TOPICAL REVIEW

Loop quantum gravity: an outside view

Hermann Nicolai, Kasper Peeters and Marija Zamaklar

Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1,
14476 Golm, Germany

E-mail: hermann.nicolai@aei.mpg.de, kasper.peeters@aei.mpg.de and
marija.zamaklar@aei.mpg.de

Abstract
We review aspects of loop quantum gravity in a pedagogical manner, with the
aim of enabling a precise but critical assessment of its achievements so far.
We emphasize that the off-shell ('strong') closure of the constraint algebra is a
crucial test of quantum spacetime covariance, and thereby of the consistency,
of the theory. Special attention is paid to the appearance of a large number
of ambiguities, in particular in the formulation of the Hamiltonian constraint.
Developing suitable approximation methods to establish a connection with
classical gravity on the one hand, and with the physics of elementary particles
on the other, remains a major challenge.

PACS numbers: 04.60.−m, 04.60.Ds, 04.60.Pp

(Some figures in this article are in colour only in the electronic version)

Contents

1. Key questions

When four-dimensional Einstein gravity is quantized canonically, in a perturbation series in the Newton constant around flat spacetime, divergences arise at two-loop order. An impressive calculation by Goroff and Sagnotti [1, 2] has demonstrated that in order to obtain a finite S-matrix, the action should contain a counterterm

$$\Gamma^{(2)}_{\text{div}} = \frac{1}{\epsilon} \frac{209}{2880} \frac{1}{(16\pi^2)^2} \int d^4x \, \sqrt{g} C_{\mu\nu\rho\sigma} C^{\rho\sigma\lambda\tau} C_{\lambda\tau}{}^{\mu\nu}, \tag{1.1}$$

a result which was later confirmed in an independent background-field calculation by van de Ven [3]. The usual conclusion drawn from this result, and from the fact that the coupling constant is dimensionful, is that an infinite number of counterterms are needed. It is generally agreed that this non-renormalizability renders perturbatively quantized Einstein gravity meaningless as a fundamental theory because an infinite number of parameters would be required to make any physical prediction.

However, this still leaves open the possibility that Einstein gravity *can* be quantized consistently, but that it is simply the perturbation series in Newton's constant which is ill-defined. This possibility has been raised and advocated not only in the context of canonical quantization, but also, and independently, in the context of suggestions that there may exist a non-trivial fixed point of the renormalization group in Einstein's theory [4–6]. Implicitly, it also underlies the path integral approach to Euclidean quantum gravity [7, 8], which provides a possible framework for the discussion of semi-classical states [9]. Indeed, to date there is no proof that such a quantization which does *not* make use of a series expansion around a fixed background is guaranteed to fail. However, quantizing without relying on perturbation theory around a free theory is hard.

Loop quantum gravity, or LQG for short, is an attempt to quantize Einstein gravity non-perturbatively. In contrast to string theory, which posits that the Einstein–Hilbert action is only an effective low energy approximation to some other, more fundamental, underlying theory, LQG takes Einstein's theory in *four* spacetime dimensions as the basic starting point. It is a Hamiltonian approach, which is background independent in the sense that its basic quantities and concepts do not presuppose the existence of a given spatial background metric. In comparison to the older geometrodynamics approach (which is also formally background independent) it makes use of many new conceptual and technical ingredients. A key role is played by the Ashtekar variables, which allow the reformulation of gravity in terms of connections and holonomies, and which (at least in the form originally proposed) greatly

simplify the constraints. A related key feature is the use of spin networks, which in turn requires other mathematical ingredients, such as non-separable ('polymer') Hilbert spaces and representations of operators, which are not weakly continuous. Undoubtedly, novel concepts and ingredients such as these will be necessary in order to circumvent the problems of perturbatively quantized gravity, but, as for any other approach to quantum gravity, it is important not to loose track of the physical questions that one is trying to answer.

The goal of the present paper is to review some essential properties of loop quantum gravity in an easily accessible way from a non-specialist's perspective, and with a non-LQG audience in mind. It is *not meant to be a comprehensive review of the subject*; readers who want to know more about the very latest developments in this field are instead referred to a number of excellent recent reviews describing the story from the specialist's point of view [10–14]. Rather, we would like to provide an *entrée* for 'outsiders', and to focus on the outstanding problems as we perceive them, and thereby initiate and enable an informed debate between the proponents of the different approaches. Accordingly, we will take the liberty to omit mathematical details that in our opinion are not truly essential to understand the physical consequences of the formalism (at least not on a first perusal of the literature), and to describe some results 'our own way'. At the same time, as we move along, we will try to make precise and clearly state the questions that are often raised about the LQG programme (for earlier reviews which mention some of these concerns, see [15–17]).

In order to focus the discussion, and for the reader's convenience, we begin with a summary of what we consider to be the main issues and open questions.

- *How do Einstein's equations appear in the classical limit?* The space of quantum states used in LQG (not necessarily a Hilbert space) is very different from that used in Fock space quantization. As a consequence, it becomes non-trivial to see how semi-classical 'coherent' states can be constructed, and how a smooth classical spacetime might emerge. In simple toy examples, such as the harmonic oscillator, it has been shown that the LQG quantization method indeed leads to quantum states whose properties are close to those of the usual Fock space coherent states [18]. In full $(3 + 1)$-dimensional LQG, the classical limit is, however, far from understood (so far only kinematical coherent states are known [19–24]). In particular, we do not know how to describe or approximate classical spacetimes in this framework that 'look' like, say, Minkowski space, or how to properly derive the classical Einstein equations and their quantum corrections. A proper understanding of the semi-classical limit is also indispensable to clarify the connection (or lack thereof) between conventional perturbation theory in terms of Feynman diagrams, and the non-perturbative quantization proposed by LQG.

- *Renormalization versus regularization: where is the 2-loop divergence?* No approach to quantum gravity can claim success that does not explain the ultimate fate of the two-loop divergence (1.1). In a consistent scheme, this divergence must be either eliminated by cancellation, or disposed of by some other mechanism. The key question raised by (1.1) is therefore what happens when one expands the results of non-perturbatively quantized Einstein gravity in Newton's constant. When such an expansion is performed about a semiclassical state (which remains to be found in LQG, see above), the two-loop divergence should manifest itself in one form or another. In its present incarnation, LQG cannot (yet?) 'see' and accommodate this divergence. Its possible 'disappearance' is occasionally argued to be due to the emergence of an effective cut-off (regulator) which might eventually make the perturbation theory finite. If this is the case, an obvious question concerns the true origin of this cut-off: is it generated dynamically, or covertly put in 'by hand' (for instance, by working with the compact group $SU(2)$ instead of the full Lorentz

group)? There are also questions concerning the meaning of 'regularization'. According to conventional (quantum field theory) wisdom, physics is supposed not to depend on the way in which the theory is regulated before the cutoff is removed; how can it be that physics predictions of LQG do depend on the chosen regularization prescription? This question is in part answered by the fact that the notions of 'finiteness' and 'regulator independence' as currently used in LQG on the one hand, and in conventional quantum field theory and perturbative quantum gravity on the other, are *not* the same (see section 4.5). Let us furthermore stress that in spite of its perturbative origin, the result (1.1) cannot be so easily dismissed as a background artefact: while it does require *some* background for its derivation (i.e. a spacetime solving Einstein's equations), the counterterm is actually independent of the particular background about which one expands, see [3], as is also evident from the manifestly spacetime covariant form in which it is written.

- *Status of the Hamiltonian constraint?* In the current LQG literature, there is surprisingly little discussion of certain basic aspects concerning the Hamiltonian constraint operator, which are of central importance for the theory (recent exceptions are [25], and the so-called master constraint programme of [26]). For this reason, we will here describe the Hamiltonian constraint operator and its action on a given spin network wavefunction in rather pedestrian detail, as far as we have been able to work it out. In particular, we will exhibit the numerous choices and ambiguities inherent in this construction, as well as the extraordinary complexity of the resulting expression for the constraint operator. The number of ambiguities can be reduced by invoking independence of the spatial background [11], and indeed, without making such choices, one would not even obtain sensible expressions, as we shall see very explicitly. In other words, the formalism is partly 'on-shell' in that the very existence of the (unregulated) Hamiltonian constraint operator depends very delicately on its 'diffeomorphism covariance', and the choice of a proper 'habitat', on which it is supposed to act in a well-defined manner. A further source of ambiguities, which, for all we know, has not been considered in the literature so far, consists of possible $\hbar$-dependent 'higher order' modifications of the Hamiltonian (5.14), which might still be compatible with all consistency requirements of LQG.

    The attitude often expressed with regard to the remaining ambiguities is that they correspond to *different* physics, and therefore the choice of the correct Hamiltonian is ultimately a matter of physics (experiment?), and not mathematics. We disagree, because we cannot believe that Nature will allow such a great degree of arbitrariness at its most fundamental level: recall that it was precisely the infinite number of parameters and the concomitant ambiguities which killed perturbative quantum gravity!

- *Does the quantum theory possess full spacetime covariance?* Spacetime covariance is a central property of Einstein's theory. Although the Hamiltonian formulation is not manifestly covariant, full covariance is still present in the classical theory, albeit in a hidden form, via the classical (Poisson or Dirac) algebra of constraints acting on phase space. However, this is not necessarily so for the quantized theory. LQG treats the diffeomorphism constraint and the Hamiltonian constraint in a very different manner. Why and how then should one expect such a theory to recover full spacetime (as opposed to purely spatial) covariance? The crucial issue here is clearly what LQG has to say about the quantum algebra of constraints. Unfortunately, to the best of our knowledge, the 'off-shell' calculation of the commutator of two Hamiltonian constraints in LQG—with an explicit operatorial expression as the final result—has never been fully carried out. Instead, a survey of the possible terms arising in this computation has led to the conclusion that the commutator vanishes on a certain restricted 'habitat' of states [27–29], and therefore the LQG constraint algebra closes without anomalies. By contrast, we will

here argue that this 'on shell closure' is not sufficient for a full proof of *quantum spacetime covariance*, but that a proper theory of quantum gravity requires a constraint algebra that closes 'off shell', i.e. without prior imposition of a subset of the constraints. The fallacies that may ensue if one does not insist on off-shell closure can be illustrated with simple examples. In our opinion, this requirement may well provide the acid test on which any proposed theory of canonical quantum gravity will stand or fail.

- *Matter couplings: anything goes?* Because LQG is claimed to be a finite and fully consistent theory of quantum gravity, it does not appear to impose any restrictions on the types of matter that are coupled to gravity, nor on their interactions. Indeed it is straightforward, though sometimes cumbersome, to extend the formalism to include matter: in this perspective, matter appears to be a mere accessory that can be added on to pure gravity as one chooses. This is in marked contrast to supergravity and superstring theory, which are based on the hypothesis that the very *raison d' être* of matter is its indispensability for curing the perturbative (and non-perturbative) inconsistencies of quantum gravity, and the desire to 'geometrize' matter in the framework of a (probably supersymmetric) unified theory. It is not inconceivable that LQG might eventually encounter new consistency requirements when descending from the kinematical to the physical Hilbert space, but we see no evidence for this so far. Therefore, the question remains whether and how LQG can recover the consistency requirements that conventional perturbative quantum field theory imposes on the matter content of the world, in particular those resulting from cancellation of *local (gauge) anomalies*. Let us recall that Nature does 'care' about such consistency requirements, in that it has chosen to put the three known generations of fermions into *anomaly free multiplets* of the standard model gauge group. Similar comments apply to global anomalies. Is it possible to obtain the correct answer for pion decay when fermions are coupled to electromagnetism in the LQG approach, or would LQG predict the neutral π meson to be a stable particle?

- *Structure of space(–time) at the smallest scales?* There is a general expectation (not only in the LQG community) that at the very shortest distances, the smooth geometry of Einstein's theory will be replaced by some quantum space or spacetime, and hence the continuum will be replaced by some 'discretuum'. LQG does not do away with conventional spacetime concepts, in that it still relies on a spatial continuum as its 'substrate'. At the kinematical level, it imposes a discrete structure which is very different from the discreteness of a lattice or naive discretization of space (i.e. of a finite or countable set), by 'polymerizing' the continuum via the scalar product (4.7). This is similar to the discrete topology 'pulverization' of the real line with countable unions of points as the open sets. Because the only notion of 'closeness' between two points in this topology is whether or not they are coincident, whence *any* function is continuous in this topology, this raises the question as to how one can recover conventional notions of continuity in this scheme.

    However, the truly relevant question here concerns the structure (and definition!) of *physical* space and time. This, and not the kinematical 'discretuum' on which holonomies and spin networks 'float', is the arena where one should try to recover familiar and well-established concepts like the Wilsonian renormalization group, with its *continuous* 'flows'. Because the measurement of lengths and distances ultimately requires an operational definition in terms of appropriate matter fields and states obeying the physical state constraints, 'dynamical' discreteness is expected to manifest itself in the spectra of the relevant physical observables (the associated *kinematical operators* will be discussed in section 4.4). It is not clear whether discreteness of (physical) space would entail a discrete structure for *time* also, hence space–*time*, as there is no *a priori* notion of 'time' in quantum gravity. Instead, 'time' must also be defined operationally in terms of a 'clock

field' (see, e.g., [30–32]). Continuity or discreteness of physical space and time thus follows from the properties of the relevant 'measuring rod fields' and 'clock fields', and their spectra.

- *Conceptual issues.* Last but not the least: although we will have nothing new to say here on the grand conceptual issues of quantum gravity and quantum cosmology we wish to emphasize that these issues must be addressed and resolved by *all* approaches to quantum gravity. This comment concerns not only the difficult interpretational problems but also more technical aspects. Among the former, we would like to mention the problem of interpreting the 'wavefunction of the universe' and associated 'matrix elements' between different such wavefunctions, or the problem of constructing 'observables' and their physical interpretation; among the latter, there is the question of whether we have any right to expect the 'wavefunction of the universe' to be normalizable, or the related question of whether the familiar Hilbert space formalism of standard quantum mechanics is really the correct mathematical framework for quantum gravity[1]. Over the past few decades, there has been slow but steady progress on several fronts (see, e.g., [16, 30, 31, 33–37], and references therein), but it is probably fair to say that we are still far from fully understanding these issues.

2. Topics omitted

As we mentioned already, there are several recent developments and advances which we cannot cover here for lack of space, or lack of expertise on our side. We will list these and briefly comment on them below, but otherwise must refer readers to the pertinent 'inside' reviews [10–14] and the more recent original references for more information and other points of view.

