procedure can be naturally extended to quantum geometry to define a candidate semi-classical state $(M|$ corresponding to the Minkowski spacetime, i.e., to the point of the phase space represented by $(A = 0, E = E^0)$ where E^0 is a flat triad. Given any graph α, one can define a Laplacian operator Δ_α on the quantum geometry Hilbert space $\mathcal{H}_\alpha$:

$$\Delta_\alpha = -\frac{\hbar}{8}\sum_{I,K} G_{IK}\eta^{ik} J_i^I J_k^K. \tag{9.11}$$

Again, this set of operators is consistent and thus defines a negative definite, self-adjoint operator Δ on the full quantum geometry Hilbert space. The desired state $(M| \in \mathrm{Cyl}^\star$ can now be defined using the heat kernel defined by this Laplacian,

$$(M|\Psi\rangle = \int_{\bar{\mathcal{A}}} \mathrm{d}\mu_0 \big(e^{\Delta}\delta(\bar{A}, A_0)|_{A_0 \mapsto A_{\mathbb{C}}^0}\big)^\star \Psi(\bar{A}) \tag{9.12}$$

for all $\Psi \in \mathrm{Cyl}$, where $A_0^{\mathbb{C}} = -\mathrm{i}bE^0$ and b is a constant with dimensions of inverse length [153]. Note, however, that the state is defined simply by analogy with the simpler systems. So far, its structure has not been analysed in any detail and there is no *a priori* guarantee that this is indeed a semi-classical state, i.e., that its shadows on sufficiently refined graphs are sharply peaked at the point $(A = 0, E = E_0)$ of the gravitational phase space. (Notions from statistical geometry [150] are likely to play an important role in selecting the appropriate family of graphs.) The 'Coulomb phase criterion' is also yet to be applied.

Thiemann has developed a systematic framework to extend this procedure to introduce semi-classical states by considering more general functions F, leading to heat kernels based on operators which are more general than Laplacians [149]. Thus, rather powerful tools are now available to explore the semi-classical regime. However, compelling candidate states are yet to emerge. Finally, the emphasis in this work is on constructing states which are peaked at

points on the constraint surface of the classical phase space. These are kinematical states; as simple examples show, these states will not solve quantum constraints. Indeed, semi-classical solutions to the quantum constraints would be peaked at points of the reduced phase space, i.e., roughly, on equivalence classes of 4-metrics where two are equivalent if they are related by a diffeomorphism. To make contact with low energy physics, what we need are states peaked at *individual classical spacetimes*, whence it is the kinematical semi-classical states considered here which are more directly relevant. The relation between the two is being explored systematically. The overall picture can be summarized as follows: (i) the kinematical states can be regarded as 'gauge fixed versions' of the semi-classical solutions to constraints; and (ii) the expectation values and fluctuations of Dirac observables agree in an appropriate sense on the two sets.

9.2. Spin foams

Spin foams can be thought of as histories traced out by 'time evolution' of spin networks and provide a path integral approach to quantum dynamics. Since an entire review article devoted to spin foams has appeared recently [10], our discussion will be very brief.

In the gravitational context, the path integral can play two roles. First, as in standard quantum field theories, it can be used to compute 'transition amplitudes'. However outside, say, perturbation theory about a background spacetime, there still remain unresolved conceptual questions about the physical meaning of such amplitudes. The second role is 'cleaner': as in the Euclidean approach of Hawking and others, it can be considered as a device to extract physical states, i.e., solutions to all the quantum constraint equations. In this role as an *extractor*, it can shed new light on the quantum Hamiltonian constraint and on the issue of finding a physical inner product on the space of solutions to all constraints.

The well-defined quantum kinematics of sections 4 and 5 has motivated specific proposals for the definition of path integrals, often called 'state sum models'. Perhaps the most successful of these is the Barrett–Crane model and its various modifications. At the classical level, one regards general relativity as a topological field theory, called the BF theory, *supplemented with an algebraic constraint*. The BF theory is itself a four-dimensional generalization of the three-dimensional Chern–Simons theory mentioned in section 8.2 and has been investigated in detail in the mathematical physics literature. However, the role of the additional constraint is very important. Indeed, BF theory has no local degrees of freedom; it is the extra constraint that reduces the huge gauge freedom, thereby recovering the local degrees of freedom of general relativity. The crux of the problem in quantum gravity is the appropriate incorporation of this constraint. At the classical level (modulo issues related to degenerate configurations) the constrained BF theory is equivalent to general relativity. To obtain Euclidean general relativity, one has to start with the BF theory associated with $SO(4)$ while the Lorentzian theory results if one uses $SO(3, 1)$ instead. The Barrett–Crane model and its extensions are specific proposals to define quantum geometry based path integrals for the constrained BF theory in either case.

Fix a 4-manifold $\mathcal{M}$ bounded by two 3-manifolds M_1 and M_2. Spin network states on the two boundaries can be regarded as 'initial' and 'final' quantum geometries. One can then consider histories, i.e., quantum 4-geometries, joining them. Each history is a spin foam. Each vertex of the initial spin network on M_1 'evolves' to give a one-dimensional edge in the spin foam and each edge, to give a two-dimensional face. Consequently, each face carries a spin label j. However, in the course of 'evolution' *new vertices* can appear, making the dynamics non-trivial and yielding a non-trivial amplitude for an 'initial' spin network with n_1 vertices to evolve into a 'final' spin network with n_2 vertices. For mathematical clarity as well as physical

intuition, it is convenient to group spin foams associated with the same four-dimensional graph but differing from one another in the labels, such as the spins j carried by faces. Each group is said to provide a *discretization* of the 4-manifold $\mathcal{M}$. Physically, a discretization has essentially just the topological information. The geometrical information—such as the area associated with each face—resides in the labels. This is an important difference from lattice gauge theories with a background metric, where a discretization itself determines, e.g., the edge lengths and hence how refined the lattice is.

A notable development is the discovery that the non-perturbative path integral, defined by a certain modification of the Barrett–Crane model, is equivalent to a manageable *group field theory* (GFT) in the sense specified below [166]. The GFT is a rather simple quantum field theory, defined on four copies of the underlying group; $SL(2, C)$ in the case of Lorentzian gravity and Spin(4) in the case of Euclidean. (Note that these are just double covers of the Lorentz group and the rotation group of Euclidean 4-space.) Thus GFTs live in high dimensions. The action has a 'free part' and an interaction term with a coupling constant λ. But the free part is non-standard and does not have the familiar kinetic term, whence the usual non-renormalizability arguments for higher dimensional, interacting theories do not apply. In fact, the first key recent result is that *this GFT is finite order by order in the Feynman perturbation expansion*. The second key result is $A_{\rm BC}(n) = A_{\rm GFT}(n)$, where $A_{\rm BC}(n)$ is the modified Barrett–Crane amplitude obtained by summing over all geometries (i.e., spin labels j) for a fixed discretization and $A_{\rm GFT}(n)$ is the coefficient of λ^n in the Feynman expansion of the GFT. Together, the two results imply that, in this approach to quantum gravity, *sum over geometries for a fixed discrete topology is finite*. This is a highly non-trivial result because, on each face, the sum over j ranges from zero to infinity; there is no cut-off[37].

However, many open issues remain. First, in the specific proposal of Perez and others, convergence is achieved at a price: the integral is dominated by 'degenerate' geometries described by spin foams where all the spins labelling faces are zero except for 'islands' of higher spin [167]. Second, in any of the finite models, it is not clear if there is a direct physical interpretation, *in gravitational terms*, of the specific amplitudes (associated with 2-faces and tetrahedra) that lead to a suppression of divergences. More importantly, while many of these developments are very interesting from a mathematical physics perspective, their significance to quantum gravity is less clear. Physical issues such as gauge fixing in the path integral are not fully understood in $3+1$ dimensions [171]. (However, recently there has been notable progress in $2+1$ dimensions [172, 173].) Finally, the discrete topology is fixed in most of this work and the issue of summing over all topologies, or a substitute thereof, remains largely unexplored. However, this is a very active area of research and the hope is that the current investigations will soon yield a sufficient intuition and control on mathematical issues to enable one to analyse in detail the deeper, physical problems. In particular, it is likely that a judicious combination of methods from the canonical treatment of the Hamiltonian constraint and spin foam models will lead to significant progress in both areas.

10. Outlook

Loop quantum gravity is a non-perturbative, background-independent approach to the problem of unification of general relativity and quantum physics. In the last nine sections, we gave a self-contained account of the core developments in this approach and then summarized the most important physical applications of the framework. However, due to space limitation we

[37] Perez's Euclidean result has the same 'flavour' as the evidence found by Lauscher, Reuter, Percacci, Perini and others [169, 170] for non-perturbative renormalizability of four-dimensional Euclidean quantum general relativity (stemming from the existence of a non-trivial fixed point).

had to leave out several interesting developments, particularly at the forefront of the field. In this section, we will discuss some of them briefly and outline a few open issues.

Quantum geometry. As mentioned in section 1 the necessity of a quantum theory of geometry was strongly motivated by the fact that, in general relativity, gravity is coded in spacetime geometry. However, the quantum geometry framework itself is more general and could be used for background-independent quantization of other theories as well. For example, in two spacetime dimensions, Yang–Mills theory requires only a background volume element, not a metric. Since the classical theory is invariant under all volume preserving diffeomorphisms, it is natural to quantize it in a way that this symmetry is manifest at every step. Quantum geometry techniques have been used to carry out this quantization and this construction has certain advantages over others [174]. Similarly, in the standard treatments of bosonic string theory, one fixes only a conformal metric on the world sheet. In two dimensions, the group of conformal isometries is an infinite-dimensional subgroup of the diffeomorphism group and one can again use the standard techniques developed in section 4 to carry out a quantization in which this symmetry is manifest [175].

A recent mathematical development is the natural emergence of quantum groups from quantum geometry considerations [176]. Suppose for a moment that quantum groups had not yet been invented and one was trying to extend the construction of Cyl, introduced in section 4, to the most general setting possible, e.g. to obtain mathematically viable generalizations of quantum gauge theories. Then, one would have naturally discovered that Cyl can be replaced by a *non-commutative $C^\star$-algebra which has precisely the same structure as a quantum group*! This is a fascinating result which brings out the naturalness of constructions underlying quantum geometry.