2.1. Loop quantum cosmology

Symmetry reduced versions of LQG have recently been studied as models of quantum cosmology [38, 39]. Two main features stand out. The first is a new mechanism for the avoidance of the big bang singularity in the framework of mini-superspace models of quantum gravity[2]. This mechanism is based on the fact that the inverse scale factor is represented by an operator that stays bounded as the classical radius of the universe shrinks to zero. Alternatively, one might say that the effective discretization of the Hamiltonian constraint enables the quantum wavefunction to 'jump over' the singularity, and to continue to evolve past the singularity. However, it is not clear whether and how these models can be derived from full-fledged LQG, and whether singularity avoidance is also a property of the full theory. In fact, a very recent investigation [41] has revealed that the spectrum of the operator corresponding to the inverse volume is *not* bounded from above in full LQG. In addition, it is by no means excluded (and some would even say likely) that inhomogeneities, and possibly other degrees of freedom, will essentially alter the nature of the quantum state near the singularity. See also [42] for other comments.

The second new feature arises in applications of loop quantum cosmology to inflationary models. It is the possibility that inflation might be triggered and eventually stopped (gracefully) by gravity itself, via an intrinsically quantum gravitational mechanism [43]. Even if scalar

---

[1] In ordinary quantum mechanics, the Hilbert space formalism and the concomitant notion of 'unitarity' are tightly linked to the physical interpretation of the theory in terms of probabilities and their conservation in time. In the absence of an *a priori* notion of time (as in quantum gravity) it is therefore by no means evident whether these will remain the relevant concepts.

[2] Let us note that 'singularity avoidance mechanisms' may exist also in more conventional mini-superspace quantum geometrodynamics [40].

fields cannot be avoided altogether, an appealing feature here is that the inflaton potential engineering characteristic of most current models of inflation could become unnecessary. Perhaps less attractive (to an outsider) is that the regularization ambiguities of LQG feed through to the symmetry reduced models and lead to physical effects (see [44] for a recent discussion). Concerning the value of the Barbero–Immirzi parameter, obtained from black hole entropy calculations (see below), one might arrive at an interesting 'internal' consistency check by matching it with the duration of the inflationary period, which is also linked to this parameter in these models. It has, furthermore, been suggested that measurements of the CMB fluctuation spectrum at large angles might provide experimental tests of LQG, but it is not clear to us whether the predicted effects, if present, might not be explained in many other ways, too.

2.2. Microscopic origin of black hole entropy

The explanation of the Bekenstein–Hawking entropy of black holes [45, 46] in terms of microstates has been claimed as a success for both LQG [47, 48] and string theory [49]. The main achievement of string theory is that it not only explains the area law, but also predicts the factor 1/4 relating entropy and area; furthermore, the agreement has been shown to extend to the higher order curvature terms predicted by string theory (see, e.g., [50]). However, the argument requires a huge extrapolation in the string coupling constant, and is essentially limited to BPS-type extremal black holes. The LQG explanation, on the other hand, succeeds in implementing the condition for isolated horizons [51] in the quantum theory, and works for ordinary (Schwarzschild and Kerr) black holes, but requires a 'fit' for the Barbero Immirzi parameter to get the prefactor right. Although the two ansätze thus both reproduce the desired result, we are faced with something of a paradox here, as the two explanations seem almost impossible to reconcile, given the very different hypotheses underlying them—pure gravity on the one hand, and an exponentially growing spectrum of D-brane states on the other hand.

2.3. Spin foams

Attempts to overcome the difficulties with the Hamiltonian constraint have led to yet another development, *spin foam models* [52–54]. These can be regarded as spacetime versions of spin networks, to wit, evolutions of spin networks in 'time'. Mathematically, these models represent a generalization of spin networks, in the sense that group theoretical objects (holonomies, representations, intertwiners, etc) are attached not only to vertices and edges (links), but also to higher dimensional faces in a simplicial decomposition of spacetime. In addition, they may pave the way towards the construction of a physical inner product, which in turn is defined by the quantum dynamics. Many of the recent advances in this area concern purely topological theories, so-called BF models, where $F(A)$ is a field strength, and B is the Lagrange multiplier (tensor) field whose variation enforces $F(A) = 0$. The formalism can thus be nicely applied to $(2 + 1)$ gravity, which *is* topological. In $(3 + 1)$ dimensions one needs extra constraints in order to recover the propagating gravitational degrees of freedom back into the theory [55]. Interesting as they are, however, these developments have so far not shed much new light on the problems with the Hamiltonian constraint, or the constraint algebra, in our opinion. A derivation of spin foam models from the Hamiltonian formulation remains incomplete due to the complexity of the Hamiltonian constraint [56]. Hence, a decisive proof of the connection between spin foam models and the full Einstein theory and its canonical formulation still appears to be lacking. In contrast, it has even been suggested that these models may provide a possible 'way out' if the difficulties with the 'orthodox' Hamiltonian approach, which we follow here, should really prove insurmountable.

3. Old versus new variables: from metric to loops

3.1. Prelude: the metric (or dreibein) approach

There are many introductory texts to which we can refer readers for a more detailed treatment of canonical gravity (see, e.g., [12, 16, 31, 34, 57–60]), but let us nevertheless briefly review the traditional way of doing canonical quantum gravity, also called *geometrodynamics*. Our exposition in this section follows [61], whose notations and conventions we adopt in the remainder. We will be using the vierbein formalism with a vierbein $E_\mu{}^A$ and spacetime metric $G_{\mu\nu} = E_\mu{}^A E_\nu{}^B \eta_{AB}$ with flat (tangent space) metric $\eta_{AB} = \mathrm{diag}(-1, +1, +1, +1)$.[3] The vierbein is covariantly constant under a derivative which is covariant w.r.t. both spacetime diffeomorphisms and local Lorentz transformations, namely

$$D_\mu E_\nu{}^A := \nabla_\mu E_\nu{}^A - \omega_\mu{}^{AB} E_{\nu B} = 0, \tag{3.1}$$

where ∇_μ is the covariant derivative which involves only the Christoffel connection,

$$\nabla_\mu V_\nu := \partial_\mu V_\nu - \Gamma_{\mu\nu}{}^\rho V_\rho \tag{3.2}$$

and $\omega_{\mu AB}$ is the spin connection. It is a standard result that both $\Gamma_{\mu\nu}{}^\rho$ and $\omega_{\mu AB}$ can be explicitly solved for the vielbein from the above equation (in the absence of torsion).

As is customary in canonical gravity, and following the standard ADM prescription [57] we assume the spacetime manifold $\mathcal{M}$ to be globally hyperbolic. It then follows from a theorem of Geroch [62] that $\mathcal{M}$ can be foliated according to $\mathcal{M} = \Sigma \times \mathbb{R}$, with a spatial manifold Σ of fixed topology (and no boundary, for simplicity). Using letters $m, n, \ldots$ and $a, b, \ldots$ for curved and flat *spatial* indices, respectively, we choose the triangular gauge for the vierbein

$$E_\mu{}^A = \begin{pmatrix} N & N^a \\ 0 & e_m{}^a \end{pmatrix} \quad \Longrightarrow \quad G_{\mu\nu} = \begin{pmatrix} -N^2 + N_a N_a & N_n \\ N_m & g_{mn} \end{pmatrix}, \tag{3.3}$$

where $g_{mn} \equiv e_m{}^a e_{na}$ is the spatial metric. The vierbein components N and N_a are Lagrange multipliers ('lapse' and 'shift'); their variation yields the canonical constraints. In the remainder, we will freely convert spatial indices by means of the spatial dreibein and its inverse.

The canonical momenta are obtained in the standard fashion from the Einstein–Hilbert action

$$\Pi_a{}^m := \frac{\delta \mathcal{L}}{\delta \partial_t e_m{}^a} = \frac{1}{2} e e_b{}^m (K_{ab} - \delta_{ab} K), \tag{3.4}$$

where $e \equiv \det(e_m{}^a)$, and

$$K_{ab} := E_a{}^\mu E_b{}^\nu \nabla_\mu E_{\nu 0} = \omega_{ab0} \tag{3.5}$$

is the extrinsic curvature of Σ in $\mathcal{M}$ (expressed in terms of flat spatial indices, with $\omega_{ab0} \equiv E_a{}^\mu \omega_{\mu b0}$), and $K \equiv K_{aa}$. The inverse formula reads

$$K_{ab} = \frac{2}{e} \left(\Pi_{ab} - \frac{1}{2} \delta_{ab} \Pi \right), \qquad (\Pi \equiv \Pi_{aa}). \tag{3.6}$$

From the symmetry $K_{ab} = K_{ba}$ we immediately deduce the Lorentz constraint

$$L_{ab} = e_{m[a} \Pi_{b]}{}^m \approx 0, \tag{3.7}$$

which is also the canonical generator of spatial rotations on the dreibein ($\approx$ means 'weakly zero' [59]). The canonical Hamiltonian (really Hamiltonian density) is

$$H = \partial_t e_a{}^m \Pi_m{}^a - \mathcal{L} = N H_0 + N_a H_a, \tag{3.8}$$

---

[3] Modulo some dimension dependent factors, the results described in this subsection are valid in any dimension.

with the diffeomorphism constraint

$$H_a \equiv D_m \Pi_a{}^m \approx 0, \tag{3.9}$$

and the Hamiltonian constraint (*alias* the scalar constraint)

$$H_0 \equiv e^{-1}\left(\Pi_{ab}\Pi_{ab} - \tfrac{1}{2}\Pi^2\right) - eR^{(3)} = \tfrac{1}{4}e(K_{ab}K_{ab} - K^2) - eR^{(3)} \approx 0, \tag{3.10}$$

where $R^{(3)}$ is the spatial Ricci scalar. The canonical equal time (Poisson) brackets are

$$\{e_{ma}(\mathbf{x}), \Pi_b{}^n(\mathbf{y})\} = \delta_{ab}\delta_m^n\delta^{(3)}(\mathbf{x}, \mathbf{y}), \tag{3.11}$$

with the other brackets vanishing (for the canonical variables with the indices in the indicated positions). Canonical quantization in the 'position space representation' now proceeds by representing the dreibein as a multiplication operator, and the canonical momentum by the functional differential operator

$$\Pi_a{}^m(\mathbf{x}) = \frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta e_m{}^a(\mathbf{x})}. \tag{3.12}$$

With these replacements, the classical constraints are converted to quantum constraint operators which act on suitable wave functionals. The diffeomorphism and Lorentz constraints become

$$H_a(\mathbf{x})\Psi[e] = 0, \qquad L_{ab}(\mathbf{x})\Psi[e] = 0. \tag{3.13}$$

They will be referred to as 'kinematical constraints' throughout. Dynamics is generated via the Hamiltonian constraint, the *Wheeler–DeWitt (WDW) equation* [63–66]:

$$H_0(\mathbf{x})\Psi[e] = 0. \tag{3.14}$$

It is straightforward to include matter degrees of freedom, in which case the constraint operators and the wave functional $\Psi[e, \ldots]$ depend on further variables (indicated by dots). The functional $\Psi[e, \ldots]$ is sometimes referred to as the 'wavefunction of the universe', and is supposed to contain the complete information about the universe 'from beginning to end'. A good way to visualize Ψ is to think of it as a film reel; 'time' and the illusion that 'something happens' emerge only when the film is played.

The substitution (3.12) turns the Hamiltonian constraint into a highly singular functional differential equation, which most likely cannot be made mathematically well defined in this form, even allowing for certain 'renormalizations'. Here we do not wish to belabour the well-known difficulties, both mathematical and conceptual, which have stymied progress with the WDW equation for over 40 years, see e.g. [31, 34, 58][4]. We should point out the (surely well-known) fact that, at least formally and ignoring subtle points of functional analysis, solutions to both the Lorentz constraint and the diffeomorphism constraint can be rather easily obtained by integrating suitably densitized invariant combinations of the spatial dreibein (or metric) and curvature components, and the matter 'position variables' over the spatial three-manifold Σ. The associated wave functionals Ψ are then automatically functionals of diffeomorphism classes of spatial metrics. In saying this, it must be stressed, however, that geometrodynamics has so far not succeeded in constructing a suitable scalar product and an appropriate Hilbert space of wave functionals.

The absence of a suitable Hilbert space is often invoked by LQG proponents as an argument against the geometrodynamics approach. While LQG is certainly 'ahead' in that it does succeed in constructing a Hilbert space (more on this later), we should like to emphasize that in all approaches there looms the larger conceptual problem of whether conventional quantum-mechanical concepts are indeed sufficient for quantum gravity [31, 33]. For instance,

---

[4] One of the inventors of (3.14) has been overheard aptly referring to it as *'that damned equation…'*.

even if we can compute matrix elements of wave function(al)s, we still have no idea what their correct physical interpretation is in the context of quantum cosmology, or whether the normalizability of the wavefunction of the universe is a necessary condition. In other words, we do not know whether these cherished concepts may not have to be amended or abandoned in the final theory. For this reason, we believe that besides the emphasis on mathematical rigour, it is equally important to develop some physical intuition for the states one is dealing with, and in this regard, we do not think that geometrodynamics lags behind. From this point of view, it appears to us that, beyond the technical subtleties, the kinematical constraints are not the real problem of quantum gravity. The core difficulties of canonical quantum gravity are all connected in one way or another to the Hamiltonian constraint—irrespective of which canonical variables are used.

3.2. Ashtekar's new variables

Much of the initial excitement over Ashtekar's discovery [67] of new canonical variables was due to the change of perspective they bring about, which fuelled hopes that they might alleviate some of the longstanding unsolved problems of quantum gravity. Let us therefore first describe what they are, and how they are obtained. There is an *ab initio* derivation based on the addition of a term $\propto \int E \wedge E \wedge R$ to the Einstein–Hilbert action [68, 69] (this term vanishes upon use of the Bianchi identity and the equations of motion for the spin connection with vanishing torsion), but we will skip this step here. Instead consider the combination[5]

$$A_{ma} := -\tfrac{1}{2}\epsilon_{abc}\omega_{mbc} + \gamma e^{-1}\left(\Pi_{ma} - \tfrac{1}{2}e_{ma}\Pi\right) = -\tfrac{1}{2}\epsilon_{abc}\omega_{mbc} + \gamma K_{ma}, \tag{3.15}$$

where ω_{mbc} is the *spatial* spin connection. The parameter $\gamma \neq 0$ is referred to as the 'Barbero–Immirzi parameter' in the LQG literature [70, 71]. Classically, γ has no physical significance, but is believed to become physically relevant upon quantization, for instance, by setting the scale for the fundamental areas and volumes (in this sense it is somewhat analogous [72] to the θ parameter of QCD). One can now show that [67]

$$\{A_{ma}(\mathbf{x}), A_{nb}(\mathbf{y})\} = 0, \qquad \left\{\tilde{E}_a{}^m(\mathbf{x}), \tilde{E}_b{}^n(\mathbf{y})\right\} = 0, \qquad \left\{A_m{}^a(\mathbf{x}), \tilde{E}_b{}^n(\mathbf{y})\right\} = \gamma\delta_m^n\delta_b^a\delta^{(3)}(\mathbf{x}, \mathbf{y}), \tag{3.16}$$

where the canonically conjugate variable to A_{ma} is the *inverse densitized spatial dreibein*

$$\tilde{E}_a{}^m := e e_a{}^m \quad \Longrightarrow \quad \tilde{E} = e^2, \tag{3.17}$$

with $e \equiv \det e_m{}^a$ and $\tilde{E} \equiv \det \tilde{E}_a{}^m$. The parameter γ is often eliminated from these brackets by absorbing it into the definition of $\tilde{E}_a{}^m$.

To rewrite the constraints in terms of the new variables, we first observe that the covariant constancy of the spatial dreibein and the Lorentz constraint imply the Gauss constraint:

$$D_m\tilde{E}_a{}^m \equiv \partial_m\tilde{E}_a{}^m + \epsilon_{abc}A_m{}^b\tilde{E}^{cm} \approx 0. \tag{3.18}$$

Defining the field strength

$$\begin{aligned}F_{mna} &:= \partial_m A_{na} - \partial_n A_{ma} + \epsilon_{abc}A_{mb}A_{nc}\\ &= -\tfrac{1}{2}\epsilon_{abc}R_{mnbc} + \gamma(D_m K_{na} - D_n K_{ma}) + \gamma^2\epsilon_{abc}K_m{}^b K_n{}^c,\end{aligned} \tag{3.19}$$

it follows that the diffeomorphism constraint takes the form

$$\tilde{E}_a{}^m F_{mna} \approx 0. \tag{3.20}$$

---

[5] We alert readers that our notational conventions differ from the ones used in most of the LQG literature, where $a, b, \ldots$ denote curved and $i, j, \ldots$ flat spatial indices.

Furthermore,

$$\epsilon_{abc}\tilde{E}_a{}^m\tilde{E}_b{}^n F_{mnc} = -\gamma^2\left(\Pi_{ab}\Pi_{ab} - \tfrac{1}{2}\Pi^2\right) - e^2 R^{(3)}$$
$$= -\gamma^2 e\mathcal{H}_0 - \tfrac{1}{4}(1+\gamma^2)e^2(K_{ab}K_{ab} - K^2). \qquad (3.21)$$

These relations immediately suggest interpreting $A_m{}^a$ as a *gauge connection* for the gauge group $SO(3)$ of spatial rotations (for $D = 2+1$ gravity, this group is replaced by its non-compact form $SO(1,2)$ [73]). Accordingly, the new variables are conveniently rewritten as [67]

$$A_{m\alpha\beta} \equiv A_{ma}\tau^a_{\alpha\beta}, \qquad\qquad (3.22)$$

where τ^a are the Pauli matrices.

For the special choice $\gamma = \pm i$ [67] the second term on the rhs of (3.21) drops out, and—save for an extra density factor of e—the Hamiltonian constraint is expressed entirely in terms of the new canonical variables, and furthermore depends on them *polynomially*[6]. In other words, this particular choice allows us to combine the diffeomorphism and Hamiltonian constraints, which are schematically of the form '$(\partial + \omega)\Pi \approx 0$' and '$(\partial\omega + \omega^2 + \Pi^2) \approx 0$', respectively, into a single expression in terms of the connection $A = \omega \pm i\Pi$, and its canonically conjugate variable. Moreover, for this choice of γ the connection $A_{ma}^{(\pm)}$ is nothing but the pullback of the four-dimensional spin connection to the spatial hypersurface Σ, with the two signs corresponding to the two chiralities (indicated by superscripts $^{(\pm)}$)

$$\omega_{mAB}\gamma^{AB} = \omega_{mab}\gamma^{ab} + 2\omega_{ma0}\gamma^a\gamma^0 = A_{ma}^{(+)}\gamma^0\gamma^a(1+\gamma^5) + A_{ma}^{(-)}\gamma^0\gamma^a(1-\gamma^5), \qquad (3.23)$$

cf equation (3.5). Equivalently, the variables $A_{ma}^{(\pm)}$ are associated with what, in a Euclidean formulation, would be the selfdual and anti-selfdual parts of the spin connection, respectively,

$$A_{ma}^{(\pm)} = \omega_{ma0}^{(\pm)} \qquad \text{with} \qquad \omega_{mAB}^{(\pm)} := \frac{1}{2}\left(\omega_{mAB} \pm \frac{i}{2}\epsilon_{ABCD}\omega_m{}^{CD}\right). \quad (3.24)$$

This is, therefore, also the natural choice for γ when one considers coupling gravity to chiral fermions. In fact, one of the authors (HN) was first enticed into learning about Ashtekar's variables when realizing that they simplify the calculation of the constraint algebra of supergravity considerably [75, 76]. This is because the local supersymmetry constraint (i.e. the time component of the Rarita–Schwinger equation) always contains a factor $D_{[m}(A)\psi_{n]}$, where ψ_n is the gravitino, and A is just the Ashtekar connection (of course with $\gamma = \pm i$). More succinctly, in supergravity, the commutation relations (3.16), as well as the polynomiality of the constraints must necessarily hold, if the commutator of two local supersymmetry constraints is to close into the scalar and diffeomorphism (and possibly other) constraints.

From the esthetical point of view $\gamma = \pm i$ is therefore clearly the preferred choice. Nevertheless, this value has been abandoned in most of the recent LQG literature, because there is a major difficulty with it: the phase space of general relativity must be *complexified* with imaginary or complex γ. To recover the *real* phase space of general relativity and to ensure that real initial data evolve into real solutions, suitable reality conditions must be imposed. This is straightforward to achieve for the classical theory—after all, we have merely changed the variables, not the theory itself—but not so for the quantum theory. There, the complexification poses subtle problems concerning the definition and imposition of appropriate hermiticity conditions on the states and operators, and no consensus has been reached so far on how to circumvent these difficulties (or on whether they can be circumvented at all).

---

[6] More recently it has, however, been appreciated that the extra density factor in (3.21) is a major source of problems for a background-independent quantization, since it makes the Hamiltonian an object of density weight two; for details see [74].

For this reason one now usually takes γ to be *real*. In this case, no problem arises with reality of solutions in either the classical or the quantum theory, but the polynomiality of the constraints, and hence one of the most attractive features of the new variables, is lost. This is because the extra term in (3.21) no longer vanishes, but must be subtracted from both sides to recover the correct WDW Hamiltonian. Accordingly, one must express the extra contributions in terms of the new canonical variables $A_m{}^a$ and $\tilde{E}_a{}^m$ via (3.15) and (3.17), and this in turn requires expressing the original dreibein as well as the extrinsic curvatures in terms of the new canonical variables. For a while this was regarded as a chief obstacle, until a way to solve it was discovered by Thiemann [74]. To this aim let us first introduce the volume associated with a region $\Omega \subset \Sigma$ (considered as a phase-space variable)[7]

$$V(\Omega) = \int_\Omega \mathrm{d}^3 x \, e = \int_\Omega \mathrm{d}^3 x \, \sqrt{\tilde{E}} \equiv \int_\Omega \mathrm{d}^3 x \, \sqrt{\frac{1}{3!} \epsilon_{mnp} \epsilon^{abc} \tilde{E}_a{}^m \tilde{E}_b{}^n \tilde{E}_c{}^p}. \quad (3.25)$$

Writing $V \equiv V(\Sigma)$, we first use the substitution

$$e_m{}^a(\mathbf{x}) = \epsilon_{mnp} \epsilon^{abc} \tilde{E}^{-1/2} \tilde{E}_b^n \tilde{E}_c^p(\mathbf{x}) = \frac{1}{4\gamma} \left\{ A_m{}^a(\mathbf{x}), V \right\} \quad (3.26)$$

to recover the spatial dreibein. The second trick is to eliminate the extrinsic curvature using a doubly nested bracket. The first bracket is introduced by rewriting

$$K_m{}^a(\mathbf{x}) = \frac{1}{\gamma} \left\{ A_m{}^a(\mathbf{x}), \bar{K} \right\} \qquad \text{where} \qquad \bar{K} \equiv \bar{K}(\Sigma) := \int_\Sigma \mathrm{d}^3 x \, K_m{}^a \tilde{E}_a{}^m. \quad (3.27)$$

The second bracket comes in through identity

$$\bar{K}(\mathbf{x}) = \frac{1}{\gamma^{3/2}} \left\{ \frac{\tilde{E}_a{}^m \tilde{E}_b{}^n}{\sqrt{\tilde{E}}} \epsilon^{abc} F_{mnc}(\mathbf{x}), V \right\}. \quad (3.28)$$

Last, we need another dreibein factor to convert the curved index m on $K_m{}^a$ to a flat one, for which we must use (3.26) once again. We will postpone writing out the Hamiltonian constraint in terms of these multiple brackets until section 5.2. But let us note already here that the above relations supply the 'tool kit' also for transcribing any given matter Hamiltonian in terms of the new variables.

For $\gamma \neq \pm\mathrm{i}$ the connection is no longer the pullback of a four-dimensional spin connection [77]. Although this does not immediately lead to problems with spacetime covariance, we will see in section 4.6 that the problem comes back when one considers coupling the theory to fermionic matter degrees of freedom.

3.3. The connection representation

Independently of the choice of γ, the reformulation of canonical gravity in terms of connection variables opens many new avenues, in particular the use of concepts, tools and techniques from Yang–Mills theory. Early attempts at quantization were based on the *connection representation* with the original Ashtekar connection, i.e. $\gamma = \pm\mathrm{i}$. Although these were ultimately not successful, let us nonetheless briefly summarize them here. In this scheme, one represents the connection $A_m{}^a$ by a multiplication operator, and sets

$$\tilde{E}_a{}^m(\mathbf{x}) = \frac{\hbar}{\mathrm{i}} \frac{\delta}{\delta A_m{}^a(\mathbf{x})}. \quad (3.29)$$

The WDW functional depending on the spatial metric (or dreibein) is accordingly replaced by a functional $\Psi[A]$ living on the space of connections (modulo gauge transformations).

---

[7] The ϵ-symbol is always the invariant *tensor density*: $\epsilon_{mnp} = e^{-1} e_m{}^a e_n{}^b e_p{}^c \epsilon_{abc}$, i.e. assumes the values $0, \pm 1$.

The price one pays is that this representation is much harder to 'visualize' because the spatial metric is no longer represented by a simple multiplication operator, but must now be determined from the operator for the inverse densitized metric

$$gg^{mn}(\mathbf{x}) = -\hbar^2 \frac{\delta}{\delta A_m{}^a(\mathbf{x})} \frac{\delta}{\delta A_n{}^a(\mathbf{x})}. \tag{3.30}$$

Even if one ignores the clash of functional differential operators at coincident points, finding suitable states and computing their expectation values is obviously not an easy task (and has not been accomplished so far). Similarly, the spatial volume density is obtained from

$$g(\mathbf{x}) = \tilde{E}(\mathbf{x}) = \frac{\mathrm{i}\hbar^3}{6} \epsilon^{abc} \epsilon_{mnp} \frac{\delta}{\delta A_m{}^a(\mathbf{x})} \frac{\delta}{\delta A_n{}^b(\mathbf{x})} \frac{\delta}{\delta A_p{}^c(\mathbf{x})}. \tag{3.31}$$

Again this operator is very singular. Equations (3.30) and (3.31) provide a first glimpse of the difficulties that the formalism has in finding semi-classical states and thereby establishing a link between the quantum theory and classical smooth spacetime geometry.

For the quantum constraints the replacement of the metric by connection variables leads to a Hamiltonian which is simpler than the original WDW Hamiltonian, but still very singular. Allowing for an extra factor of e (and assuming $e \neq 0, \infty$) the WDW equation becomes

$$\epsilon^{abc} F_{mna}(A(\mathbf{x})) \frac{\delta}{\delta A_m{}^b(\mathbf{x})} \frac{\delta}{\delta A_n{}^c(\mathbf{x})} \Psi[A] = 0. \tag{3.32}$$

Here we have adopted a particular ordering, which however is by no means singled out. No viable solutions to this constraint have been found, but there is at least one interesting solution if one allows for a non-vanishing cosmological constant Λ. Namely, using an ordering opposite to that above, and including a term Λg with the volume density (3.31), the WDW equation reads

$$\epsilon_{abc} \frac{\delta}{\delta A_m{}^a(\mathbf{x})} \frac{\delta}{\delta A_n{}^b(\mathbf{x})} \left(F_{mnc}(A(\mathbf{x})) - \frac{\mathrm{i}\hbar\Lambda}{6} \epsilon_{mnp} \frac{\delta}{\delta A_p{}^c(\mathbf{x})} \right) \Psi_\Lambda[A] = 0. \tag{3.33}$$

This is solved by

$$\Psi_\Lambda[A] = \exp\left(\frac{\mathrm{i}}{\hbar\Lambda} \int_\Sigma \mathrm{d}^3\mathbf{x}\, \mathcal{L}_{CS}(A) \right), \tag{3.34}$$

with the Chern–Simons Lagrangian $\mathcal{L}_{CS} = A \wedge \mathrm{d}A + \mathrm{i}A \wedge A \wedge A$ (actually, Ψ_Λ is already annihilated by the operator in parentheses, so the first factor in the Hamiltonian constraint operator (3.33) is not 'needed' for this result). In the literature this state is known as the Kodama state [78], but the solution had been known for a long time in Yang–Mills theory [79], where however it has rather unusual physical properties [80]. The difficulties with this solution have been much discussed recently [81, 80]; an obvious one concerns the flat space limit $\Lambda \to 0$ (idem for the 'semiclassical' limit $\hbar \to 0$).