From conceptual considerations, an important issue is the *physical* significance of discreteness of eigenvalues of geometric operators (see, e.g., [72]). Recall first that in the classical theory differential geometry simply provides us with formulae to compute areas of surfaces and volumes of regions in a Riemannian manifold. To turn these quantities into physical observables of general relativity, one has to define the surfaces and regions *operationally*, e.g. using matter fields. Once this is done, one can simply calculate values of these observables using formulae supplied by differential geometry. The situation is the same in quantum theory. For instance, the area of the isolated horizon is a Dirac observable in the classical theory and the application of the quantum geometry area formula to *this* surface leads to physical results. In $2+1$ dimensions, point particles have recently been incorporated and the physical distance between them is again a Dirac observable [173]. When used in this context, the spectrum of the length operator has direct physical meaning. In all these situations, the operators and their eigenvalues correspond to the 'proper' lengths, areas and volumes of physical objects, measured in the rest frames. Finally, sometimes questions are raised about compatibility between discreteness of these eigenvalues and Lorentz invariance. There is no tension whatsoever [177]: it suffices to recall that discreteness of eigenvalues of the angular momentum operator $\hat{J}_z$ of non-relativistic quantum mechanics is perfectly compatible with the rotational invariance of that theory.

Quantum Einstein's equations. The challenge of quantum dynamics in the full theory is to find solutions to the quantum constraint equations and endow these physical states with the structure of an appropriate Hilbert space. The general consensus in the loop quantum gravity community is that while the situation is well understood for Gauss and diffeomorphism constraints, it is very far from being definitive for the scalar (i.e., the Hamiltonian) constraint. It *is* non-trivial that well-defined candidate operators representing the scalar constraint exist on the space $\mathcal{H}_{\rm diff}$ of solutions to the Gauss and diffeomorphism constraints. However, as

section 6.3 shows there is a host of ambiguities and none of the candidate operators has been shown to lead to a 'sufficient number of' semi-classical states in $3+1$ dimensions. A second important open issue is to find restrictions on matter fields and their couplings to gravity for which this non-perturbative quantization can be carried out to a satisfactory conclusion. In the renormalization group approach, for example, the situation is as follows. There is significant evidence for a non-trivial fixed point for pure gravity in four dimensions [169] but when matter sources are included it continues to exist only when the matter content and couplings are suitably restricted. For scalar fields in particular, Percacci and Perini [170] have found that polynomial couplings (beyond the quadratic term in the action) are ruled out, an intriguing result that may 'explain' the triviality of such theories in Minkowski spacetimes. Are there similar constraints coming from loop quantum gravity?

To address these core issues, at least four different approaches are being followed. The first, and the closest to ideas discussed in section 6.3, is the 'master constraint program' recently introduced by Thiemann [178]. The idea here is to avoid using an infinite number of constraints $\mathcal{C}(N)$, each smeared by a lapse function N. Instead, one squares the integrand $C(x)$ itself in an appropriate sense and then integrates it on the 3-manifold M. In simple examples, this procedure leads to physically viable quantum theories [179]. In the gravitational case, however, the procedure does not seem to remove any of the ambiguities. Rather, its principal strength lies in its potential to resolve the difficult issue of finding the physically appropriate scalar product on physical states. The general philosophy is similar to that advocated by Klauder [180] over the years in his very interesting approach based on coherent states. However, there are two key differences. First, Klauder seeks solutions to constraints in the original, kinematical Hilbert space rather than in a larger space such as $\mathrm{Cyl}^\star$. Consequently, when zero is in the continuous part of the spectrum of constraint operators his physical states are only approximately annihilated by the constraints. Second, in Klauder's proposal *all* constraints are to be imposed in this manner. In loop quantum gravity, this does not seem to be feasible for several important technical reasons; one needs to first solve the Gauss and the diffeomorphism constraint and work on the Hilbert space $\mathcal{H}_{\mathrm{diff}}$. Indeed, to our knowledge, the proposal has not been implemented in sufficient detail to know if the original strategy can be employed to solve the diffeomorphism constraint rigorously, even by itself. But the program has a key advantage that, since it is based on coherent states, the semi-classical sector can be readily located. A cross-fertilization of this program and loop quantum gravity is likely to be fruitful in the analysis of low energy physics.

A second approach to quantum scalar constraint is due to Gambini, Pullin and their collaborators [181]. It builds on their earlier extensive work [4] on the interplay between quantum gravity and knot theory. The more recent developments use the relatively new invariants of *intersecting* knots discovered by Vassiliev. This is a novel approach which furthermore has the potential of enhancing the relation between topological field theories and quantum gravity. As our knowledge of invariants of intersecting knots deepens, this approach is likely to provide increasingly significant insights. In particular, it has the potential of leading to a formulation of quantum gravity which does not refer even to a background manifold (see footnote 5). The third approach comes from spin foam models [10] discussed briefly in section 9.2. Here, amplitudes used in the path integrals can be used to restrict the choice of the scalar constraint operator in the canonical theory. This is a promising direction and the detailed analysis of restrictions is already in progress in $2+1$ dimensions [173]. In the fourth approach, also due to Gambini and Pullin, one first constructs consistent discrete theories at the classical level and then quantizes them [182]. In this program, there are no constraints; they are solved to find lapse and shift fields. It has already been applied successfully to gauge theories and certain cosmological models. An

added bonus here is that one can revive a certain proposal made by Page and Wootters to address the difficult issues of interpretation of quantum mechanics which become especially acute in quantum cosmology, and more generally in the absence of a background physical geometry.

Applications. As we saw in sections 7 and 8, loop quantum gravity has resolved some of the long-standing physical problems of quantum gravity. As in other approaches to black-hole entropy [183–186], concrete progress could be made because the constructions do not require detailed knowledge of how quantum dynamics is implemented in the *full* quantum theory. Recently, the first law of black-hole mechanics has been extended to fully dynamical situations [187]. Its form suggests that the entropy is given by the area of the dynamical horizon. Can the quantum entropy calculation be extended to these non-equilibrium situations? This may even provide an input to non-equilibrium statistical mechanics where the notion of entropy is still rather poorly understood.

In quantum cosmology, there is ongoing work on obtaining 'effective field equations' which incorporate quantum corrections [117, 188, 189]. Quantum geometry effects significantly modify the effective field equations which in turn lead to new physics in the early universe. In particular, not only is the initial singularity resolved but the (Belinski–Khalatnikov–Lifschitz-type) chaotic behaviour predicted by classical general relativity and supergravity also disappears! As explained in [189], this is to be expected on rather general grounds if the underlying geometry exhibits quantum discreteness because even in the classical theory chaos disappears if the theory is truncated at any smallest, non-zero volume. There are also less drastic but interesting modifications of the inflationary scenario with potentially observable consequences [188, 190]. While the technical steps used in these analyses of effective equations are not as clean as those of section 7, it is encouraging that loop quantum cosmology is already yielding some phenomenological results.

As explained in section 9.1, a frontier area of research is contact with low energy physics. Here, a number of fascinating challenges appear to be within reach. Fock states have been isolated in the polymer framework [152, 153, 160] and elements of quantum field theory on quantum geometry have been introduced [148]. These developments lead to concrete questions. For example, in quantum field theory in flat spacetimes, the Hamiltonian and other operators are regularized through normal ordering. For quantum field theory on quantum geometry, on the other hand, the Hamiltonians are expected to be manifestly finite (see, e.g., appendix A). Can one then show that, in a suitable approximation, normal ordered operators in the Minkowski continuum arise naturally from these finite operators? Can one 'explain' why Hadamard states of quantum field theory in curved spacetimes are special? These issues also provide valuable hints for construction of viable semi-classical states of quantum geometry. The final and much more difficult challenge is to 'explain' why perturbative quantum general relativity fails if the theory exists non-perturbatively. As explained in section 1, heuristically the failure can be traced back to the insistence that the continuum spacetime geometry is a good approximation even below the Planck scale. But a more detailed answer is needed. Is it because, as recent developments in Euclidean quantum gravity indicate [169, 170], the renormalization group is a non-trivial fixed point?

Finally, there is the issue of unification. At a kinematical level, there is already a unification because the quantum configuration space of general relativity is the same as in gauge theories which govern the strong and electro-weak interactions. But the non-trivial issue is that of dynamics. We will conclude with a speculation. One possibility is to use the 'emergent phenomena' scenario where new degrees of freedom or particles, which were not

in the initial Lagrangian, emerge when one considers excitations of a non-trivial vacuum. For example, one can begin with solids and arrive at phonons; start with superfluids and find rotons; consider superconductors and discover Cooper pairs. In loop quantum gravity, the micro-state representing Minkowski spacetime will have a highly non-trivial Planck scale structure. The basic entities are one dimensional and polymer-like. Even in the absence of a detailed theory, one can tell that the fluctuations of these one-dimensional entities will correspond not only to gravitons but also to other particles, including a spin-1 particle, a scalar and an anti-symmetric tensor. These 'emergent states' are likely to play an important role in Minkowskian physics derived from loop quantum gravity. A detailed study of these excitations may well lead to interesting dynamics that includes not only gravity but also a select family of non-gravitational fields.

Acknowledgments

Discussions with a large number of colleagues have added clarity and precision to this review. Among them, we would especially like to thank J Baez, J Barrett, M Bojowald, D Buchholz, A Corichi, A Ghosh, S Fairhurst, C Fleischhack, K Fredenhagen, L Freidel, R Gambini, J Hartle, G Horowitz, J Jain, T Jacobson, K Krasnov, D Lüst, D Maison, P Majumdar, D Marolf, J Mourão, H Nicolai, K Noui, A Okołów, R Penrose, A Perez, J Pullin, C Rovelli, J Samuel, H Sahlmann, D Sudarsky, T Thiemann, C van den Broeck, M Varadarajan, J Wisniewski, J Willis, W Unruh, R Wald and J A Zapata. This work was supported in part by the NSF grant PHY 0090091, the KBN grant 2 P03B 12724, the Alexander von Humboldt foundation and the Eberly research funds of The Pennsylvania State University.

Appendix A. Inclusion of matter fields: the Einstein–Maxwell theory

In section 5, to bring out the main ideas we simplified the discussion of dynamics by ignoring matter fields. Inclusion of these fields does not require a major modification of the underlying framework. In this appendix, we will illustrate the procedure using Einstein–Maxwell theory.

A.1. Classical framework

The point of departure for canonical quantization is again a Hamiltonian framework. One can easily repeat the procedure used in section 2 by carrying out a 3 + 1 decomposition also of the Maxwell action. The phase space now consists of two pairs of canonically conjugate fields (A, P) describing geometry and $(\mathbb{A}, \mathbb{P})$ describing the Maxwell field, where $\mathbb{A}$ is our Maxwell vector potential and $\mathbb{P}$, our Maxwell electric field. As usual the only non-vanishing Poisson bracket in the Maxwell sector is

$$\{\mathbb{A}_a(x), \mathbb{P}^b(y)\} = \delta_a^b \delta(x, y). \tag{A1}$$

As in the geometrical sector, the basic configuration variables are taken to be holonomies $\mathbb{A}(e) := \exp(-\mathrm{i}\int_e \mathbb{A})$. However, because the Maxwell gauge group $U(1)$ is Abelian, it turns out that the electric field $\mathbb{P}^a$ can be smeared either along 2-surfaces (as was done for the gravitational P^a in section 4.3.1), or directly in three dimensions. It is more convenient to use three-dimensional smearing and set $\mathbb{P}(g) := \int_M \mathrm{d}^3x\, g_a(x)\mathbb{P}^a(x)$ for all test 1-forms g_a on M.