What happens when we choose γ to be different from $\pm\mathrm{i}$, and real in particular? As we explained already, the extra term in (3.21) then no longer vanishes, must be dealt with separately [74]. The nice polynomial form of the Hamiltonian constraint operator (3.32) is lost. When implementing the translation rules at the end of the foregoing subsection in the connection formulation, one finds that the new Hamiltonian is not significantly simpler any more than the original one of geometrodynamics in terms of metric or dreibein variables.

3.4. From connections to holonomies

The *loop representation* is an attempt to overcome the difficulties with the connection representation which we sketched above. The transition between the connection and the

loop representation was originally obtained via the *loop transform*, which can be thought of as a kind of functional Fourier transform [82]. We will not describe that construction here, but turn immediately to the formulation in terms of holonomies, on which the modern formulation of LQG—spin networks and spin foams—are based.

Whereas in the connection representation one works with functionals $\Psi[A]$ which are supported 'on all of Σ', one now switches to the *holonomies* as the basic variables. These are gauge covariant functionals supported on one-dimensional links, or 'edges', which we will designate by e (following established LQG notation). For a given edge, i.e. some (open) curve embedded in Σ, we set

$$h_e[A] = \mathcal{P}\exp\int_e A_m\,\mathrm{d}x^m, \qquad \text{with} \quad A \equiv A^a\tau_a. \tag{3.35}$$

Hence, $h_e[A]$ is a matrix-valued functional. The holonomy transforms under the action of $SU(2)$ at each end of the edge e:

$$h_e[A] \to h_e^g[A] = g(e(0))h_e[A]g^{-1}(e(1)), \qquad \text{with} \quad g(e(0)),\ g(e(1)) \in SU(2). \tag{3.36}$$

For the remainder it is important that the holonomies are to be regarded as variables *in their own right*, subject to these transformation properties (so in some sense one can 'forget' about the original connection A defining the holonomy). The *distributional nature* of the holonomy is not only evident from its singular support (on a line rather than all of Σ), but also from the fact that we do not assume $h_e[A]$ to be close to the identity if the edge e is 'small' (this terminology has to be used with due care, as there is no *a priori* measure that tells us when e is 'small', but we can still imagine making it 'small' by chopping e into as many 'subedges' as we like). The fact that the typical field configuration is generically a distribution rather than some smooth function is well known from constructive quantum field theory [83].

The holonomies are taken to transform in $SU(2)$ representations ρ_{j_e} of *arbitrary spin* $j_e = \frac{1}{2}, 1, \frac{3}{2}, \ldots$ for each link e (with the convention that $j_e = 0$ means that there is no edge). We will denote such a spin-j_e valued holonomy by

$$\left(\rho_{j_e}(h_e[A])\right)_{\alpha\beta}, \tag{3.37}$$

with indices $\alpha, \beta, \ldots$ as appropriate for the representation at hand. To make the notation less cumbersome, we will occasionally suppress ρ_{j_e} and the representation labels, and simply denote the above matrix as $(h_e[A])_{\alpha\beta}$.

To define the conjugate variable, we recall that the area element for the spatial manifold Σ can be expressed as a Lorentz vector (i.e. with flat $SO(3)$ indices) via

$$\mathrm{d}F_a := \epsilon_{abc}\theta^b \wedge \theta^c \qquad \text{with} \quad \theta^a \equiv \mathrm{d}x^m\,e_m{}^a. \tag{3.38}$$

Happily, this can be nicely rewritten in terms of the new canonical variables

$$\mathrm{d}F_a = \epsilon_{abc}e_n{}^b e_p{}^c\,\mathrm{d}x^n \wedge \mathrm{d}x^p = \epsilon_{mnp}\tilde{E}_a^m\,\mathrm{d}x^n \wedge \mathrm{d}x^p. \tag{3.39}$$

As the conjugate variable to $h_e[A]$ one takes the 'flux' vector

$$F_S^a[\tilde{E}] := \int_S \mathrm{d}F^a \tag{3.40}$$

through any two-dimensional surface S embedded in Σ. There is also a smeared version of this variable, with a test function f_a to soak up the free index a, which reads

$$F_S[\tilde{E}, f] := \int_S f^a\epsilon_{abc}e_n{}^b e_p{}^c\,\mathrm{d}x^n \wedge \mathrm{d}x^p = \int_S f_a\epsilon_{mnp}\tilde{E}^{ma}\,\mathrm{d}x^n \wedge \mathrm{d}x^p. \tag{3.41}$$

We note that the standard notation for this variable in the LQG literature is $E[S, f]$, but we prefer the one above because it is in parallel with the notation for the holonomy itself.
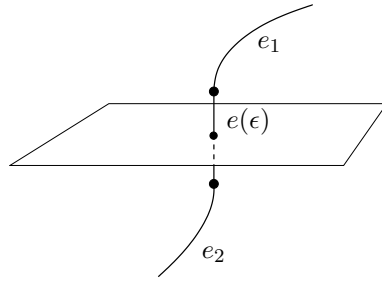
Figure 1. Setup used for the computation of the bracket (3.42). In the limit $\epsilon \to 0$ the edge $e(\epsilon)$ shrinks to zero and the two nodes just above and below the surface coincide.

Both $F_S[\tilde{E}, f]$ and $F_S^a[\tilde{E}]$ are *distributional* in the sense that they are supported on a two-dimensional submanifold of Σ.

To compute the Poisson brackets between the new canonical variables introduced above, we consider a surface S and an edge e that 'pierces' S at the point P (if e does not intersect S, the bracket simply vanishes). We next subdivide this edge into three pieces as shown in figure 1: two subedges e_1 and e_2, with associated holonomies $h_{e_1}^\epsilon$ and $h_{e_2}^\epsilon$, which touch S only in the limit $\epsilon \to 0$, and a third 'infinitesimal' edge $e(\epsilon)$ intersecting S, for which the path-ordered exponential can be approximated by the linear term. Then

$$\{(h_e[A])_{\alpha\beta}, F_S[\tilde{E}, f]\}$$
$$= \lim_{\epsilon \to 0} \left[(h_{e_1}^\epsilon[A])_{\alpha\gamma} \left\{ \int_{e(\epsilon)} \mathrm{d}x^m \, A_m{}^a(\mathbf{x}) \tau_{\gamma\delta}^a, \int_S \mathrm{d}y^n \wedge \mathrm{d}y^p \, \epsilon_{npq} f_b(\mathbf{y}) \tilde{E}^{bq}(\mathbf{y}) \right\} (h_{e_2}^\epsilon[A])_{\delta\beta} \right]$$
$$= \iota(e, S) \gamma f_a(P) \left(h_{e_1}[A] \tau^a h_{e_2}[A] \right)_{\alpha\beta}. \tag{3.42}$$

Here, the intersection number

$$\iota(e, S) := \int_e \mathrm{d}x^m \int_S \mathrm{d}y^n \, \mathrm{d}y^p \epsilon_{mnp} \delta^{(3)}(\mathbf{x}, \mathbf{y}) = \pm 1 \ \text{ or } \ 0 \tag{3.43}$$

encodes the information on how e intersects S in a coordinate independent way. The integral is equal to ± 1, depending on the orientation of e and S, if e intersects S transversally. When the edge e intersects or touches S tangentially, a little care must be exercised; one then finds that $\iota(e, S) = 0$, and the above bracket vanishes [84]. The fact that $h_e[A]$ and $F_S[A, f]$ are, respectively, supported on one-dimensional and two-dimensional subsets of Σ is thus precisely what is required to perform the integral over the three-dimensional δ-function. Evidently, the integral is ill-defined when an entire segment of e lies within S; for a discussion of how to deal with this difficulty, see [84].

4. Quantization: kinematics

Having determined the classical canonical variables, one would now like to promote them to quantum operators obeying the appropriate commutation relations. The essential assumption of LQG is that this quantization should take place at the level of the bounded Hermitian operators $h_e[A]$ rather than the connection A itself. This is analogous to ordinary quantum mechanics, when one replaces the Heisenberg operators x and p by Weyl operators $\mathrm{e}^{\mathrm{i}x}$ and $\mathrm{e}^{\mathrm{i}p}$; the spin network representation actually uses the analogue of a hybrid formulation with x and $\mathrm{e}^{\mathrm{i}p}$. The Stone–von Neumann theorem [85–87] is usually invoked to argue that it makes no difference whether one quantizes the Heisenberg or the Weyl algebra, i.e. that these

quantizations are equivalent. The theorem does require, however, that the representations which are used are 'weakly continuous'. In the case of ordinary quantum mechanics, for example, this means that matrix elements of the operators corresponding to $e^{i\alpha x}$ and $e^{i\beta p}$ are smooth functions of the parameters α and β. In LQG the representations of operators do not satisfy this requirement[8].

The failure of operators to be weakly continuous can, as we will see, be traced back to the very special choice of the scalar product (4.7) below, which LQG employs to define its kinematical Hilbert space $\mathcal{H}_{kin}$. This Hilbert space does not admit a countable basis, hence is *non-separable*, because the set of all spin network graphs in Σ is uncountable, and non-coincident spin networks are orthogonal w.r.t. (4.7). Therefore, any operation (such as a diffeomorphism) which moves around graphs continuously corresponds to an uncountable sequence of mutually orthogonal states in $\mathcal{H}_{kin}$. That is, no matter how 'small' the deformation of the graph in Σ, the associated elements of $\mathcal{H}_{kin}$ always remain a finite distance apart, and consequently, the continuous motion in 'real space' gets mapped to a highly discontinuous one in $\mathcal{H}_{kin}$. Although unusual, and perhaps counter-intuitive, as they are, these properties constitute a cornerstone for the hopes that LQG can overcome the seemingly unsurmountable problems of conventional geometrodynamics: if the representations used in LQG were equivalent to the ones of geometrodynamics, there would be no reason to expect LQG not to end up in the same quandary.

It is perhaps also instructive to contrast the LQG approach with the standard lattice approach to field theory. In the lattice approach, all quantities depend explicitly on the lattice spacing (i.e. the regulating parameter). In the limit in which the lattice spacing is taken to zero, one recovers the continuum results (all expectation values are smooth functions of the regulating parameter). In the LQG approach the 'discretuum' is instead built in by the very construction of the scalar product, rather than by introducing a regulating parameter. LQG is therefore quite different from 'quantum gravity on the lattice'. While the radical modification underlying LQG has certain appealing properties, it makes it hard to recover long-distance physics and the usual notion of continuity.

In short, one thus implements the quantization not by the replacement (3.29) but rather by promoting the above Poisson bracket (3.42) to a quantum commutator:

$$[(h_e[A])_{\alpha\beta}, \hat{F}_S[\tilde{E}, f]] = i\hbar l_P^2 \gamma \iota(e, S) f^a(P)(h_{e_1}[A]\tau_a h_{e_2}[A])_{\alpha\beta} \tag{4.1}$$

or equivalently,

$$[(h_e[A])_{\alpha\beta}, \hat{F}_S^a[\tilde{E}]] = i\hbar l_P^2 \gamma \iota(e, S)(h_{e_1}[A]\tau^a h_{e_2}[A])_{\alpha\beta}. \tag{4.2}$$

On the spin network representation the holonomies $h_e[A]$ will be represented as multiplication operators; the action of the canonically conjugate operators will be explained below. The inequivalence of LQG quantization with Fock-space quantization arises through a special scalar product, to be discussed below.

4.1. The Hilbert space of spin networks

After defining the basic variables in which the theory should be quantized, the next step is to choose a Hilbert space in which the operators act. Starting from this space, one should construct a Hilbert space of physical states, i.e. space of states for which all constraints hold. The initial Hilbert space of LQG is the space of spin networks. While the Gauss constraint

---

[8] This is also the reason why the kinematical Hilbert space employed in loop quantum cosmology is different from the standard one already for a *finite* number of degrees of freedom [39]. When the number of degrees of freedom is infinite (as in quantum field theory), the Stone–von Neumann theorem anyhow does not apply.

is easily solved in this space it turns out that a solution of the diffeomorphism constraints lies *outside* this 'naive' initial space, and one is forced to introduce a larger space.

The intuitive idea behind spin networks is that the geometry at the Planck scale is foam-like. Physical gravitational and matter degrees of freedom are excited only on so-called spin networks, i.e. one-dimensional 'edges' or 'links', and the vertices connecting them. Geometries which look smooth at large scales are supposed to arise only from complicated spin network states with many edges. In order to find the Hilbert space of these objects, one has to find a basis of wavefunctions over the configuration space, which associate a complex number to each and every configuration of the gauge connection. LQG makes use of wavefunctions which have singular support in the sense that they only probe the gauge connection on one-dimensional networks embedded in the three-dimensional spatial hypersurface Σ, which is a (not necessarily differentiable) manifold, i.e. can be mapped out by local charts. This three-dimensional 'reference space', or 'carrier space' of the spin networks, does not carry any physical metric. LQG makes occasional use of local coordinates, or fiducial background metrics for certain intermediate steps in the construction, but physical quantities must not depend on such background data. We will here avoid their use as far as possible. Let us emphasize again that the 'discreteness' of the spin networks does not correspond to a naive discretization of space. Rather, the underlying continuum, on which the spin networks 'float', the spatial manifold Σ, is still present. As we will see, the setup furthermore requires the *a priori* exclusion of infinite spin networks, which might contain Cantor-like or 'fractal' sets.

By definition, each network is a (not necessarily connected) graph Γ embedded in Σ and consisting of *finitely many* edges $e_i \in \Gamma$ and vertices $v \in \Gamma$. The edges are connected at the vertices. Each edge e carries a holonomy $h_e[A]$ of the gauge connection A (this connection does not have to be smooth on the edge). The wavefunction on the spin network over the graph Γ can be written as

$$\Psi_{\Gamma,\psi}[A] = \psi\big(h_{e_1}[A], h_{e_2}[A], \dots\big), \tag{4.3}$$

where the ψ is some function of the basic holonomies associated with the edges $e \in \Gamma$. If, in addition, the wavefunction ψ is invariant under arbitrary $SU(2)$ gauge transformations it satisfies the *Gauss constraint*, and vice versa. A gauge invariant function ψ thus takes care of joining the collection of holonomies into an $SU(2)$-invariant complex number by contracting all $SU(2)$-indices of the holonomies with invariant tensors 'located' at the vertices v (see figure 2). The basic building blocks of the spin network wavefunctions are therefore expressions of the following type. A 3-valent vertex connects three edges according to

$$\psi[\text{figure 2}(a)] = \big(\rho_{j_1}(h_{e_1}[A])\big)_{\alpha_1\beta_1} \big(\rho_{j_2}(h_{e_2}[A])\big)_{\alpha_2\beta_2} \big(\rho_{j_3}(h_{e_3}[A])\big)_{\alpha_3\beta_3} C^{j_1 j_2 j_3}_{\beta_1\beta_2\beta_3} \dots, \tag{4.4}$$

where dots represent the remainder of the graph. While the contraction is obviously unique if only two or three edges meet at a vertex, there may be more and independent choices for vertices of valence four and higher, depending on the way in which the edges are connected with the Clebsch–Gordan coefficients. For this reason, any given spin network will, in general, admit several independent wavefunctions of the above type. As an example, consider a 4-valent vertex: one first has to decide on how to pair the edges into two groups of two. One such choice leads to, e.g.

$$\psi[\text{figure 2}(b)] = \big(\rho_{j_1}(h_{e_1}[A])\big)_{\alpha_1\beta_1} \big(\rho_{j_2}(h_{e_2}[A])\big)_{\alpha_2\beta_2} \big(\rho_{j_3}(h_{e_3}[A])\big)_{\alpha_3\beta_3} \big(\rho_{j_4}(h_{e_4}[A])\big)_{\alpha_4\beta_4}$$
$$\times C^{j_1 j_2 k}_{\beta_1\beta_2\beta} C^{j_3 j_4 k}_{\beta_3\beta_4\beta} \dots. \tag{4.5}$$

In these equations $C^{j_1 j_2 k}_{\beta_1\beta_2\beta}$ are intertwiners (Clebsch–Gordan coefficients). In (4.5) the intermediate spin k can be freely chosen in accordance with the standard rules for the vector addition of angular momenta: $|j_1 - j_2| \leqslant k \leqslant j_1 + j_2$. In other words, we can graphically
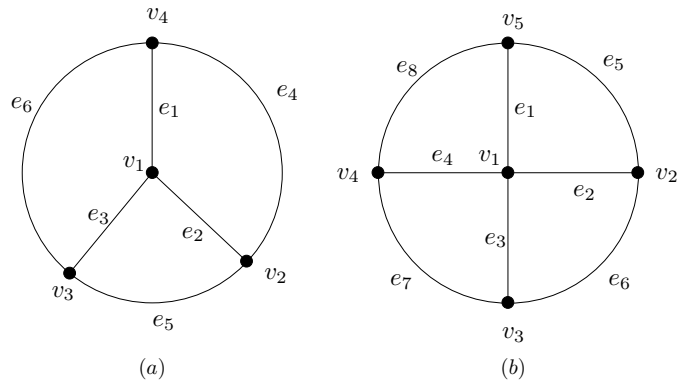
Figure 2. Examples of spin network states. For the 3-valent vertices on the left, the three incoming edges at each vertex are connected by a Clebsch–Gordan coefficient. For the 4-valent vertex on the right, one has to decide on a given way to construct a higher-order invariant tensor from two Clebsch–Gordan coefficients.
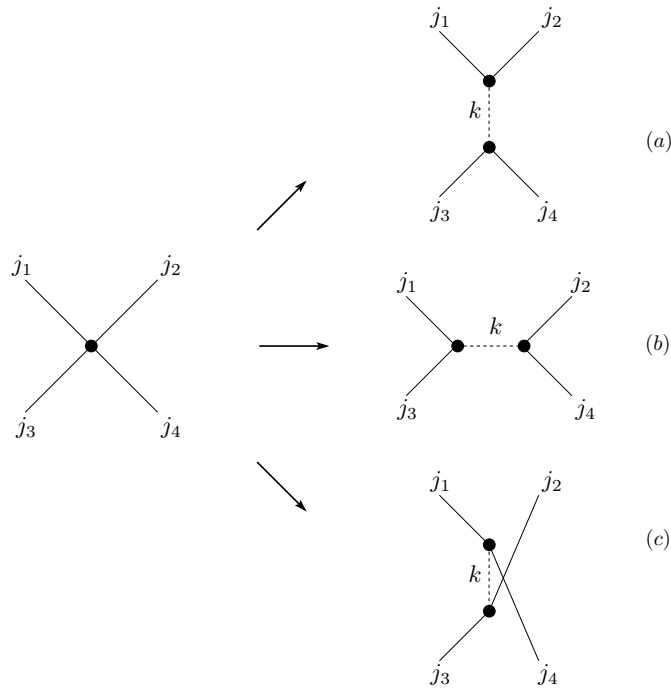

Figure 3. A 4-valent vertex is defined by a particular way of connecting 3-valent vertices. The spin k has to satisfy the triangle inequalities, but is otherwise arbitrary.

represent the 4-valent vertex by splitting it into two 'virtual' 3-valent vertices and adding a 'virtual' edge, carrying angular momentum k (see figure 3). The same wavefunction can be re-expressed by performing this split in a different 'channel' by means of recoupling relations for the Clebsch–Gordan coefficients (see, for instance, [88])

$$C_{\alpha_1\alpha_2\beta}^{j_1 j_2 k} C_{\alpha_3\alpha_4\beta}^{j_3 j_4 k} = \sum_m \sqrt{2k+1}\sqrt{2m+1} \begin{Bmatrix} j_1 & j_2 & k \\ j_3 & j_4 & m \end{Bmatrix} C_{\alpha_1\alpha_2\beta}^{j_1 j_3 m} C_{\alpha_3\alpha_4\beta}^{j_2 j_4 m}, \qquad (4.6)$$

where the object with curly brackets is the Wigner $6j$ symbol. To make the dependence of a general spin network wave functional on the spins j_e associated with the edges and the intertwiners C_v associated with the vertices explicit, one occasionally designates functionals such as (4.3) also by $|\Gamma, \{j\}, \{C\}\rangle$. The space of finite linear combinations of such states is denoted by $\mathcal{S}$ (for 'spin networks'). Although 3-valent spin networks are obviously simplest, we will see later (see section 4.4.2) that higher valence is even generic because 3-valent networks correspond to 'zero volume', and hence are deemed to be not of much interest.

The wave functionals (4.3) are called *cylindrical*, because they probe the connection A only 'on a set of measure zero' (like the δ-function does for ordinary functions). Similar 'cylindrical functionals' were formerly used in constructive quantum field theory in order to rigorously define the functional measures for free and certain interacting models of quantum field theory as limits of finite dimensional integration measures [83]. The space $\mathcal{S}$ spanned by finite linear combinations of such cylindrical functions over all possible graphs is the starting point for the construction of the Hilbert space of spin networks. Obviously, the product of two cylindrical functions supported on the same or different spin networks is again a cylindrical function.

To complete the definition of the space of spin network states, we must introduce a suitable scalar product. In LQG this is not the standard scalar product induced by a Fock space representation (see the following section for more on this); instead, the scalar product of two cylindrical functions $\Psi_{\Gamma, \{j\}, \{C\}}[A]$ and $\Psi'_{\Gamma', \{j'\}, \{C'\}}[A]$ is defined as

$$\langle \Psi_{\Gamma, \{j\}, \{C\}} | \Psi'_{\Gamma', \{j'\}, \{C'\}} \rangle$$

$$= \begin{cases} 0 & \text{if } \Gamma \neq \Gamma' \\ \int \prod_{e_i \in \Gamma} dh_{e_i} \, \bar{\psi}_{\Gamma, \{j\}, \{C\}}(h_{e_1}, \dots) \psi'_{\Gamma', \{j'\}, \{C'\}}(h_{e_1}, \dots) & \text{if } \Gamma = \Gamma', \end{cases}$$

(4.7)

where the integrals $\int dh_e$ are to be performed with the $SU(2)$ Haar measure. The above scalar product may look artificial, but given a few reasonable assumptions, there in fact exists a strong uniqueness theorem; the reader is referred to [89–94] for details.

From (4.7) we see that the inner product vanishes if the graphs Γ and Γ' do not coincide (even if they are 'very close to each other' in any given fiducial background metric). If Γ and Γ' coincide, the product may still vanish, depending on the choice of spins and intertwiners. At any rate, for $\Gamma = \Gamma'$, the inner product is given by integrating the product of $\bar{\psi}$ and ψ' over the holonomies at all edges, using the standard Haar measure. Because Γ was assumed to consist of *finitely* many edges, this is always a *finite-dimensional* integral, with one $SU(2)$ integral for every edge in the graph. It is noteworthy that the above scalar product is *invariant under spatial diffeomorphisms*, even if the states Ψ_1 and Ψ_2 themselves are not, because the statement whether two graphs coincide or not is diffeomorphism invariant. Whereas the original wave functionals $\Psi[A]$ probe the value of the connection and therefore also depend on the position of the graph, this information is 'lost' in the scalar product (4.7) which makes no more reference to the underlying space of connections or the 'shape' of the spin network graph.

One now defines the *kinematical Hilbert space* $\mathcal{H}_{\text{kin}}$ as the completion of the space of gauge invariant spin network states $\mathcal{S}$ w.r.t. to this norm. $\mathcal{H}_{\text{kin}}$ thus consists of all linear superpositions of spin network states $\Psi_n \equiv |\Gamma_n, \{j_n\}, \{C_n\}\rangle$ such that they have finite norm,

$$\Psi = \sum_{n=1}^{\infty} a_n \Psi_n \quad, \qquad \|\Psi\|^2 = \sum_{n=1}^{\infty} |a_n|^2 \|\Psi_n\|^2 < \infty \tag{4.8}$$

where the norm $\|\Psi_n\|^2$ is defined by (4.7). One of the distinctive features of the Hilbert space $\mathcal{H}_{\text{kin}}$, which will play a key role in the further development of the theory, is its *non-separability*. This non-separability can be traced back to the existence of an underlying continuum, the spatial manifold Σ, and accounts for the difference between this scheme and the standard (Fock space) quantization of gauge theories, see the following subsection. Although each Ψ_n in the above sum is associated with a *finite* graph Γ_n the expected number of edges need not be finite, because the sum

$$\langle \#(\text{edges}) \rangle \propto \sum_{n=1}^{\infty} |a_n|^2 L(\Gamma_n) \|\Psi_n\|^2, \tag{4.9}$$

with $L(\Gamma_n)$ the number of edges in Γ_n, can be made to diverge if $L(\Gamma_n)$ increases sufficiently rapidly with n, even if $\sum_n |a_n|^2 \|\Psi_n\|^2 < \infty$. Idem for the expectation value for the number of vertices. Let us, however, caution readers already at this point that $\mathcal{H}_{\text{kin}}$ is *not* the relevant Hilbert space for solving the quantum constraints in LQG. The physical Hilbert space, consisting of those states which satisfy the constraints, *is* expected to be separable; we will have more to say about this in the following sections.

A second unusual feature, at least for a traditional quantum field theorist, is the *ab initio* absence of negative norm states, despite the fact that at this stage the constraints have not even appeared yet. This is in stark contrast to the usual covariant Fock space quantization of gauge theories, where negative norm states are unavoidable, and can only be eliminated (if they can be eliminated at all) by restricting the Hilbert space to the subspace of *physical states*, which by definition are annihilated by the constraints. Curiously, there does exist a reformulation of (free) second quantized Maxwell theory with similar properties [95]. Roughly speaking, this consists in (i) defining the scalar product to be identical with the standard Fock space one for gauge-invariant expectation values (as is well known, in second quantized Maxwell theory this restriction indeed eliminates negative norm states), and (ii) by declaring gauge-variant scalar products to be zero. In this way one preserves Lorentz covariance (because a gauge-variant state is never transformed into a gauge-invariant one by a Lorentz transformation, and vice versa), and eliminates negative norm states, in such a way that *for observable quantities the results are identical with the usual Fock space quantization*. The price one pays is the loss of weak continuity (because the expectation values for gauge-variant expressions always vanish, no matter how close they are to gauge-invariant ones) and the loss of separability of the Hilbert space. These are precisely the features encountered in the LQG quantization procedure above. However, the crucial difference is that the prescription [95] is tailored to reproduce the standard Fock space quantization in the observable sector, whereas LQG does not. Also, it is not known how to extend the prescription of [95] to full QED or Yang–Mills theory, or how to implement the procedure in a perturbative approach (with unphysical intermediate states in Feynman diagrams).

4.2. Comparison with the loop approach in gauge theories

The (Wilson) loop approach has been advocated in the past in the context of gauge theories, as an attempt to non-perturbative quantization. This approach eventually did not lead to the expected success, see [96, 97] for detailed reviews of the subject. Nevertheless LQG practitioners argue that their approach, despite its similarity to the loop approach in gauge theories, should work for gravity. Because this relates to one of the frequently asked questions about the LQG programme, we here briefly contrast some key features of the Wilson loops in gauge theory with the behaviour of spin networks in LQG. Not unexpectedly, the essential difference turns out to lie in the special scalar product (4.7) used in LQG.

The first difference between Wilson loops and spin networks is that, due to the product (4.7), any spin network is orthogonal to any other spin network except itself, including the trivial empty one (i.e. the vacuum). Hence,

$$\langle \Psi_{\Gamma, \{C\}}[A] \rangle = \langle \mathbf{1} | \Psi_{\Gamma, \{C\}}[A] \rangle = 0. \tag{4.10}$$

By contrast, in gauge theory, the expectation value of a single Wilson loop encodes key physical information: it determines the behaviour and properties of quark–anti-quark potentials, and can signal confinement of quarks. More specifically, in a perturbative approach we have

$$\langle W_C \rangle \equiv \left\langle \exp\left[\mathrm{i}e \oint_C A_\mu \mathrm{d}x^\mu \right] \right\rangle = \exp\left[-\frac{\mathrm{i}e}{2} \oint_C \oint_C \mathrm{d}x^\mu \, \mathrm{d}x^\nu \, \Delta_{\mu\nu}(x - y) \right], \tag{4.11}$$

where $\Delta_{\mu\nu}$ is the gluon propagator. In conventional quantum field theory (QCD), this expression is divergent, in contrast to (4.10) because of the self-interaction of the quarks and gluons propagating in the loop. The way in which it diverges generically depends on the shape of the loop. Nevertheless, the UV divergences in (4.11) can be consistently removed, for example by point splitting regularization, and a well-defined answer obtained. All loops can be regularized in this way (while it is only known how to renormalize, *smooth* and *non-self-intersecting* loops [96]). A better, and entirely non-perturbative scheme to calculate the Wilson loops and other quantities of interest is, of course, provided by lattice gauge theory [98–100].