The Poisson bracket between these elementary variables is given by

$$\{\mathbb{A}(e), \mathbb{P}(g)\} = -\mathrm{i}\left(\int_e g\right)\mathbb{A}(e). \tag{A2}$$

Thus the Poisson algebra of elementary variables is closed as needed[38].

As in section 2, one can obtain the Hamiltonian through a Legendre transform. As expected, the total Hamiltonian density h_{EM} is a sum of constraints:

$$h_{\mathrm{EM}} = N(C + \mathbf{C}) + N^a(C_a + \mathbf{C}_a) + \omega_t^i G_i + \mathbb{A}_t\mathbf{G}, \tag{A3}$$

where the lapse N and shift N^a are the same as in the gravitational sector (see (2.23)); $\mathbb{A}_t$ is a freely specifiable function, the Lagrange multiplier for the Maxwell–Gauss constraint $\mathbf{G} = D_a\mathbb{P}^a$; and $\mathbf{C}$ and $\mathbf{C}_a$ are functionals of P, $\mathbb{A}$, $\mathbb{P}$, representing the Maxwell energy and momentum density, respectively. Specifically, the electromagnetic contribution to the scalar constraint is

$$\mathcal{C}^{\mathrm{Max}}(N) = \frac{1}{8\pi}\int \mathrm{d}^3x\, N(x)\frac{q_{ab}(x)}{\sqrt{\det q(x)}}(\mathbb{P}^a(x)\mathbb{P}^b(x) + \mathbb{B}^a(x)\mathbb{B}^b(x)), \tag{A4}$$

and

$$\mathbb{B}^a = \tfrac{1}{2}\eta^{abc}(\partial_a\mathbb{A}_b - \partial_b\mathbb{A}_a) \tag{A5}$$

is the magnetic vector density. In comparison with the Hamiltonian of the Maxwell field in Minkowski spacetime, the presence of the inverse square root of q may seem surprising. Note, however, that the electric and magnetic fields naturally carry density weight 1, whence this factor is quite essential. In Minkowski space, the background metric is implicitly used to remove the density weight.

A.2. Quantum kinematics

One can just use the procedure of section 4.3 to carry out quantization using $G = SU(2) \times U(1)$, where $SU(2)$ refers to geometry and $U(1)$ to the Maxwell field. The kinematical Hilbert space of the Einstein–Maxwell theory is given by

$$\mathcal{H}_{\mathrm{EM}} = \mathcal{H} \otimes \mathcal{H}_{\mathrm{Max}} \tag{A6}$$

where $\mathcal{H}$ is the Hilbert space of states of the quantum geometry of section 4.3.2 and $\mathcal{H}_{\mathrm{Max}}$ is the corresponding Hilbert space for the case $G = U(1)$.

Since we discussed the structure of $\mathcal{H}$ and of the operators thereon in detail in sections 4.3 and 5, let us focus just on the Maxwell sector. Convenient orthonormal basis states $F_{\alpha,\vec{n}}$, called *flux networks*, in $\mathcal{H}_{\mathrm{Max}}$ can be constructed as follows. Given a graph α, assign an orientation to the edges $(e_1, \ldots, e_N)$, label them by integers $(n_1, \ldots, n_N)$ and set

$$F_{\alpha,\vec{n}}(\mathbb{A}) = [\mathbb{A}(e_1)]^{n_1} \ldots [\mathbb{A}(e_n)]^{n_N}. \tag{A7}$$

Note that if the orientation of an edge e_I is reversed, the state is unchanged if n_I is replaced by $-n_I$.

The Poisson bracket relation (A2) leads to the definition of the smeared electric operator $\hat{\mathbb{P}}(g)$:

$$\hat{\mathbb{P}}(g)\Psi = \mathrm{i}\hbar\{\mathbb{P}(g), \Psi\} \tag{A8}$$

[38] The physical dimensions of the Maxwell variables are the same as those of their gravitational analogues. Thus, $[\mathbb{A}] = L^{-1}$ and $[\mathbb{P}] = ML^{-1}$. The magnetic potential $\tilde{\mathbb{A}}$ and electric field $\tilde{\mathbf{E}}$ of classical electrodynamics are given by $\tilde{\mathbb{A}} = \mathbf{e}\mathbb{A}$ and $\tilde{\mathbf{E}} = (1/\mathbf{e})\mathbb{P}$. In quantum electrodynamics, the holonomy is generally written as $\exp(-\mathrm{i}\,\mathbf{e}/\hbar)\int \mathbb{A}'$. Therefore, the vector potential $\mathbb{A}'$ used there is given by $\mathbb{A}' = (\hbar/\mathbf{e})\mathbb{A}$.

capturing the expectation that $\hat{\mathbb{P}}(x)$ should be represented by $i\hbar\delta/\delta\mathbb{A}(x)$. On the flux network states, the action reduces to

$$\hat{\mathbb{P}}(g)F_{\gamma,\vec{n}} = -\hbar\left(\sum_I n_I \int_{e_I} g\right) F_{\gamma,\vec{n}}. \tag{A9}$$

If the support of g has non-trivial intersection just with a single edge e_I of α, then the flux network $F_{\alpha,\vec{n}}$ is an eigenstate of $\hat{\mathbb{P}}(g)$ and the eigenvalue just measures n_I, the 'electric flux carried by the oriented edge e_I'. Thus the electric flux is quantized and each edge of the flux network $F_{\alpha,\vec{n}}$ can be thought of as carrying an integral multiple of the fundamental quantum.

A.3. The quantum constraints

As noted in appendix A.1, the Einstein–Maxwell theory again has a set of three first class constraints. The action of the Gauss constraint for the group $SU(2) \times U(1)$ naturally factors on $\mathcal{H}_{\rm EM} = \mathcal{H} \times \mathcal{H}_{\rm Max}$: $\hat{\mathcal{C}}_G^{\rm EM} = \hat{\mathcal{C}}_G \otimes \hat{\mathcal{C}}_G^{\rm Max}$, where $\hat{\mathcal{C}}_G^{\rm Max}$ is the Gauss constraint operator on the quantum geometry Hilbert space $\mathcal{H}$ and $\hat{\mathbf{G}}$ that on the Maxwell–Hilbert space. Imposition of this constraint selects the *gauge invariant subspace* of $\mathcal{H}_{\rm EM}$. The gauge invariant subspace of $\mathcal{H}$ was obtained in section 6.1. On the Maxwell Hilbert space $\mathcal{H}_{\rm Max}$, the constraint simply restricts the flux network states as follows: at each vertex the sum of the labels n_I assigned to the incoming edges is equal to the sum of the labels assigned to the outgoing edges. Note that the solution space is a *subspace* on $\mathcal{H}_{\rm EM}$.

The diffeomorphism constraint $\int \mathrm{d}^3x N^a(C_a + \mathbf{C}_a)$ is also straightforward to impose in the exponentiated version. The general procedure is the same as that of section 6.2. Again, the solutions lie in the dual $\mathrm{Cyl}_{\rm EM}^\star$ of $\mathrm{Cyl}_{\rm EM} = [\mathrm{Cyl} \otimes \mathrm{Cyl}_{\rm Max}]$ where $\mathrm{Cyl}_{\rm Max}$ is the space of the cylindrical functions of $U(1)$ connections.

Finally, we have to impose the scalar constraint. Regularization of the Einstein part $\hat{\mathcal{C}}(N)$ of the constraint was discussed in detail in section 6.3. Here we will focus just on the Maxwell part $\hat{\mathcal{C}}^{\rm Max}(N)$. We have organized the discussion so that it will serve a dual purpose. On the one hand, it will provide us the Maxwell part of the total Hamiltonian constraint that must be imposed to select the physical states of the Einstein–Maxwell theory. For this purpose, we will construct an operator which is well defined on the (gauge and) diffeomorphism invariant sector $(\mathrm{Cyl}_{\rm EM}^\star)_{\rm diff}$ of $\mathrm{Cyl}_{\rm EM}^\star$. On the other hand, in the framework of field theory in a given classical spacetime, $\mathcal{C}^{\rm Max}(N)$ can also be regarded as the *physical Hamiltonian* of the Maxwell field. Therefore, it is natural to ask if one can construct from $\mathcal{C}^{\rm Max}(N)$ a well-defined operator which will act on $(\Psi|_{\rm geo} \otimes \mathcal{F}_{\rm Max}$, where $(\Psi|_{\rm geo}$ is a given semi-classical state of quantum geometry and $\mathcal{F}_{\rm Max}$ the Fock-space of photons on this geometry. We will show that this is also possible. The result will be a Hamiltonian governing the dynamics of a test quantum Maxwell field on a fixed, semi-classical quantum geometry.

Regularization of the 3-geometry part in $\mathcal{C}^{\rm Max}(N)$. In contrast to the Maxwell parts $\mathcal{C}_G^{\rm Max}$ and $\mathcal{C}_{\rm Diff}$ of the Gauss and the vector constraints, the Maxwell part $\mathcal{C}^{\rm Max}(N)$ of the scalar constraint contains a coefficient $q_{ab}/\sqrt{q}$ that explicitly depends on geometry. We will first 'regularize' this term, i.e., express it using variables which have direct operator analogues on the quantum geometry Hilbert space. This discussion will bring out the role played by quantum geometry in regulating the quantum matter Hamiltonians. For simplicity, we will work with just the electric term; by inspection all equations of this subsection continue to hold if the electric fields are replaced by magnetic.

Consider then the term

$$\int_M \mathrm{d}^3x N(x) \frac{q_{ab}(x)}{\sqrt{\det q(x)}} \mathbb{P}^a(x)\mathbb{P}^b(x). \tag{A10}$$

We can express the metric q_{ab} using (a slight generalization of) expression (6.16) for the orthonormal co-frame e^i_a,

$$e^i_a(x) = \frac{2}{k\gamma}\{A^i_a(x), V_{\mathcal{R}}\} \tag{A11}$$

where $\mathcal{R}$ is an arbitrary open neighbourhood of x and $V_{\mathcal{R}}$ is its volume with respect to q_{ab}. The Poisson bracket is independent of $\mathcal{R}$ and for our regularization purposes, it is convenient to choose it to be the ball $\mathcal{R}_\epsilon$ of coordinate volume ϵ^3, centred at x. Denote the geometric volume of this ball (with respect to q_{ab}) by $V(x,\epsilon)$. Approximating $\sqrt{\det q}$ by $V(x,\epsilon)/\epsilon^3$ it is easy to verify

$$\frac{q_{ab}(x)}{\sqrt{\det q(x)}} = \frac{16}{k^2\gamma^2} \lim_{\epsilon\to 0} \epsilon^3 \{A^i_a(x), \sqrt{V(x,\epsilon)}\}\{A^i_b(x), \sqrt{V(x,\epsilon)}\}. \tag{A12}$$

Now, in the quantum Maxwell theory in Minkowski spacetime, the electric field becomes an operator-valued distribution whence the product of electric fields at the same point, such as that in (A10), is ill-defined. Therefore, with an eye towards quantization, let us point-split the product by introducing a 2-point smearing function $\chi_\epsilon(x,y)$:

$$\chi_\epsilon(x,y) = \begin{cases} 1, & \text{if } y \in \mathcal{R}_\epsilon \\ 0, & \text{otherwise.} \end{cases} \tag{A13}$$

Then, we obtain

$$\int \mathrm{d}^3x N(x) \frac{q_{ab}}{\sqrt{\det q}} \mathbb{P}^a(x)\mathbb{P}^b(x) = \frac{16}{k^2\gamma^2} \lim_{\epsilon\to 0} \int_M \mathrm{d}^3x N(x) \int_M \mathrm{d}^3y\, \mathbb{P}^a(x)\mathbb{P}^b(y) \times \chi_\epsilon(x,y)\{A^i_a(x), \sqrt{V(x,\epsilon)}\}\{A^i_a(y), \sqrt{V(y,\epsilon)}\}. \tag{A14}$$

Note that the point-splitting procedure requires us to set $\mathbb{P}^a(x) = (1/\epsilon^3)\int_{\mathcal{R}_\epsilon} \mathrm{d}^3y \mathbb{P}^a(y)$ but the factor $1/\epsilon^3$ in the denominator is cancelled by the factor ϵ^3 in the geometric term (A12). We will see that, thanks to point splitting, this regulated classical version has a well-defined operator analogue in the quantum theory. Had we worked in Minkowski spacetime, the geometric term $q_{ab}/\sqrt{q}$ would simply be a smooth field on M. Then, the $1/\epsilon^3$ factor required in the point-splitting procedure would have remained and led to a divergence in the limit as $\epsilon \mapsto 0$, i.e., when the regulator is removed. This divergence is now avoided because the quantum geometry operator corresponding to $\{A^i_a(x), \sqrt{V(x,\epsilon)}\}$ has a well-defined limit as ϵ tends to infinity. *In this precise sense, the quantum nature of geometry provides a natural regulator for matter Hamiltonians* [9].

Quantization of the electric part of $\mathcal{C}^{\mathrm{Max}}(N)$. We now wish to find the quantum analogue of the expression on the right-hand side of (A14). Following a strategy introduced by Thiemann [98], we will proceed in two steps. In the first, we replace the classical electric field by the corresponding operators and in the second we do the same for geometric fields[39].

Consider a flux network state $F_{\gamma,\vec{n}}$ in the Maxwell–Hilbert space. Then, using the action (A9) of smeared electric field operators, we immediately obtain the action of the regulated operator $\hat{\mathcal{C}}^\epsilon_{\mathrm{elec}}$ (see equation (A4)),

$$\hat{\mathcal{C}}^\epsilon_{\mathrm{elec}}(N) F_{\gamma,\vec{n}} = \mathbf{C}^\epsilon_{\mathrm{elec}} F_{\gamma,\vec{n}} \tag{A15}$$

[39] It is also possible to quantize simultaneously the electric and geometric fields in the constraint. However, then subtleties arise in the choice of the holonomies replacing the $SU(2)$ connection A.

where the eigenvalue $\mathbf{C}^{\epsilon}_{\mathrm{elec}}$ is given by

$$\mathbf{C}^{\epsilon}_{\mathrm{elec}}(N) = \frac{2\hbar^2}{k^2\gamma^2\pi}\int_{e_I}\mathrm{d}t\int_{e_J}\mathrm{d}t'\sum_{I,J}N(e_I(t))\chi_\epsilon(e_I(t),e_J(t'))n_I n_J \times\big\{A^i_a(e_I(t))\dot{e}^a_I(t),\sqrt{V(e_I(t),\epsilon)}\big\}\big\{A^i_b(e_J(t'))\dot{e}^b_J(t'),\sqrt{V(e_J(t'),\epsilon)}\big\}. \quad \text{(A16)}$$

The second step is now facilitated because the gravitational connection A appears only through its edges along the graph. By dividing the edges into segments of coordinate length ϵ', replacing the integrals by sums of holonomies and taking the limit $\epsilon'\mapsto 0$ first, followed by the limit $\epsilon\mapsto 0$, one obtains the action of the electric part of $\mathcal{C}^{\mathrm{Max}}(N)$ on a state $\Psi_\gamma\otimes F_{\gamma,\vec{n}}$ in the full Hilbert space $\mathcal{H}_{\mathrm{EM}}$,

$$\hat{\mathcal{C}}_{\mathrm{elec}}(N)[\Psi_\gamma\otimes F_{\gamma,\vec{n}}] = \frac{4}{k^2\gamma^2\pi}\sum_v N(v)\sum_{I,J}n_I n_J \times\mathrm{Tr}\big(\hat{A}\big(e_I^{-1}\big)[\hat{A}(e_I),\sqrt[4]{\hat{q}_v}]\hat{A}\big(e_J^{-1}\big)[\hat{A}(e_J),\sqrt[4]{\hat{q}_v}]\big)[\Psi_\gamma\otimes F_{\gamma,\vec{n}}] \quad \text{(A17)}$$

where v ranges over the vertices of the graph γ and, given v, I, J run over the labels of the edges intersecting v. Note that this action preserves each subspace $(\mathcal{H}_{\mathrm{EM}})_\gamma$ of $\mathcal{H}_{\mathrm{EM}}$; it does *not* require us to extend the graph γ. Finally, it is straightforward to check that $\hat{\mathcal{C}}_{\mathrm{elec}}$ admits a self-adjoint extension to $\mathcal{H}_{\mathrm{EM}}$.

Quantization of the magnetic part. The starting point is again expression (A14), but with $\mathbb{P}^a$ replacing $\mathbb{B}^a$. We need to define operator analogues of the magnetic field. The strategy is the same as that used in section 6.3 for curvature F_{ab} of the $SU(2)$ connection of quantum geometry: approximate the dual $\mathbf{F}_{ab}$ of the magnetic field $\mathbb{B}^a$ by holonomies around small closed loops. For this purpose, as in section 6.3.1, we again cover M with cells $\square$ (possibly with arbitrary shape) and, in every cell, introduce edges $s^I_\square$ and loops $\alpha^{IJ}_\square$. Let us label this structure by $\mathcal{T}$. The idea now is to replace the double integral on the left-hand side of (A14) by a generalized Riemann sum in which the gravitational connections A^i_a are approximated by holonomies along the edges $s^I_\square$ and the magnetic field is approximated by the holonomies along $\alpha^{IJ}_\square$. We are then led to define the approximate expression of the magnetic part of $\mathcal{C}^{\mathrm{max}}(N)$ as

$$\mathcal{C}^{(\epsilon,\mathcal{T})}_{\mathrm{mag}}(N) := T_{IJI'J'}\sum_{\square,\square'}N(x_\square)\chi_\epsilon(x_\square,x_{\square'})\,\mathrm{Tr}\big(\mathbb{A}\big(\alpha^I_\square\big)A\big(\big(s^J_\square\big)^{-1}\big)\big\{A\big(s^J_\triangle\big),\sqrt{V_\mathcal{R}}\big\} \times\mathbb{A}\big(\alpha^{I'}_{\square'}\big)A\big(\big(s^{J'}_{\square'}\big)^{-1}\big)\big\{A\big(s^{J'}_{\square'}\big),\sqrt{V'_\mathcal{R}}\big\}\big) \quad \text{(A18)}$$

which converges to the magnetic part of $\mathcal{C}^{\mathrm{Max}}(N)$ as we shrink $\mathcal{T}$ and take ϵ to zero. (Here $\mathcal{T}^{IJI'J'}$ are constants determined by the geometry of $\mathcal{T}$.) It is now straightforward to pass to the regulated operator $\hat{\mathbf{C}}^{(\epsilon,\mathcal{T})}_{\mathrm{mag}}(N)$. While this operator is well defined in $\mathrm{Cyl}_{\mathrm{EM}}$, as in the gravitational case, its limit as $\epsilon\mapsto 0$ fails to be well defined on $\mathcal{H}_{\mathrm{EM}}$. Therefore, to define the constraint operator, as in section 6.3.2, we pass to $(\mathrm{Cyl}^\star_{\mathrm{EM}})_{\mathrm{diff}}$. Given $(\Psi|\in(\mathrm{Cyl}^\star_{\mathrm{EM}})_{\mathrm{diff}}$, the procedure used in section 6.3.2 leads to the following well-defined action,

$$[(\Psi|\hat{C}_{\mathrm{mag}}(N)]|\Psi_\gamma F_{\gamma,\vec{n}}\rangle = -\hbar^{-2}T^{IJI'J'}(\Psi|\sum_v N(v) \times\mathrm{Tr}\big(\mathbb{A}\big(\alpha^I_v\big)\hat{A}\big(\big(s^J_v\big)^{-1}\big)\big[\hat{A}\big(s^J_v\big),\sqrt[4]{\hat{q}_v}\big]\mathbb{A}\big(\alpha^{I'}_v\big)\hat{A}\big(\big(s^{J'}_v\big)^{-1}\big)\big[\hat{A}\big(s^{J'}_v\big),\sqrt[4]{\hat{q}_{v'}}\big]\big)|\Psi_\gamma\rangle \quad \text{(A19)}$$

where for every vertex v of the graph γ, s^I_v and α^I_v are the edges and loops of $\mathcal{T}$ originating at v. Recall from section 6.3 that the geometric structure of $\mathcal{T}$ is such that the edges s^I_v

themselves do not appear in the final result; the graph changes only through the loops α_v^I. But the geometric operators that appear on the right-hand side do not refer to the loops α_v^I. Therefore, while the action of $\hat{\mathcal{C}}_{\rm mag}(N)$ does add new edges to the graph, the spin labels of these edges vanish; only the Maxwell flux quantum numbers $\vec{n}$ are non-trivial on these edges.