In the case of correlation functions of several Wilson loops, no new types of divergences appear. For two non-intersecting loops C and C' 'self-energy divergences' can be removed by defining the 'connected correlator'

$$\langle W_C W_{C'} \rangle_{\mathrm{conn}} := \langle W_C W_{C'} \rangle - \langle W_C \rangle \langle W_{C'} \rangle = \mathrm{finite}, \tag{4.12}$$

which can be thought of as the analogue of the scalar product of two single loops in LQG. In first approximation we then obtain

$$\langle W_{C_1} | W_{C_2} \rangle_{\mathrm{conn}} = \exp\left[-\frac{\mathrm{i}e}{2} \oint_{C_1} \mathrm{d}x^\mu \oint_{C_2} \mathrm{d}y^\nu \, \Delta_{\mu\nu}(x - y) \right]. \tag{4.13}$$

The main point about this result is that it depends continuously on the shape of the loops C_1 and C_2—unlike the scalar product or correlator of two spin networks (4.7), which is non-zero if and only if the two networks completely overlap. Indeed, different shapes of Wilson loops in gauge theory are inequivalent, such that the relative change can be expressed via the Yang–Mills fields strength and the 'non-Abelian Stokes' theorem'. The value of the Wilson loop is invariant under continuous deformations only for vanishing field strength (as in a topological theory, such as Chern–Simons gauge theory). In LQG on the other hand, the physical states (for which we have to enlarge the space $\mathcal{S}$, as we said) are supposed to be diffeomorphism invariant. Hence there is no physical information whatsoever in the shape of the loop, since two networks of different shape but identical topology can always be related by a suitable diffeomorphism. This equivalence holds independently of the 'value' of the Ashtekar field strength in this state.

When matter is included in the LQG formalism, consistency would seem to require that standard gauge fields must be treated analogously to the Ashtekar connection, i.e. with holonomies associated with edges, and a scalar product similar to (4.7), leading to a non-separable kinematical Hilbert space for the matter sector as well. It is not known how (and whether) this formalism can recover known results such as the ones sketched above, and in particular the shape dependence of the Wilson loops, and we do not know of any attempts in this direction. At the very least, this will require a sophisticated analogue of the 'shadow
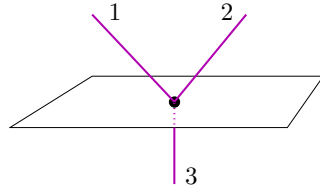
Figure 4. A flux operator intersecting a 3-valent node. The action is given by equation (4.15).

states' introduced in [18]. For kinematical states (which do not satisfy all the constraints) progress along these lines has been made recently in [19–24, 101, 102].

4.3. The action of elementary operators on a spin network

We would now like to implement the canonical variables as operators on spin network wavefunctions in such a way that the canonical commutation relations (4.1) and (4.2) are satisfied. This will provide us with the 'building blocks' that will allow us to define the action of the basic kinematical operators (area and volume operators) and the Hamiltonian constraint operator on any spin network wavefunction. The elementary holonomy and flux operators defined in (3.35) and (3.41) act in a rather simple way on the spin network states. Namely, the holonomy $(h_e[A])_{\alpha\beta}$ simply acts as a multiplication operator, and thus adds the edge e to an existing network Γ. This edge may be disconnected from or 'overlaid' on Γ; usually it will appear in some gauge-invariant combination, whose action on the spin network wavefunction produces another such wavefunction. The action of the flux $\hat{F}_S[\tilde{E}, f]$ on a network state is either zero (if the surface element S does not intersect the graph Γ anywhere), or otherwise cuts in two any edge which pierces S, and inserts a new intertwiner vertex τ_a into the spin network at the point of intersection. Consequently, on such a single edge (which is part of some spin network wavefunction $\Psi_{\Gamma,\{C\}}$) one has

$$\hat{F}_S^a[\tilde{E}](\dots(h_e[A])_{\alpha\beta}\dots) = 8\pi i l_P^2 \hbar\gamma\iota(e,S)\big(\dots\big(h_{e_1}[A]\tau^a h_{e_2}[A]\big)_{\alpha\beta}\dots\big) \tag{4.14}$$

where $e = e_1 \cup e_2$ and we have now omitted the symbol ρ_{j_e}. Dots indicate the remaining part of the spin network wavefunction.

If $\hat{F}_S^a[\tilde{E}]$ cuts the network at a vertex, the result depends on the choice for the intertwiner as well as the position and orientation of the surface S with respect to the edges. Let us illustrate this by considering the 3-valent vertex depicted in figure 4. The flux acts at the end of every edge that emanates from the node. Edges which are located above the surface S contribute with opposite sign from the ones that are located below, and for the example of the 3-valent vertex the net result is

$$\begin{aligned}
\hat{F}_S^a[\tilde{E}]\big(&C_{\alpha_1\alpha_2\alpha_3}^{j_1 j_2 j_3}\big(h_{e_1}[A]\big)_{\alpha_1\beta_1}\big(h_{e_2}[A]\big)_{\alpha_2\beta_2}\big(h_{e_3}[A]\big)_{\alpha_3\beta_3}\big) \\
&= 8\pi l_P^2 \hbar\gamma\Big(\big(\rho_{j_1}(\tau^a)\big)_{\alpha_1\gamma_1} C_{\gamma_1\alpha_2\alpha_3}^{j_1 j_2 j_3} + \big(\rho_{j_2}(\tau^a)\big)_{\alpha_2\gamma_2} C_{\alpha_1\gamma_2\alpha_3}^{j_1 j_2 j_3} \\
&\quad - \big(\rho_{j_3}(\tau^a)\big)_{\alpha_3\gamma_3} C_{\alpha_1\alpha_2\gamma_3}^{j_1 j_2 j_3}\Big)\big(\big(h_{e_1}[A]\big)_{\alpha_1\beta_1}\big(h_{e_2}[A]\big)_{\alpha_2\beta_2}\big(h_{e_3}[A]\big)_{\alpha_3\beta_3}\big) \\
&= -16\pi l_P^2 \hbar\gamma\big(\rho_{j_3}(\tau^a)\big)_{\alpha_3\beta_3} C_{\alpha_1\alpha_2\beta_3}^{j_1 j_2 j_3}\big(\big(h_{e_1}[A]\big)_{\alpha_1\beta_1}\big(h_{e_2}[A]\big)_{\alpha_2\beta_2}\big(h_{e_3}[A]\big)_{\alpha_3\beta_3}\big)
\end{aligned} \tag{4.15}$$

(the uncontracted indices $\beta_1, \beta_2, \beta_3$ connect to the remainder of the spin network wavefunction). In the last step, we have used the invariance property of the Clebsch–Gordan

coefficients,

$$\left(\rho_{j_1}(\tau^a)\right)_{\alpha_1\beta_1} C^{j_1 j_2 j_3}_{\beta_1\alpha_2\alpha_3} + \left(\rho_{j_2}(\tau^a)\right)_{\alpha_2\beta_2} C^{j_1 j_2 j_3}_{\alpha_1\beta_2\alpha_3} + \left(\rho_{j_3}(\tau^a)\right)_{\alpha_3\beta_3} C^{j_1 j_2 j_3}_{\alpha_1\alpha_2\beta_3} = 0. \tag{4.16}$$

Likewise, when the node has higher valence, the action of the flux operator $\hat{F}^a_S[\tilde{E}]$ inserts factors $\pm\rho_{j_e}(\tau^a)$ in the appropriate place in the spin network wavefunction, with a '+' for edges e 'above' S, and a '−' for the edges 'below' S. This action is sometimes symbolically represented in terms of angular momentum operators $\hat{J}^a_{(e,v)}$. Consequently, the flux operator can be re-expressed as a sum of angular momentum operators, namely

$$\hat{F}^a_S[\tilde{E}] = \sum_e \sigma(e, S)\hat{J}^a_{(e,v)}, \tag{4.17}$$

where the sum runs over all edges entering the vertex v, and the signs $\sigma(e, S) = \pm1$ are determined by the position of e in relation to S. When the signs are all '+', the sum vanishes for gauge invariant wave functionals by angular momentum conservation, i.e. Gauss' law.

4.4. The kinematical operators: area and volume operators

Two important *kinematical operators* in quantum gravity are the volume and the area operators. The area operator is relevant in the application of LQG to black hole physics, while the volume operator is essential for the very definition and the implementation of the scalar (Hamiltonian) constraint. It is important to stress that neither of these operators is an observable *à la* Dirac, in the sense that neither commutes with all the canonical constraints (classically or quantum mechanically)[9]. Their supposed physical relevance is based on the general belief that proper Dirac observables corresponding to area and volume should exist, provided suitable matter degrees of freedom are included [103, 104]. The reason is that, for an operational definition of area and volume, one needs 'measuring rod fields' in the same way that the operational definition of 'time' in quantum gravity requires a 'clock field', and that the true observables are appropriate combinations of gravitational and matter fields (which then would commute with the constraints). Although there has been much discussion in the literature about these 'relational' aspects (see [16, 30, 31, 34] and references therein, and [105, 106] for a more recent discussion in the context of LQG), these issues are by no means conclusively settled, and we are not aware of any fully concrete realization of this intuitive idea (see also [107]).

Setting aside such conceptual worries, the main strategy in constructing the area and volume operators is to first write the classical expressions for these quantities using the basic variables (holonomies and electric field), and then promote those into operators. To do this one first has to regularize the classical expressions by a discretization of the integrals [108, 109]. It will turn out that the ambiguity in the way one does this is irrelevant for the area operator, while it does affect the result for the volume operator.

We should point out that the area and volume operators appear to be of no particular relevance in conventional geometrodynamics. There, they would simply be represented by multiplication operators (defining a 'length operator' would also be straightforward, as it is in LQG [110]). As long as these act on smooth wave functionals, there would be no problem, and issues of regularization and renormalization, like the ones we are about to discuss, would not even arise. On the other hand, on more singular 'distributional' functionals, these operators would not be well defined, as they depend nonlinearly on the basic metric variables. Whether it is possible or not to properly define these operators in geometrodynamics, therefore, depends on what the 'good' and the 'bad' wave functionals are. But this question is impossible to

---

[9] As pointed out in [16], such observables might perhaps better be called *'perennials'*, to distinguish them from the kinematical 'observables' which commute only with the spatial diffeomorphism generators.

answer in the absence of a suitable scalar product and a Hilbert space of wave functionals (see e.g. [31] for a discussion of these and other difficulties in geometrodynamics).

4.4.1. The area operator. A two-dimensional surface $S \subset \Sigma$ possesses a unit normal vector field $\vec{N}$ and its classical area is given by the determinant of the induced metric $h_{mn} = g_{mn} - N_m N_n$. To compute the area, it is however simpler to work directly with the area element (3.38) (cf subsection 3.4), so that

$$A_S[g] = \int_S \sqrt{dF^a \cdot dF^a}. \tag{4.18}$$

To convert this into an expression that can act on the spin network wavefunction, we subdivide the surface into infinitesimally small surfaces S_I (with $I = 1, \ldots, N$), such that $S = \cup_I S_I$. Next, we approximate the area by a Riemann sum (which, of course, converges for well-behaved surfaces S), using

$$\int_{S_I} \sqrt{dF^a \cdot dF^a} \approx \sqrt{F_{S_I}^a[\tilde{E}] F_{S_I}^a[\tilde{E}]}. \tag{4.19}$$

Therefore

$$A_S[\tilde{E}_m^a] = \lim_{N \to \infty} \sum_{I=1}^{N} \sqrt{F_{S_I}^a[\tilde{E}] F_{S_I}^a[\tilde{E}]}. \tag{4.20}$$

The Riemann sum on the right-hand side of (4.20) can now, for fixed N, be promoted unambiguously to an operator. Namely, when acting on an arbitrary spin network wavefunction, the area operator will receive a contribution each time an edge of the spin network pierces an *elementary* surface S_I in the sum (4.20). More precisely, one can use equation (4.14) to derive (always in the appropriate representations $\rho_{j_{e_i}}$ and assuming $\iota(e, S) = \pm 1$)

$$\hat{F}_{S_I}^a[\tilde{E}] \hat{F}_{S_I}^a[\tilde{E}] (h_e[A])_{\alpha\beta} = i^2 \left(8\pi l_P^2 \hbar \gamma\right)^2 \left(h_{e_1}[A] \tau_a \tau^a h_{e_2}[A]\right)_{\alpha\beta}$$

$$= \left(8\pi l_P^2 \hbar \gamma\right)^2 (j_e(j_e + 1))(h_e[A])_{\alpha\beta}. \tag{4.21}$$

Note also that if the surface S_I had been pierced by two edges (carrying representations j_1 and j_2) instead of just one, the previous expression would generalize to

$$\hat{F}_{S_I}^a[\tilde{E}] \hat{F}_{S_I}^a[\tilde{E}] \left(h_{e_1}[A]\right)_{\alpha\beta} \left(h_{e_2}[A]\right)_{\gamma\delta}$$

$$= \left(8\pi l_P^2 \hbar \gamma\right)^2 \left(j_{e_1}(j_{e_1} + 1) + j_{e_2}(j_{e_2} + 1)\right) \left(h_{e_1}[A]\right)_{\alpha\beta} \left(h_{e_2}[A]\right)_{\gamma\delta}. \tag{4.22}$$

Hence if one applies the operator (4.20) to wavefunction associated with a fixed graph Γ with $L(\Gamma)$ edges, and refines it in such a way that each elementary surface S_I is pierced by only *one* edge of the network, one obtains

$$\hat{A}_S \Psi = 8\pi l_p^2 \hbar \gamma \sum_{p=1}^{L(\Gamma)} \sqrt{j_p(j_p + 1)} \Psi. \tag{4.23}$$

It is obvious that further refinement of the area operator then does not change the result. It is also clear that the final result, therefore, does not depend on the shape of the elementary area cells. From (4.23), one sees that all spin networks are eigenstates of the area operator (when the S_I intersect only edges, see below). Due to the 'discreteness' of the spin network states, the spectrum of the area operator is also discrete. The quantization scale is set not only by the Planck length l_p^2, but also depends on the parameter γ.[10]

---

[10] We emphasize that this discreteness, which is often invoked as the underlying reason for the absence of divergences in LQG, hinges on the compactness of the group $SU(2)$, and would not hold if this group were replaced by a non-compact or complex group (as would be the case for the original Ashtekar variables with imaginary γ).
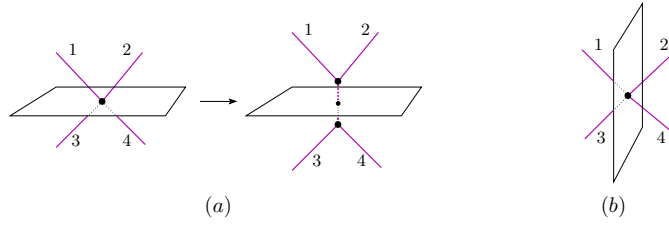
Figure 5. The action of the area operator on a node with intertwiner $C^{j_1 j_2 k}_{\alpha_1 \alpha_2 \beta} C^{j_3 j_4 k}_{\alpha_3 \alpha_4 \beta}$. In the figure on the left, the location of the edges with respect to the surface is such that the invariance of the Clebsch–Gordan coefficients (4.16) can be used to evaluate the action of the area operator. The result can be written in terms of a 'virtual' edge. In the figure on the right, however, the edges are located in a different way with respect to the surface. The invariance property (4.16) does not apply, one has to use the recoupling relation (4.6), and the spin network state is therefore not an eigenstate of the corresponding area operator.