Summary. Collecting the results of the last two sections, solutions to the quantum scalar constraint are elements $(\Psi|$ of $(\mathrm{Cyl}^\star_{\rm EM})_{\rm diff}$ satisfying

$$(\Psi|[\hat{\mathcal{C}}(N) + \hat{\mathcal{C}}_{\rm elec}(N) + \hat{\mathcal{C}}_{\rm mag}(N)] = 0$$

where the action of the geometrical part $\hat{\mathcal{C}}(N)$ is as in section 6.3 and the electric and magnetic Maxwell operators are given by (A17) and (A19).

Finally, as mentioned in the beginning of appendix A.3, $\mathcal{C}^{\rm Max}(N)$ is also the Hamiltonian of the Maxwell field propagating on a fixed, static background (where the 4-metric is determined completely by the 3-metric q_{ab} and the lapse N). Can we use this operator to 'derive', *in a suitable approximation*, the quantum theory of Maxwell fields on static spacetimes? Let us use for $(\Psi|$ the tensor product $(\Psi|_{\rm geo} \otimes (\Psi|_{\rm Max}$ where $(\Psi|_{\rm geo}$ is a quantum geometry state peaked at a static spacetime and $(\Psi|_{\rm Max}$ is (the image in $\mathrm{Cyl}^\star_{\rm Max}$ of) a Fock state of photons associated with the static background. Note that these states are *not* diffeomorphism invariant. However, we can exploit the availability of a background metric and use in place of $\hat{\mathcal{C}}_{\rm mag}(N)$ the regulated operator $\hat{\mathbf{C}}^{(\epsilon,\mathcal{T})}(N)$, where the area of the loops $\alpha_\Box^I$ is given by the minimum non-zero eigenvalue of the area operator. The resulting Maxwell Hamiltonian $\mathbf{C}(N)$ has a well-defined action on $\mathrm{Cyl}^\star_{EM}$. Therefore, we can analyse the evolution of the resulting state and compare it with the standard evolution in the Fock space. An important viability criterion for this strategy to work is that the geometry part of the state does not change appreciably under the action of $\mathcal{C}^{\rm Max}(N)$. To analyse whether this condition is met, we can expand $(\Psi|_{\rm geo}$ in terms of spin network states $(s_{\gamma,\vec{j},\vec{I}}|$. It is easy to check that the action of the geometric operators in $\hat{\mathbf{C}}(N)$ leaves γ and the spin labels $\vec{j}$ of *each* of these spin network components invariant; only the intertwiners change. Therefore, it is plausible that $(\Psi|_{\rm geo}$ does not change appreciably. However, so far it is not obvious that any of the candidate semi-classical states proposed to date satisfy this condition; this issue is being investigated. Thus, considerations involving matter fields provide detailed, quantitative criteria for viability of candidate semi-classical states of quantum geometry.

Appendix B. List of symbols

$a, b, \ldots$	spatial indices for tensor fields on the 3-manifold M
$\alpha, \beta, \ldots$	spacetime indices in sections 2 and 8.1
$\alpha, \beta, \ldots$	labels for graphs on M in the rest of the sections
A_a^i	a connection 1-form on M
$A(e)$	holonomy along an edge e defined by a connection A
$\mathcal{A}$	space of smooth connections on M for a given gauge group G
$\bar{A}$	a generalized connection
$\bar{A}(e)$	holonomy along an edge e defined by a generalized connection $\bar{A}$
$\hat{A}(e)$	its corresponding quantum operator
$\bar{\mathcal{A}}$	quantum configuration space (of generalized connections)
A_S	area of a 2-surface (without boundary) S
$\hat{A}_S$	its corresponding quantum operator
$\mathbb{A}_a$	Maxwell vector potential

$\mathbb{A}(e)$	its corresponding holonomy along an edge e
$\mathbb{B}^a$	Maxwell magnetic (vector density) field
$\mathbb{C}$	the set of complex numbers
$\mathcal{C}_{\rm Diff}(\vec{N})$	diffeomorphism constraint smeared with N^a
$\hat{\mathcal{C}}_{\rm Diff}(\vec{N})$	its corresponding quantum operator
$\mathcal{C}(N)$	scalar constraint smeared with N
$\hat{\mathcal{C}}(N)$	its corresponding quantum operator
$\mathcal{C}_{\rm G}(\Lambda)$	Gauss constraint smeared with Λ^i
$\hat{\mathcal{C}}_{\rm G}(\Lambda)$	its corresponding quantum operator
$C^{(n)}$	a differentiability class
Cyl	algebra of cylindrical functions on $\mathcal{A}$
Cyl_α	algebra of the cylindrical functions defined by a graph α
$\mathrm{Cyl}^\star$	space of linear functionals on Cyl
$\mathrm{Cyl}^\star_{\rm diff}$	the image of Cyl under the diffeomorphism averaging map
Diff	group of certain diffeomorphisms of M (defined in section 4.3.5)
e	a closed-piecewise analytic edge (defined in section 4.3.1)
E^a_i	triads with density weight 1, defining the Riemannian geometry on M
ϵ^i_{jk}	structure constants of $su(2)$ (of a general $\mathfrak{g}$ in section 4)
η_{ij}	the Killing form on $su(2)$ (on a general Lie algebra $\mathfrak{g}$ in section 4)
η^{abc}	metric independent, totally skew pseudo-tensor density of weight 1 on M
η_{abc}	metric independent, totally skew pseudo-tensor density of weight -1 on M
η	diffeomorphism averaging map (defined in section 6.2)
F^i_{ab}	curvature of A^i_a
G	a compact Lie group
$\mathfrak{g}$	its Lie algebra
G	Newton's constant
γ	Barbero–Immirzi parameter
$\mathcal{H}$	kinematical Hilbert space of quantum geometry
$\mathcal{H}_\alpha$	subspace of $\mathcal{H}$ defined by cylindrical functions compatible with graph α
$\mathcal{H}'_\alpha$	subspace of $\mathcal{H}_\alpha$ used in the spin network decomposition of $\mathcal{H}$
$i, j, \ldots$	internal indices for $so(3) = su(2)$ (in section 4, for a general $\mathfrak{g}$)
$I, J, \ldots$	four-dimensional internal indices in section 2
$I, J, \ldots$	labels (e.g., for edges, punctures, etc) in sections 4–9
$\mathcal{I}_E$	map from the space of connections on a graph with n edges into G^n
$\mathcal{I}_V$	map from the space of gauge transformations on a graph with m vertices into G^m
$\hat{J}^{(v,e)}_i$	operator on Cyl_α associated with an edge e and a vertex v of α
k	8π times Newton's constant
κ	surface gravity of isolated horizons
$\kappa(S, e)$	a constant $(0, \pm - 1)$ assigned to a surface S and edge e intersecting it
$\ell_{\rm Pl}$	Planck length
L^2	space of square integrable functions
M	three-dimensional ('spatial') manifold (generally assumed to be compact)
$\mathcal{M}$	four-dimensional spacetime manifold
$\mathbb{N}$	the set of natural numbers
P^i_a	momentum canonically conjugate to A^i_a
$P(S, f)$	flux across a 2-surface S of P^a_i smeared with a test field f^i
$\hat{P}(S, f)$	quantum operator corresponding to $P(S, f)$
$\mathbb{P}^a$	momentum conjugate to the Maxwell connection $\mathbb{A}_a$
$\mathbb{P}(g)$	Maxwell momentum smeared against a test field g_a

$\hat{\mathbb{P}}(g)$	its corresponding quantum operator
q_{ab}	positive definite metric on M
$\hat{q}_{v,\alpha}$	the quantum operator representing the determinant of $q_{ab}(v)$, restricted to Cyl_α
$\mathbb{R}$	the set of real numbers
S	a closed-piecewise analytic sub-manifold of M (defined in section 4.3.1)
Σ^i_{ab}	Hodge-dual of the gravitational momentum P^a_i $\left(\Sigma^i_{ab} = \eta_{abc}\eta^{ij}E^c_j\right)$
Tr	trace
$V_\mathcal{R}$	the volume of a region $\mathcal{R}$ defined by q_{ab}
$\hat{V}_\mathcal{R}$	its corresponding quantum volume

References

Books and recent reviews

[1] Creutz M 1983 *Quarks, Gluons and Lattices* (Cambridge: Cambridge University Press)

[2] Ashtekar A 1991 *Lectures on Non-Perturbative Canonical Gravity* (Singapore: World Scientific) notes prepared in collaboration with R S Tate

[3] Baez J and Muniain J P 1994 *Gauge Fields, Knots and Gravity* (Singapore: World Scientific)

[4] Gambini R and Pullin J 1996 *Loops, Knots, Gauge Theories and Quantum Gravity* (Cambridge: Cambridge University Press)

[5] Carlip S 1998 *Quantum Gravity in 2 + 1 Dimensions* (Cambridge: Cambridge University Press)

[6] Rovelli C 1998 Loop quantum gravity *Liv. Rev. Rel* **1** 1

[7] Bojowald M 2001 *Quantum Geometry and Symmetry* (Aachen: Saker-Verlag)

[8] Ashtekar A 2002 Quantum geometry and gravity: recent advances *General Relativity and Gravitation* ed N T Bishop and S D Maharaj (Singapore: World Scientific) (*Preprint* gr-qc/0112038)

[9] Thiemann T 2002 Lectures on loop quantum gravity *Preprint* gr-qc/0210094 (2003 *Aspects of Quantum Gravity: From Theory to Experimental Search (Lecture Notes in Physics)* (Berlin: Springer) at press)
Thiemann T 2001 *Preprint* gr-qc/0110034 (2004 *Modern Canonical Quantum General Relativity* (Cambridge: Cambridge University Press) at press)

[10] Perez A 2003 Spin foam models for quantum gravity *Class. Quantum Grav.* **20** R43–104

[11] Rovelli C 2004 *Quantum Gravity* (Cambridge: Cambridge University Press) at press

Classical theory

[12] Ashtekar A 1986 New variables for classical and quantum gravity *Phys. Rev. Lett.* **57** 2244–7

[13] Ashtekar A 1987 New Hamiltonian formulation of general relativity *Phys. Rev.* D **36** 1587–602

[14] Samuel J 1987 A Lagrangian basis for Ashtekar's reformulation of canonical gravity *Pramana J. Phys.* **28** L429–32
Jacobson T and Smolin L 1987 The left handed spin connection as a variable for canonical gravity *Phys. Lett.* B **196** 39–42

[15] Ashtekar A, Mazur P and Torre C G 1987 BRST structure of general relativity in terms of new variables *Phys. Rev.* D **36** 2955–62

[16] Ashtekar A, Romano J D and Tate R S 1989 New variables for gravity: inclusion of matter *Phys. Rev.* D **40** 2572–87

[17] Jacobson T 1988 New variables for canonical supergravity *Class. Quantum Grav.* **5** 923–35
Matschull H J and Nicolai H 1994 Canonical quantum supergravity in three-dimensions *Nucl. Phys.* B **411** 609–46

[18] Ashtekar A, Balachandran A P and Jo S 1989 The CP problem in quantum gravity *Int. J. Mod. Phys.* A **4** 1493–514

[19] Plebański J F 1977 On the separation of Einsteinian substructures *J. Math. Phys.* **18** 2511

[20] Capovilla R, Dell J and Jacobson T 1989 General relativity without a metric *Phys. Rev. Lett.* **63** 2325–8
Capovilla R, Dell J, Jacobson T and Mason L 1991 Selfdual two forms and gravity *Class. Quantum Grav.* **8** 41–57

[21] Holst S 1996 Barbero's Hamiltonian derived from a generalized Hilbert–Palatini action *Phys. Rev.* D **53** 5966–9
Wisniewski J 2001 Symplectic structure from Holst action *Preprint*

[22] Samuel J 2000 Is Barbero's Hamiltonian formulation a gauge theory of Lorentzian gravity? *Class. Quantum Grav.* **17** L141–8

[23] Robinson D C 1995 A Lagrangian formalism for the Einstein–Yang–Mills equations *J. Math. Phys.* **36** 3733–42
Lewandowski J and Okol'ow A 2000 2-form gravity of the Lorentzian signature *Class. Quantum Grav.* **17** L47–51
[24] Alexandrov S and Livine E R 2003 $SU(2)$ loop quantum gravity seen from covariant theory *Phys. Rev.* D **67** 044009
[25] Freidel L and Krasnov K 1999 BF description of higher dimensional gravity theories *Adv. Theor. Math. Phys.* **3** 1289–324
[26] Jacobson T and Romano J D 1992 Degenerate extensions of general relativity *Class. Quantum Grav.* **9** L119–24
[27] Matschull H J 1996 Causal structure and diffeomorphisms in Ashtekar's gravity *Class. Quantum Grav.* **13** 765–82
[28] Jacobson T 1996 (1 + 1) sector of (3 + 1) gravity *Class. Quantum Grav.* **13** L111–6
Jacobson T 1996 *Class. Quantum Grav.* **13** 3269 (erratum)
[29] Lewandowski J and Wisniewski J 1997 (2 + 1) sector of (3 + 1) gravity *Class. Quantum Grav.* **14** 775–82
[30] Lewandowski J and Wisniewski J 1999 Degenerate sectors of the Ashtekar gravity *Class. Quantum Grav.* **16** 3057–69
[31] Penrose R 1976 Non-linear graviton and curved twistor theory *Gen. Rel. Grav.* **7** 31–52
[32] Ko M, Ludvigsen M, Newman E T and Tod P 1981 The theory of $\mathcal{H}$ space *Phys. Rep.* **71** 51–139

Connections and loops

[33] Rovelli C and Smolin L 1988 Knot theory and quantum gravity *Phys. Rev. Lett.* **61** 1155–8
[34] Rovelli C and Smolin L 1990 Loop representation for quantum general relativity *Nucl. Phys.* B **331** 80–152
[35] Ashtekar A, Husain V, Rovelli C, Samuel J and Smolin L 1989 2 + 1 quantum gravity as a toy model for the 3 + 1 theory *Class. Quantum Grav.* **6** L185–93
[36] Ashtekar A, Rovelli C and Smolin L 1991 Gravitons and loops *Phys. Rev.* D **44** 1740–55
[37] Smolin L 1992 Recent developments in non-perturbative quantum gravity *Quantum Gravity and Cosmology* ed J Mercader, J Solá and R Verdaguer (Singapore: World Scientific)
[38] Goldberg J N, Lewandowski J and Stornaiolo C 1992 Degeneracy in loop variables *Commun. Math. Phys.* **148** 377–402

Background-independent quantization of theories of connections

[39] Ashtekar A and Isham C J 1992 Representation of the holonomy algebras of gravity and non-Abelian gauge theories *Class. Quantum Grav.* **9** 1433–67
[40] Ashtekar A and Lewandowski J 1994 Representation theory of analytic holonomy algebras *Knots and Quantum Gravity* ed J C Baez (Oxford: Oxford University Press)
[41] Baez J C 1994 Generalized measures in gauge theory *Lett. Math. Phys.* **31** 213–23
[42] Lewandowski J 1994 Topological measure and graph-differential geometry on the quotient space of connections *Int. J. Mod. Phys.* D **3** 207–10
[43] Ashtekar A, Lewandowski L, Marolf D, Mourão J and Thiemann T 1995 A manifestly gauge invariant approach to quantum gauge theories *Geometry of Constrained Dynamical Systems* ed J M Charap (Cambridge: Cambridge University Press) pp 60–72
[44] Marolf D and Mourão J 1995 On the support of the Ashtekar–Lewandowski measure *Commun. Math. Phys.* **170** 583–606
[45] Ashtekar A and Lewandowski J 1995 Projective techniques and functional integration *J. Math. Phys.* **36** 2170–91
[46] Ashtekar A, Lewandowski L, Marolf D, Mourão J and Thiemann T 1996 Coherent state transforms for spaces of connections *J. Funct. Anal.* **135** 519–51
[47] Baez J C and Sawin S 1997 Functional integration on spaces of connections *J. Funct. Anal.* **150** 1–27
[48] Ashtekar A, Corichi A and Zapata J A 1998 Quantum theory of geometry: III. Non-commutativity of Riemannian structures *Class. Quantum Grav.* **15** 2955–72
[49] Marolf D, Mourão J and Thiemann T 1997 The status of diffeomorphism superselection in Euclidean 2 + 1 gravity *J. Math. Phys.* **38** 4730–40
[50] Mourão J M, Thiemann T and Velhinho J M 1999 Physical properties of quantum field theory measures *J. Math. Phys.* **40** 2337–53
[51] Velhinho J M 2002 A groupoid approach to spaces of generalized connections *J. Geom. Phys.* **41** 166–80
[52] Fleischhack C 2000 Stratification of the generalized gauge orbit space *Commun. Math. Phys.* **214** 607–49
[53] Fleischhack C 2003 On the Gribov problem for generalized connections *Commun. Math. Phys.* **234** 423–54
[54] Fleischhack C 2003 Hyphs and the Ashtekar–Lewandowski measure *J. Geom. Phys.* **45** 231–51
[55] Sahlmann H 2002 When do measures on the space of connections support the triad operators of loop quantum gravity? *Preprint* gr-qc/0207112

Sahlmann H 2002 Some comments on the representation theory of the algebra underlying loop quantum gravity *Preprint* gr-qc/0207111

[56] Okolów A and Lewandowski J 2003 Diffeomorphism covariant representations of the holonomy-flux star-algebra *Class. Quantum Grav.* **20** 3543–68

[57] Sahlmann H and Thiemann T 2003 On the superselection theory of the Weyl algebra for diffeomorphism invariant quantum gauge theories *Preprint* gr-qc/0302090

Sahlmann H and Thiemann T 2003 Irreducibility of the Ashtekar–Isham–Lewandowski representation *Preprint* gr-qc/0303074

[58] Lewandowski J, Okolów A, Sahlmann H and Thiemann T 2003 Uniqueness of the diffeomorphism invariant state on the quantum holonomy-flux algebra *Preprint*

Spin networks

[59] Penrose R 1971 Angular momentum: an approach to combinatorial space-time *Quantum Theory and Beyond* ed T Bastin (Cambridge: Cambridge University Press)

[60] Rovelli C and Smolin L 1995 Spin networks and quantum gravity *Phys. Rev.* D **52** 5743–59

[61] Baez J C 1996 Spin networks in non-perturbative quantum gravity *The Interface of Knots and Physics* ed L Kauffman (Providence, RI: American Mathematical Society) pp 167–203

Baez J C 1996 Spin networks in gauge theory *Adv. Math.* **117** 253–72

[62] Thiemann T 1998 The inverse loop transform *J. Math. Phys.* **39** 1236–48

[63] Baez J C and Sawin S 1998 Diffeomorphism-invariant spin network states *J. Funct. Anal.* **158** 253–66

[64] Fleischhack C 2003 Proof of a conjecture by Lewandowski and Thiemann *Preprint* math-ph/0304002

Geometric operators and their properties

[65] Rovelli C and Smolin L 1995 Discreteness of area and volume in quantum gravity *Nucl. Phys.* B **442** 593–622
Rovelli C and Smolin L 1995 *Nucl. Phys.* B **456** 753 (erratum)

[66] Ashtekar A and Lewandowski L 1995 Differential geometry on the space of connections using projective techniques *J. Geom. Phys.* **17** 191–230

[67] Loll R 1995 The volume operator in discretized quantum gravity *Phys. Rev. Lett.* **75** 3048–51

[68] Loll R 1995 Spectrum of the volume operator in quantum gravity *Nucl. Phys.* B **460** 143–54

[69] Frittelli S, Lehner L and Rovelli C 1996 *Class. Quantum Grav.* **13** 2921–32

[70] De Pietri R and Rovelli C 1996 *Phys. Rev.* D **54** 2664–90

[71] Lewandowski J 1997 Volume and quantizations *Class. Quantum Grav.* **14** 71–6

[72] Ashtekar A and Lewandowski J 1997 Quantum theory of geometry: I. Area operators *Class. Quantum Grav.* **14** A55–81

[73] Ashtekar A and Lewandowski J 1997 Quantum theory of geometry: II. Volume operators *Adv. Theor. Math. Phys.* **1** 388–429

[74] Loll R 1997 Simplifying the spectral analysis of the volume operator *Nucl. Phys.* B **500** 405–20

[75] Loll R 1997 Further results on geometric operators in quantum gravity *Class. Quantum Grav.* **14** 1725–41

[76] De Pietri R 1997 Spin networks and recoupling in loop quantum gravity *Nucl. Phys. Suppl.* **57** 251–4

[77] De Pietri R 1997 On the relation between the connection and the loop representation of quantum gravity *Class. Quantum Grav.* **14** 53–70

[78] Thiemann T 1998 Closed formula for the matrix elements of the volume operator in canonical quantum gravity *J. Math. Phys.* **39** 3347–71

[79] Thiemann T 1998 A length operator for canonical quantum gravity *J. Math. Phys.* **39** 3372–92

[80] Major S A 1999 Operators for quantized directions *Class. Quantum Grav.* **16** 3859–77