We have so far assumed that the surface elements S_I in the above computation 'meet' only the edges (hence insert a bivalent vertex into the spin network), but not the vertices of the spin network. Making use of the general result (4.17) we can also evaluate the area operator when the surface elements S_I intersect a vertex. In this case, the result depends on the particular intertwiner at that node and on the orientation of the surface element. Let us illustrate this on a 4-valent vertex, with intertwiner $C^{j_1 j_2 k}_{\alpha_1 \alpha_2 \beta} C^{j_3 j_4 k}_{\alpha_3 \alpha_4 \beta}$ (see figure 5). Consider first the situation in which the edges 1 and 2 are both located on the same side of the surface element and the same is true for edges 3 and 4 (figure 5(a)). In this case, one can use the invariance of the Clebsch–Gordan coefficients (4.16) to move around the inserted τ-matrices, and re-write the action of the area operator in terms of a 'virtual' edge. For our example, this yields the result

$$\hat{A}_S |\text{figure } 5(a)\rangle = 8\pi l_p^2 \hbar \gamma \sqrt{2k(k+1)} |\text{figure } 5(a)\rangle. \tag{4.24}$$

Observe the relative factor of $\sqrt{2}$ with respect to (4.23): the result for a virtual edge is not the same as that for a real edge. The 4-valent node is thus an eigenstate of this area operator. Next, consider the situation in which the surface element is oriented as in figure 5(b). In this case, one first has to rewrite the product of Clebsch–Gordan coefficients by means of the recoupling relation (4.6). Independently of the normalization of the $6j$ symbol, the important thing to note is that the area operator will *change* the relative coefficients of the various terms in the sum (4.6), picking up a contribution $\propto \sqrt{m(m+1)}$ for each admissible value of the angular momentum $|j_1 - j_3| \leqslant m \leqslant j_1 + j_3$ in the new channel $1 + 3 \rightarrow 2 + 4$. Hence, the 4-valent vertex is generally *not* an eigenstate of the area operator associated with the surface of figure 5(b). For higher valences, the number of different contributions increases rapidly with the number of accessible channels. Therefore, the area operator in general does *not* act diagonally on an arbitrary spin network state.

We conclude this subsection with three remarks. Although infinite spin networks have been excluded *by fiat*, it is nonetheless instructive to see what would happen if we did include them. In this case it may not be possible to achieve the refinement of one edge per one cell in a *finite* number of steps. To see this, consider the example of a strongly fluctuating network, of 'shape' $z = 0$, $y = \sin(1/x)$ in the fiducial background coordinates. It is clear that if one considers the area of the $y = 0$ surface, then no matter how small the size of cells, the cell around zero will always be pierced by an *infinite* number of edges, and hence the area operator would not be well defined. Because classical space(–time?) presumably emerges in the limit of infinite spin networks, the exclusion of such 'ill-behaved' networks really amounts to a

prescription for the order in which limits are to be taken, when they would otherwise produce an ambiguous answer.

Secondly, the area operator is unbounded, and not every state in $\mathcal{H}_{kin}$ possesses finite area (whereas every element of $\mathcal{S}$ has finite area). This can happen even for compact spatial manifold Σ, and the wavefunction can then be thought of as associated with an 'infinitely crumpled surface' in Σ. To see this more explicitly, consider a sequence of spin network states Ψ_n with edges labelled by $p = 1, \ldots, L(\Gamma_n)$, and spins $j_p^{(n)}$; then we have

$$\hat{A}_S \left(\sum_{n=1}^{\infty} a_n \Psi_n \right) = 8\pi l_p^2 \hbar \gamma \sum_{n=1}^{\infty} \left(\sum_{p=1}^{L(\Gamma_n)} \sqrt{j_p^{(n)}(j_p^{(n)} + 1)} a_n \Psi_n \right). \tag{4.25}$$

The expectation value of $\hat{A}_S$ can be made to diverge by letting the spins $\{j_p^{(n)}\}$ grow sufficiently fast with n, even if $\sum_n |a_n|^2 \|\Psi_n\|^2 < \infty$.

Thirdly, the formula (4.23) plays an important role in the microscopic explanation of the Bekenstein–Hawking entropy of black holes [45, 46], which is now reduced to a counting problem. As an important prerequisite for this calculation, one must first *define* the notion of (isolated) horizon in the quantized theory, i.e. introduce a condition on the quantum states that singles out the 'horizon states'. This has been accomplished in [47]. Fitting the solution of this problem to the known factor relating entropy and area of the horizon then yields a prediction for the value of the parameter γ. While the original value given was $\gamma = \frac{\ln 2}{\pi \sqrt{3}} \approx 0.127\,384\,02\ldots$, a more recent calculation gives $\gamma = 0.237\,532\,95\ldots$ [111, 112] (see also [113, 114]).

4.4.2. The volume operator. The construction of the volume operator is less 'clean cut' than that of the area operator, because it is fraught with ambiguities, which can only be eliminated by making certain choices and by averaging, see [109] for a comprehensive discussion. For a different derivation using a more conventional point-splitting method, which yields the same final result, see [115]. In a first approximation one might try to use the functional differential operator (3.31), but its direct application would lead to (square roots of) singular factors $\delta^{(3)}(\mathbf{x}, \mathbf{x})$. The 'detour' in the construction to be described, in particular the smearing over the surfaces S_I^a in (4.30) below, is necessary mainly in order to eliminate these singular factors. One starts with the classical expression for the volume of a three-dimensional region $\Omega \subset \Sigma$,

$$V(\Omega) = \int_{\Omega} \mathrm{d}^3 x \sqrt{\left| \frac{1}{3!} \epsilon_{abc} \epsilon^{mnp} \tilde{E}_m^a \tilde{E}_n^b \tilde{E}_p^c \right|}. \tag{4.26}$$

Just as with the area operator, one partitions Ω into small cells $\Omega = \cup_I \Omega_I$, so that the integral can be replaced with a Riemann sum. To express the volume operator in terms of canonical quantities, we rewrite the volume element $\mathrm{d}V = \theta^1 \wedge \theta^2 \wedge \theta^3$ in terms of the area elements $\mathrm{d}F^1 = \theta^2 \wedge \theta^3, \ldots$ (see (3.38) as

$$\mathrm{d}V = \sqrt{|\mathrm{d}F^1 \cdot \mathrm{d}F^2 \cdot \mathrm{d}F^3|}. \tag{4.27}$$

In order to express this volume element in terms of the canonical quantities introduced before, we next approximate the area elements $\mathrm{d}F^a$ by the small but finite area operators $F_S^a[\tilde{E}]$, such that the volume is obtained as the limit of a Riemann sum

$$V(\Omega) = \lim_{N \to \infty} \sum_{I=1}^{N} \sqrt{\left| \frac{1}{3!} \epsilon_{abc} F_{S_I^1}^a[\tilde{E}] F_{S_I^2}^b[\tilde{E}] F_{S_I^3}^c[\tilde{E}] \right|}. \tag{4.28}$$

Now, in order to make sense of this expression, we must for each cell Ω_I choose three small non-coincident surfaces S_I^a, which subdivide Ω_I, as shown in figure 6. This should be done
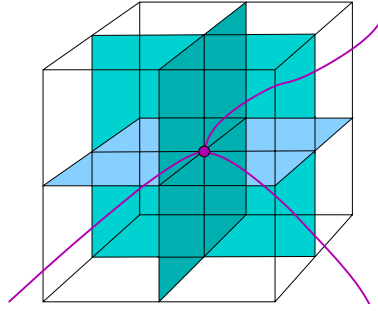
Figure 6. An elementary cell in the Riemann sum (4.29), together with the three surfaces on which the fluxes are evaluated.

in such a way that the rhs of (4.28) reproduces the correct classical value. For instance, one can choose a point inside each cube Ω_I, then connect these points by lines and 'fill in' the faces. In each cell Ω_I we then have three lines labelled by $a = 1, 2, 3$; the surface S_I^a is then the one that is traversed by the ath line. With this choice it can be shown that the result is insensitive to small 'wigglings' of the surfaces, hence independent of the shape of S_I^a, and the above expression converges to the desired result[11].

Having constructed a sequence of refinements with the correct classical limit, we can now attempt to promote (4.28) to a quantum operator by invoking the known action of each factor $\hat{F}_{S_I^a}^a$ on a spin network wavefunction, namely

$$\hat{V}(\Omega) = \lim_{N \to \infty} \sum_{I=1}^{N} \hat{V}(\Omega_I) = \lim_{N \to \infty} \sum_{I=1}^{N} \sqrt{\left|\hat{q}_I^{(N)}\left(S_I^1, S_I^2, S_I^3\right)\right|}, \qquad (4.29)$$

with the operator (= square of the local volume)

$$\hat{q}_I^{(N)}\left(S_I^1, S_I^2, S_I^3\right) := \frac{1}{3!} \epsilon_{abc} \hat{F}_{S_I^1}^a[\tilde{E}] \hat{F}_{S_I^2}^b[\tilde{E}] \hat{F}_{S_I^3}^c[\tilde{E}]. \qquad (4.30)$$

Ignoring the square root for the moment, (4.30) is indeed well defined as an operator at this point. However, unlike the classical volume, the action of $\hat{q}_I^{(N)}$ on the spin network wavefunction depends on the way we partition Ω, and more specifically on the position of the faces S_I^a in relation to the edges and vertices. As a simple example of the inherent ambiguities consider a cube which is traversed by one edge e, but not containing any vertex of the network. If the surfaces S_I^a are positioned in such a way that the edge intersects each one of them with $\iota(e, S_I^a) \neq 0$ at three different points in Ω_I, the operator simply inserts the matrices τ^1, τ^2 and τ^3 at the appropriate places in $h_e[A]$. If the cell is sufficiently small, we can bring these matrices together, using $\tau^1\tau^2\tau^3 = i$, and thus pick up a contribution of order one. It is easy to construct sequences of refinements such that an arbitrarily large number of cells Ω_I give such a contribution, no matter how small the cells are. In order to prevent catastrophic divergent sums from appearing on the rhs of (4.29) one must, therefore, rule out such configurations 'by hand' and stipulate that each cell Ω_I may contain at most two such intersections with any edge, in which case one of the three factors in (4.30) will give zero.

With this proviso, the sum receives contributions only from those cells Ω_I which contain a vertex v. If the vertex v does not coincide with the intersection point of the three faces S_I^a, further refinement will yield more cells, which do not contribute by the above requirement. Hence, with sufficient refinement only those cells will remain and contribute for which the

---

[11] Alternatively, one can partition Σ with other types of cells, with analogous results.
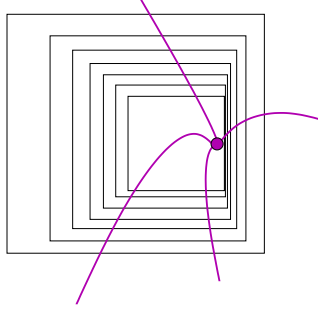
Figure 7. A series of refining steps for one cell in a Riemann sum, such that the indicated node is never at the centre of any of the boxes except in the strict $N \to \infty$ limit. A volume operator based on cells of this kind would always yield a vanishing result for any finite cell size.