Barbero–Immirzi ambiguity

[81] Barbero F 1996 Real Ashtekar variables for Lorentzian signature spacetimes *Phys. Rev.* D **51** 5507–10

[82] Immirzi G 1997 Quantum gravity and Regge calculus *Nucl. Phys. Proc. Suppl.* **57** 65–72

[83] Rovelli C and Thiemann T 1998 The Immirizi parameter in quantum general relativity *Phys. Rev.* D **57** 1009–14

[84] Gambini R, Obregon O and Pullin J 1999 Yang–Mills analogues of the Immirzi ambiguity *Phys. Rev.* D **59** 047505

Quantum Einstein's equation I

[85] Kuchař K 1993 Canonical quantum gravity *General Relativity and Gravitation 1992* ed R J Gleiser, C N Kozameh and O M Moreschi (Bristol: Institute of Physics Publishing) pp 119–50

[86] Ashtekar A, Tate R S and Uggla C 1993 Minisuperspaces: observables and quantization *Int. J. Phys.* D **2** 15–50
Ashtekar A, Tate R S and Uggla C 1993 Minisuperspaces: symmetries and quantization *Misner Festschrift* ed B L Hu *et al* (Cambridge: Cambridge University Press)

[87] Ashtekar A and Tate R S 1994 An algebraic extension of Dirac quantization: examples *J. Math. Phys.* **35** 6434–70

[88] Marolf D 1995 Refined algebraic quantization: systems with a single constraint *Preprint* gr-qc/9508015
[89] Ashtekar A, Lewandowski J, Marolf D, Mourão J and Thiemann T 1995 Quantization of diffeomorphism invariant theories of connections with local degrees of freedom *J. Math. Phys.* **36** 6456–93
[90] Marolf D, Mourão J and Thiemann T 1997 The status of diffeomorphism super-selection in Euclidean 2 + 1 gravity *J. Math. Phys.* **38** 4730–40
[91] Guilini N and Marolf D 1999 On the generality of refined algebraic quantization *Class. Quantum Grav.* **16** 2479–88
Guilini N and Marolf D 1999 A uniqueness theorem for constraint quantization *Class. Quantum Grav.* **16** 2489–505
[92] Lewandowski J and Thiemann T 1999 Diffeomorphism invariant quantum field theories of connections in terms of webs *Class. Quantum Grav.* **16** 2299–322
[93] Corichi A and Zapata J A 1997 On diffeomorphism invariance for lattice theories *Nucl. Phys.* B **493** 475–90

Quantum Einstein's equation II

[94] Rovelli C and Smolin L 1994 The physical Hamiltonian in nonperturbative quantum gravity *Phys. Rev. Lett.* **72** 446–9
[95] Thiemann T 1996 Anomaly-free formulation of non-perturbative, four-dimensional Lorentzian quantum gravity *Phys. Lett.* B **380** 257–64
[96] Thiemann T 1998 Quantum spin dynamics (QSD) *Class. Quantum Grav.* **15** 839–73
[97] Thiemann T 1998 QSD: III. Quantum constraint algebra and physical scalar product in quantum general relativity *Class. Quantum Grav.* **15** 1207–47
[98] Thiemann T 1998 QSD: V. Quantum gravity as the natural regulator of matter quantum field theories *Class. Quantum Grav.* **15** 1281–314
[99] Thiemann T 2001 Quantum spin dynamics (QSD): VII. Symplectic structures and continuum lattice formulations of gauge field theories *Class. Quantum Grav.* **18** 3293–338
[100] Gambini R, Lewandowski J, Marolf D and Pullin J 1998 On the consistency of the constraint algebra in spin network quantum gravity *Int. J. Mod. Phys.* D **7** 97–109
[101] Lewandowski J and Marolf D 1998 Loop constraints: a habitat and their algebra *Int. J. Mod. Phys.* D **7** 299–330
[102] Gaul G and Rovelli C 2001 Generalized Hamiltonian constraint operator in loop quantum gravity and its simplest Euclidean matrix elements *Class. Quantum Grav.* **18** 1593–624

Big-bang

[103] Kodama H 1988 Specialization of Ashtekar's formalism to Bianchi cosmology *Prog. Theor. Phys.* **80** 1024–40
Kodama H 1990 Holomorphic wavefunction of the universe *Phys. Rev.* D **42** 2548–65
[104] Bojowald M 2001 Absence of singularity in loop quantum cosmology *Phys. Rev. Lett.* **86** 5227–30
[105] Bojowald M 2001 Inverse scale factor in isotropic quantum geometry *Phys. Rev.* D **64** 084018
[106] Bojowald M 2001 Dynamical initial conditions in quantum cosmology *Phys. Rev. Lett.* **87** 121301
[107] Bojowald M 2001 Loop quantum cosmology: IV. Discrete time evolution *Class. Quantum Grav.* **18** 1071–88
[108] Bojowald M 2001 Loop quantum cosmology: III. Wheeler–DeWitt operators *Class. Quantum Grav.* **18** 1055–70
Bojowald M 2002 The semiclassical limit of loop quantum cosmology *Class. Quantum Grav.* **18** L109–16
[109] Bojowald M 2001 The inverse scale factor in isotropic quantum geometry *Phys. Rev.* D **64** 084018
[110] Bojowald M 2002 Isotropic loop quantum cosmology *Class. Quantum Grav.* **19** 2717–42
[111] Bojowald M and Hinterleitner F 2002 Isotropic loop quantum cosmology with matter *Phys. Rev.* D **66** 104003
[112] Bojowald M and Vandersloot K 2003 Loop quantum cosmology, boundary proposals, and inflation *Phys. Rev.* D **67** 124023
[113] Bojowald M 2003 Homogeneous loop quantum cosmology *Class. Quantum Grav.* **20** 2595–615
[114] Bojowald M 2003 *Gen. Rel. Grav.* **35** 1877–83 first prize in the Gravity Research Foundation Essay Contest (Initial conditions for a universe *Preprint* gr-qc/0305069)
[115] Ashtekar A, Bojowald M and Lewandowski J 2003 Mathematical structure of loop quantum cosmology *Adv. Theor. Math. Phys.* **7** 233–68
[116] Bojowald M, Date G and Vandersloot K 2004 Homogeneous loop quantum cosmology: the role of the spin connection *Class. Quantum Grav.* **21** 1253–78
[117] Ashtekar A, Bojowald M and Willis J 2004 Quantum corrections to Friedmann equations in loop quantum cosmology (in preparation)

Black holes

[118] Bekenstein J D 1973 Black holes and entropy *Phys. Rev.* D **7** 2333–46
Bekenstein J D 1974 Generalized second law of thermodynamics in black hole physics *Phys. Rev.* D **9** 3292–300

Bekenstein J D and Meisels A 1977 Einstein A and B coefficients for a black hole *Phys. Rev.* D **15** 2775–81

[119] Bardeen J W, Carter B and Hawking S W 1973 The four laws of black hole mechanics *Commun. Math. Phys.* **31** 161–70

[120] Hawking S W 1975 Particle creation by black holes *Commun. Math. Phys.* **43** 199–220

[121] Gibbons G and Hawking S W 1977 Cosmological event horizons, thermodynamics, and particle creation *Phys. Rev.* D **15** 2738–51

[122] Smolin L 1995 Linking topological quantum field theory and nonperturbative quantum gravity *J. Math. Phys.* **36** 6417–55

[123] Barreira M, Carfora M and Rovelli C 1996 Physics with non-perturbative quantum gravity: radiation from a quantum black hole *Gen. Rel. Grav.* **28** 1293–9

Rovelli C 1996 Black hole entropy from loop quantum gravity *Phys. Rev. Lett.* **14** 3288–91

Rovelli C 1996 Loop quantum gravity and black hole physics *Helv. Phys. Acta.* **69** 582–611

[124] Krasnov K 1997 Geometrical entropy from loop quantum gravity *Phys. Rev.* D **55** 3505–13

Krasnov K 1998 On statistical mechanics of Schwarzschild black holes *Gen. Rel. Grav.* **30** 53–68

[125] Jacobson T, Kang G and Myers R C 1994 On black hole entropy *Phys. Rev.* D **49** 6587–98

[126] Wald R 1993 Black hole entropy is Noether charge *Phys. Rev.* D **48** 3427–31

Iyer V and Wald R 1994 Some properties of Noether charge and a proposal for dynamical black hole entropy *Phys. Rev.* D **50** 846–64

[127] Ashtekar A, Baez J C, Corichi A and Krasnov K 1998 Quantum geometry and black hole entropy *Phys. Rev. Lett.* **80** 904–7

[128] Ashtekar A, Corichi A and Krasnov K 1999 Isolated horizons: the classical phase space *Adv. Theor. Math. Phys.* **3** 418–71

[129] Ashtekar A, Baez J C and Krasnov K 2000 Quantum geometry of isolated horizons and black hole entropy *Adv. Theor. Math. Phys.* **4** 1–95

[130] Kaul R K and Majumdar P 2000 Logarithmic corrections to the Bekenstein–Hawking entropy *Phys. Rev. Lett.* **84** 5255–7

[131] Ashtekar A, Beetle C and Fairhurst S 1999 Isolated horizons: a generalization of black hole mechanics *Class. Quantum Grav.* **16** L1–7

Ashtekar A, Beetle C and Fairhurst S 2000 Mechanics of isolated horizons *Class. Quantum Grav.* **17** 253–98

[132] Lewandowski J 2000 Spacetimes admitting isolated horizons *Class. Quantum Grav.* **17** L53–9

[133] Ashtekar A, Fairhurst S and Krishnan B 2000 Isolated horizons: Hamiltonian evolution and the first law *Phys. Rev.* D **62** 104025

[134] Ashtekar A, Beetle C and Lewandowski J 2001 Mechanics of rotating isolated horizons *Phys. Rev.* D **64** 044016

[135] Ashtekar A, Beetle C and Lewandowski J 2002 Geometry of generic isolated horizons *Class. Quantum Grav.* **19** 1195–225

[136] Ashtekar A, Corichi A and Sudarski D 2003 Non-minimally coupled scalar fields and isolated horizons *Class. Quantum Grav.* **20** 3413–25

[137] Ashtekar A and Corichi A 2003 Non-minimal couplings, quantum geometry and black hole entropy *Class. Quantum Grav.* **20** 4473–84

[138] Ashtekar A, Engle J, Pawlowski T and van der Broeck C 2004 Multipole moments of isolated horizons *Class. Quantum Grav.* **21** 2549–70

[139] Ashtekar A 2003 Black hole entropy: inclusion of distortion and angular momentum, online at http://www.phys.psu.edu/events/index.html?event_id=517&event_type=17

Ashtekar A, Engle J and van der Broeck C 2004 in preparation

[140] Wheeler J A 1992 *It* from *bit Sakharov Memorial Lectures on Physics* vol 2, ed L Keldysh and V Feinberg (Moscow: Nova Science)

[141] Ghosh A and Mitra P 2004 A bound on the log correction to the black hole area law *Preprint* gr-qc/0401070

Low energy physics:

[142] Ashtekar A, Rovelli C and Smolin L 1992 Weaving a classical geometry with quantum threads *Phys. Rev. Lett.* **69** 237–40

[143] Arnsdorf M and Gupta S 2000 Loop quantum gravity on non-compact spaces *Nucl. Phys.* B **577** 529–46

[144] Corichi A and Reyes J M 2001 A Gaussian weave for kinematical loop quantum gravity *Int. J. Mod. Phys.* D **10** 325–38

[145] Hall B C 1994 The Segal–Bergmann coherent state transform for compact Lie groups *J. Funct. Anal.* **122** 103–51

[146] Sahlmann H, Thiemann T and Winkler O 2001 Coherent states for canonical quantum general relativity and the infinite tensor product extension *Nucl. Phys.* B **606** 401–40

[147] Thiemann T 2001 Gauge field theory coherent states (GCS): I. General properties *Class. Quantum Grav.* **18** 2025–64
Thiemann T and Winkler O 2001 Gauge field theory coherent states (GCS): II. Peakedness properties *Class. Quantum Grav.* **18** 2561–636
Thiemann T and Winkler O 2001 Gauge field theory coherent states (GCS): III. Ehrenfest theorems *Class. Quantum Grav.* **18** 4629–82
Thiemann T and Winkler O 2001 Gauge field theory coherent states (GCS): IV. Infinite tensor product and thermodynamical limit *Class. Quantum Grav.* **18** 4997–5054
[148] Sahlmann H and Thiemann T 2002 Towards the QFT on curved spacetime limit of QGR: I. A general scheme *Preprint* gr-qc/0207030
Sahlmann H and Thiemann T 2002 Towards the QFT on curved spacetime limit of QGR: II. A concrete implementation *Preprint* gr-qc/0207031
[149] Thiemann T 2002 Complexifier coherent states for quantum general relativity *Preprint* gr-qc/0206037
[150] Bombelli L 2002 Statistical geometry of random weave states *Proc. 9th Marcel Grossmann Meeting on General Relativity* ed V G Gurzadyan, R T Jantzen and R Ruffini (Singapore: World Scientific) (*Preprint* gr-qc/0101080)
[151] Ashtekar A and Isham C J 1992 *Phys. Lett.* B **274** 393–8
[152] Varadarajan M 2000 Fock representations from U(1) holonomy algebras *Phys. Rev.* D **61** 104001
Varadarajan M 2001 Photons from quantized electric flux representations *Phys. Rev.* D **64** 104003
[153] Ashtekar A and Lewandowski J 2001 Relation between polymer and Fock excitations *Class. Quantum Grav.* **18** L117–27
[154] Velhinho J M 2002 Invariance properties of induced Fock measures for U(1) holonomies *Commun. Math. Phys.* **227** 541–50
[155] Ashtekar A, Fairhurst S and Willis J 2003 Quantum gravity, shadow states, and quantum mechanics *Class. Quantum Grav.* **20** 1031–61
[156] Ashtekar A, Fairhurst S and Ghosh A 2004 (in preparation)
[157] Varadarajan M 2002 Gravitons from a loop representation of linearised gravity *Phys. Rev.* D **66** 024017
[158] Ashtekar A, Ghosh A and Van Den Broeck C 2004 (in preparation)
[159] Bobienski M, Lewandowski J and Mroczek M 2002 A 2-surface quantization of the Lorentzian gravity *Proc. 9th Marcel Grossmann Meeting on General Relativity* ed V G Gurzadyan, R T Jantzen and R Ruffini (Singapore: World Scientific) (*Preprint* gr-qc/0101069)
[160] Ashtekar A, Lewandowski J and Sahlmann H 2003 Polymer and Fock representations of a scalar field *Class. Quantum Grav.* **20** L11–21

Spin foams and finiteness

[161] Baez J C 2000 An introduction to spin foam models of quantum gravity and BF theory *Lect. Notes Phys.* **543** 25–94
[162] Baez J C 1998 Spin foam models *Class. Quantum Grav.* **15** 1827–58
[163] Reisenberger M P 1997 A lattice worldsheet sum for 4-d Euclidean general relativity *Preprint* gr-qc/9711052
Reisenberger M P 1999 On relativistic spin network vertices *J. Math. Phys.* **40** 2046–54
[164] Reisenberger M P and Rovelli C 2001 Spacetime as a Feynman diagram: the connection formulation *Class. Quantum Grav.* **18** 121–40
Reisenberger M P and Rovelli 2002 Spacetime states and covariant quantum theory *Phys. Rev.* D **65** 125016
[165] Barrett J W and Crane L 1998 Relativistic spin networks and quantum gravity *J. Math. Phys.* **39** 3296–302
Barrett J W and Crane L 2000 A Lorentzian signature model for quantum general relativity *Class. Quantum Grav.* **17** 3101–18
[166] Perez A 2001 Finiteness of a spinfoam model for Euclidean quantum general relativity *Nucl. Phys.* B **599** 427–34
Perez A and Rovelli C 2001 Spin foam model for Lorentzian general relativity *Phys. Rev.* D **63** 041501
Crane L, Perez A and Rovelli C 2001 Perturbative finiteness in spin-foam quantum gravity *Phys. Rev. Lett.* **87** 181301
Crane L, Perez A and Rovelli C 2001 3 + 1 spinfoam model of quantum gravity with spacelike and timelike components *Phys. Rev.* D **64** 064002
[167] Baez J, Christensen D, Halford T R and Tsang D C 2002 Spin foam models of Riemannian quantum gravity *Class. Quantum Grav.* **19** 4627–48
[168] Gambini R and Pullin J 2002 A finite spin-foam-based theory of three and four dimensional quantum gravity *Phys. Rev.* D **66** 024020
[169] Lauscher O and Reuter M 2002 Is quantum Einstein gravity non-perturbatively renormalizable? *Class. Quantum Grav.* **19** 483–92

[170] Percacci R and Perini D 2003 Asymptotic safety of gravity coupled to matter *Phys. Rev.* D **68** 044018
Perini D 2003 Gravity and matter with asymptotic safety *Preprint* hep-th/0305053
[171] Bojowald M and Perez A 2003 Spin foam quantization and anomalies *Preprint* gr-qc/0303026
[172] Freidel L and Louapre D 2003 Diffeomorphisms and spin foam models *Nucl. Phys.* B **662** 279–98
Freidel L and Louapre D 2003 Non-perturbative summation over 3D discrete topologies *Phys. Rev.* D **68** 104004 (*Preprint* hep-th/0211026)
[173] Perez A and Noui K 2004 Three dimensional loop quantum gravity: coupling to point particles *Preprint* gr-qc/0402111
Perez A and Noui K 2004 Three dimensional loop quantum gravity: physical scalar product and spin foam models *Preprint* gr-qc/0402110

Outlook

[174] Ashtekar A, Lewandowski J, Marolf D, Mourão J and Thiemann T 1997 $SU(N)$ quantum Yang–Mills theory in two dimensions: a complete solution *J. Math. Phys.* **38** 5453–82
[175] Ashtekar A and Sahlmann H 2004 (in preparation)
[176] Okolw A 2004 Representations of quantum geometry *PhD Thesis* University of Warsaw
[177] Rovelli C and Speziale S 2003 Reconcile Planck-scale discreteness and the Lorentz–Fitzgerald contraction *Phys. Rev.* D **67** 064019
[178] Thiemann T 2003 The Phoenix project: master constraint programme for loop quantum gravity *Preprint* gr-qc/0305080
[179] Dittrich B 2004 Testing the master constraint programme for loop quantum gravity, online at http://www.phys.psu.edu/events/index.html?event_id=850&event_type=17
[180] Klauder J 2003 Affine quantum gravity *Int. J. Mod. Phys.* D **12** 1769–74
Klauder J 2004 The utility of coherent states and other mathematical methods in the foundations of affine quantum gravity *Preprint* hep-th/0401214
[181] Di Bartolo C, Gambini R, Griego J and Pullin J 2000 Consistent canonical quantization of general relativity in the space of Vassiliev knot invariants *Phys. Rev. Lett.* **84** 2314–7
Di Bartolo C, Gambini R, Griego J and Pullin J 2000 Canonical quantum gravity in the Vassiliev invariants arena: I. Kinematical structure *Class. Quantum Grav.* **17** 3211–38
Di Bartolo C, Gambini R, Griego J and Pullin J 2000 Canonical quantum gravity in the Vassiliev invariants arena: II. Constraints, habitats and consistency of the constraint algebra *Class. Quantum Grav.* **17** 3239–64
[182] Gambini R and Pullin J 2003 Canonical quantization of general relativity in discrete spacetimes *Phys. Rev. Lett.* **90** 021301
Gambini R, Porto R and Pullin J 2004 Loss of coherence from discrete quantum gravity *Class. Quantum Grav.* **21** L51–7
Gambini R, Porto R and Pullin J 2004 A relational solution to the problem of time in quantum mechanics and quantum gravity induces a fundamental mechanism for quantum decoherence *Preprint* gr-qc/0402118
[183] Horowitz G T 1998 Quantum states of black holes *Black Holes and Relativistic Stars* ed R M Wald (Chicago, IL: University of Chicago Press)
[184] Carlip S 1999 Entropy from conformal field theory at Killing horizons *Class. Quantum Grav.* **16** 3327–48
[185] Dreyer O, Ghosh A and Wisniewski J 2001 Black hole entropy calculations based on symmetries *Class. Quantum Grav.* **18** 1929–38
[186] Carlip S 2002 Near-horizon conformal symmetry and black hole entropy *Phys. Rev. Lett.* **88** 241301
[187] Ashtekar A and Krishnan B 2002 Dynamical horizons: energy, angular momentum, fluxes and balance laws *Phys. Rev. Lett.* **89** 261101
Ashtekar A and Krishnan B 2003 Dynamical horizons and their properties *Phys. Rev.* D **68** 104030
[188] Bojowald M 2002 Inflation from quantum geometry *Phys. Rev. Lett.* **89** 261301
[189] Bojowald M and Date G 2004 Quantum suppression of the generic chaotic behavior close to cosmological singularities *Phys. Rev. Lett.* **92** 071302
[190] Tsujikawa S, Singh P and Maartens R 2003 Loop quantum gravity effects on inflation and the CMB *Preprint* astro-ph/0311015