vertex v sits precisely at the intersection: $v = S_I^1 \cap S_I^2 \cap S_I^3$. This means that we must also exclude refinements such as in figure 7 for which $S_I^1 \cap S_I^2 \cap S_I^3$ 'misses' v by an amount that gets smaller with N, but never vanishes for $N < \infty$. With these assumptions, the resulting action of the operator $\hat{q}_I$ on the vertex v reduces to

$$\hat{q}_I = \left(8\pi l_P^2 \gamma\right)^3 \frac{1}{48} \sum_{e_1, e_2, e_3} \epsilon(e_1, e_2, e_3)\epsilon_{abc} \hat{J}_a^{(e_1)} \hat{J}_b^{(e_2)} \hat{J}_c^{(e_3)}, \qquad (4.31)$$

where the sum runs over all possible triplets of edges in the vertex v, and the angular momentum operator $\hat{J}_a^{(e,v)}$ (defined in (4.17) by definition inserts a matrix $\rho_{j_e}(\tau^a)$ at the vertex v into the spin network wavefunction. The coefficients $\epsilon(e_1, e_2, e_3)$ are orientation factors which are equal to ± 1. They depend on how the faces S_I^a cut across the vertex, and on which edges are 'above' and 'below' S_I^a. More precisely, in terms of the factors $\sigma(e, S)$ introduced in (4.17) they are given by

$$\epsilon(e_1, e_2, e_3) = \sigma\left(e_1, S_I^1\right)\sigma\left(e_2, S_I^2\right)\sigma\left(e_3, S_I^3\right) \qquad (4.32)$$

and thus completely fixed by the relative positions of the edges and the surfaces. Let us also stress they are independent of any fiducial metric background structure, and purely combinatorial—hence invariant under diffeomorphisms. Equation (4.32) is not yet the final answer, but let us first make the above formulae a little more concrete by working out the action of the above operator on a 4-valent vertex (4.5). Its net effect is the replacement of the intertwiner according to

$$C_{\alpha_1 \ldots \alpha_4}^{j_1 \ldots j_4} \to \text{const.} \sum_{\text{perm.}\{1,2,3,4\}} \epsilon(e_1, e_2, e_3)\epsilon_{abc}$$

$$\times \left(\rho_{j_{e_1}}(\tau^a)\right)_{\alpha_1 \beta_1} \left(\rho_{j_{e_2}}(\tau^b)\right)_{\alpha_2 \beta_2} \left(\rho_{j_{e_3}}(\tau^c)\right)_{\alpha_3 \beta_3} \delta_{\alpha_4 \beta_4} C_{\beta_1 \ldots \beta_4}^{j_{e_1} \ldots j_{e_4}}, \qquad (4.33)$$

where the sum runs over all choices of three edges out of four. From this formula it is obvious that in general the operator $\hat{q}_I$ acts *non-diagonally* on a spin network wavefunction $|\Gamma, \{j\}, \{C\}\rangle$, because it changes the intertwiners $\{C\}$, although it does not affect the spin network itself (i.e. the graph Γ). The operator vanishes when all the edges enter v through the same octant, in which case $\epsilon = +1$ for all choices, and

$$\epsilon_{abc}\left(\hat{J}_1^a \hat{J}_2^b \hat{J}_3^c + \hat{J}_1^a \hat{J}_2^b \hat{J}_4^c + \hat{J}_1^a \hat{J}_3^b \hat{J}_4^c + \hat{J}_2^a \hat{J}_3^b \hat{J}_4^c\right) = 0 \qquad (4.34)$$

by angular momentum conservation, $\hat{J}_1^a + \hat{J}_2^a + \hat{J}_3^a + \hat{J}_4^a = 0$. Note that we can in principle arrange any 4-valent vertex to conform with this choice (see figure 8). Similar results hold true when yet more edges enter the vertex from one octant.
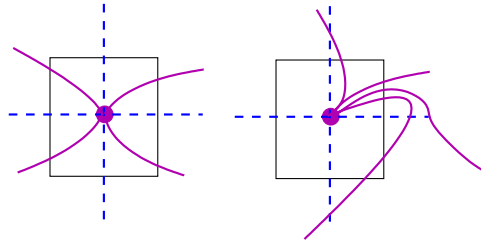
Figure 8. Two different choices for the relative orientation of the edges emanating from a 4-valent node and the cell used in the definition of the volume operator (prior to averaging over the positions of the surfaces S_I). It is always possible to choose the division of the cell such that all four edges emanate from the same octant. We have here drawn the 'active' version in which the edges rather than the surfaces are deformed. Note that these two configurations are not homotopically related.
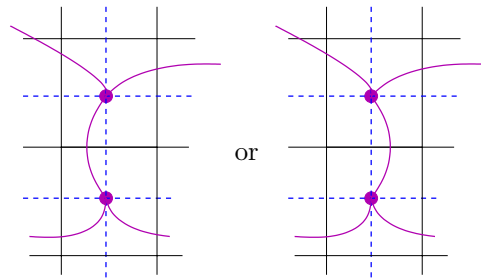

Figure 9. Sign ambiguities in the volume operator, present because the relative location of the three surfaces S_I and the edges of the spin network can be chosen in different ways (again prior to averaging, as in figure 8).

Let us now return to the coefficients $\epsilon(e_1, e_2, e_3)$ in (4.32). Though manifestly background independent they do depend on how we choose to position the faces and edges: for instance, there is no *a priori* argument that would force a given edge linking two vertices in adjacent cells to appear 'to the left' or 'to the right' of the face connecting the two vertices, as illustrated in figure 9, and in fact almost any combination of factors ± 1 is possible. To eliminate this remaining ambiguity in the definition of the volume operator, one now averages over the various possibilities[12]. More specifically, this is done by integrating over the positions of the surfaces with some measure $d\mu(\theta_1, \theta_2, \theta_3)$, where the θ_i are suitable angular coordinates, namely

$$\hat{\epsilon}(e_1, e_2, e_3) := \int d\mu(\theta_1, \theta_2, \theta_3)\sigma\left(e_1, S_I^1(\theta_1)\right)\sigma\left(e_2, S_I^2(\theta_2)\right)\sigma\left(e_3, S_I^3(\theta_3)\right). \tag{4.35}$$

This leads to a well-defined combinatorial answer depending on the relative orientation of the three edges (e_1, e_2, e_3) ($=$ the sign of the determinant of the dreibein spanned by the tangents to the three edges at the vertex), but only modulo an overall factor which encapsulates the freedom in the choice of the measure $d\mu$. In principle such a factor can arise for each vertex, but in order to preserve diffeomorphism invariance of the volume these factors should be

___

[12] We might note that this averaging is 'undemocratic' insofar as it is applied only to cells containing a vertex. If it were extended to other cells, some of the unwanted features which we just disposed of might re-appear. Vertex-less cells which satisfy the condition that edges intersect at most two surfaces do not necessarily satisfy this condition for all surface orientations which appear in the averaging procedure.

chosen to be *the same* for all vertices. This factor is the one ambiguity which remains in the definition of the volume operator, and this ambiguity cannot be eliminated [109].

Finally, given an assignment of $\hat{\epsilon}(e_1, e_2, e_3)$ for each vertex, how do we actually compute the spectrum of the volume operator in (4.29)? First of all, it is easy to see that this operator always vanishes on 3-valent vertices which satisfy the Gauss constraint. This is a direct consequence of angular momentum conservation, or equivalently

$$\epsilon_{abc}\big(\rho_{j_{e_1}}(\tau^a)\big)_{\alpha_1\beta_1}\big(\rho_{j_{e_2}}(\tau^b)\big)_{\alpha_2\beta_2}\big(\rho_{j_{e_3}}(\tau^c)\big)_{\alpha_3\beta_3}C^{j_{e_1}j_{e_2}j_{e_3}}_{\beta_1\beta_2\beta_3} = 0. \tag{4.36}$$

In other words, the volume operator vanishes on gauge invariant 2- and 3-valent vertices (but *not* on gauge variant ones), and picks up contributions only from spin networks with vertices of valence $\geqslant 4$. Given any spin network wavefunction, the first step is to work out the action of $\hat{q}_I$ according to the above rules. In a second step this action must be diagonalized for each I on the spin network wavefunctions. This can be done because the operators $\hat{q}_I$ commute amongst each other, i.e. $[\hat{q}_I, \hat{q}_{I'}] = 0$ for $I \neq I'$. Only now can we define the absolute value $|\hat{q}_I|$ and take the square root $\sqrt{|\hat{q}_I|}$ using the spectral decomposition of this operator (note that $\hat{q}_I$ is not necessarily a positive operator). Unfortunately, these steps are all but straightforward to carry out in practice [116]. The analysis can be simplified by applying the operator identity [115, 117]

$$\epsilon_{abc}J_a^1 J_b^2 J_c^3 = \frac{i}{4}[(\mathbf{J}^1 + \mathbf{J}^2)^2, (\mathbf{J}^2 + \mathbf{J}^3)^2] \tag{4.37}$$

to any three edges entering the vertex, enabling a numerical analysis of simple configurations with low-valence nodes. This procedure has recently been used to show that the operator corresponding to the inverse volume is not bounded from above when acting on zero-volume spin network states [41] (in marked contrast to the symmetry reduced models of [38], where the avoidance of an initial singularity hinges on the fact that the operator corresponding to the 'inverse volume' is bounded from above[13]). Nevertheless, the resulting formulae remain very complicated. The lack of a more explicit representation for the operators $\sqrt{\hat{q}_I}$ is also a serious impediment towards a better understanding of the Hamiltonian constraint.

4.5. Some comments on regularization dependence

At this point some comments regarding the regularization procedures for kinematical operators which we have outlined above, are in order. Similar comments will apply to the regularization of the Hamiltonian constraint operator. First of all it should be evident from the foregoing discussion that the notions of 'finiteness' and 'regulator independence' as currently used in LQG are not the same as in the conventional quantum field theory. There one insists that renormalized quantities should come out the same (modulo a finite number of normalizations) after removal of the regulator, *no matter how the latter is chosen*. By contrast, LQG does not claim that different prescriptions for constructing an operator will always lead to the same physical result. This is well illustrated by the volume operator, whose definition and whose very existence hinge on various choices that have been made along the way.

Let us nevertheless briefly summarize what is meant here by 'regularization'. In order to regulate a given classical phase space quantity defined by an integral (such as the volume), one first approximates that integral by a Riemann sum (which is of course assumed in the classical theory to converge to the desired quantity in the limit of infinite refinement). One then replaces in this finite Riemann sum all terms by operators, usually in such a way that

---

[13] The operator $\hat{X}$ corresponding to the inverse volume is *not* the same as the inverse of the volume operator. The latter does not exist on all of $\mathcal{H}_{\text{kin}}$ because of the occurrence of zero eigenvalues in its (discrete) spectrum. Instead, the operator $\hat{X}$ obeys $\hat{X} \circ \hat{V} = \mathbf{1} + \mathcal{O}(\hbar)$, and can therefore in principle stay bounded 'near the singularity'.

the various area or volume elements are absorbed into the definition of these operators (as exemplified by the 'small' operators $\hat{F}^a_{S^a_I}[\tilde{E}]$ entering the definition of the area and volume operators). Acting with a sequence of these discretized operators on an arbitrary, but fixed spin network state, the resulting sequence of refinements (labelled by N, say) eventually becomes stationary (i.e. constant for N sufficiently large), because the state is kept fixed as the limit $N \to \infty$ is taken. This follows straightforwardly for the area operator, but requires a choice of 'preferred' regularizations for the volume operator. With these choices, the 'correct' result is always obtained already after a finite number of refinements. A good analogy would be the Riemann integral of a step function, for which a suitably refined Riemann sum attains the exact value already after finitely many steps, after which the result no longer depends on the 'regulator' (i.e. the size of the intervals defining the Riemann sum). In this sense, the results of LQG are indeed 'regulator independent'.

From what has just been said, it is evident that infinities can never appear in the LQG regularization procedure, and in this sense the resulting theory is 'finite', at least as far as the kinematical operators are concerned. LQG, nevertheless, requires the regularization of the area and volume operators in order even to be able to define the quantum counterparts of the classical constraints via Thiemann's trick. The outlined regularization is, therefore, not introduced to remove divergences. Standard short distance, QFT divergences 'disappear' in the LQG approach by the very construction of the theory: all states are discrete, and at any step of the calculation one deals only with a finite number of objects. The price one pays are ambiguities of the type encountered above, some of which can only be eliminated by making *ad hoc* choices, such as the rule prescribing the way in which an edge has to traverse a cell. We shall encounter more such ambiguities when we attempt to define the Hamiltonian constraint operator.

Of course, these considerations do not necessarily imply any inconsistency. Namely, one can always adopt the above prescriptions for the kinematic operators as *definitions*, which can be plausibly related to the corresponding classical objects, and in this sense the definitions of the above operators in LQG are perfectly acceptable. What we would like to emphasize, however, is that with this construction the work is hardly begun. The far more difficult task is now to show that these definitions, when put together and inserted in the constraints, yield a sensible quantum theory, and a consistent quantum constraint algebra, in particular. The situation here is reminiscent of the one in the early days of constructive quantum field theory (see, e.g., [118] and references therein). The problem then was not so much finding non-perturbative regulators and establishing the existence of selected limits when the regulator is removed. Rather, the main problem was to establish that the resulting theory has all the desired physical properties, to wit, full Poincaré invariance, locality and causality. As is well known, no interacting quantum field theory in four dimensions satisfying all these requirements is known to this day[14]. And for those low dimensional models which could actually be shown to exist and possess all the requisite properties, the tightly knit construction left no room for ambiguities other than the normalization of a finite number of physical parameters (masses and coupling constants).

4.6. Coupling to matter fields

Any realistic theory of quantum gravity should at least allow for a coupling to the matter fields of the standard model of elementary particles. In the framework of geometrodynamics it is fairly easy to extend the relevant formulae to incorporate matter fields (but, of course, the

---

[14] Even the best candidate model for realizing the quantum field theorist's dream of a UV finite theory, $N = 4$ supersymmetric Yang–Mills theory, has not been rigorously shown to exist beyond perturbation theory.

difficulties with quantization remain the same as before). Likewise, it is possible to include gauge fields and fermions in the LQG formalism, essentially without any restriction on the gauge group or the number and structure of fermion families—as we already pointed out LQG does not appear to put any restrictions on the matter couplings (such as renormalizability). Armed with the translation rules of section 3.2, the transition from the 'old' metric or dreibein variables to the loop variables is in principle straightforward, though usually cumbersome (and, we may add, of considerably diminished esthetical appeal in comparison with Ashtekar's original reformulation of *pure* gravity). Because there is an ample literature on this subject (see, e.g., [119–124]), and because the introduction of matter couplings does not alleviate the problems with the Hamiltonian constraint in LQG, we will be very brief and only mention two selected aspects here.

Consider, for example, the coupling of fermion fields in the LQG setting [119, 125]. The starting point is the classical continuum action of a Dirac fermion coupled to gravity

$$S = \int d^4x \, E \, \bar{\chi} \gamma^A E_A{}^\mu \mathcal{D}_\mu \chi \equiv \int d^4x \, E \, \bar{\chi} \gamma^A E_A{}^\mu (\partial_\mu + \omega_{\mu BC} \gamma^{BC}) \chi, \quad (4.38)$$

where we have omitted a possible mass term for simplicity, as its inclusion would pose no new conceptual problems. There are two immediate questions one faces when trying to construct quantum Hamiltonian for this system. One is the choice of the conjugate pair of variables for fermions; finding the proper fermionic variables may also require redefinitions of the other fields so as to 'diagonalize' the canonical (Poisson or Dirac) brackets. For instance, for the above model it turns out [120] that a good choice for the conjugate pair of fermionic variables is the pair of *half-densitized* spinors

$$\tilde{\chi} \equiv \sqrt[4]{\det \tilde{E}} \, \psi, \qquad \pi = i \tilde{\chi}^\dagger \gamma^0. \quad (4.39)$$

If instead of densitized spinors we used bare spinors (without any factors of the determinant of the dreibein), one would have to modify the gravitational connection [120] in order to obtain diagonal canonical brackets, but at the expense of making the connection complex again

$$\tilde{A}_m^a = -\frac{1}{2} \epsilon^{abc} \omega_{mbc} + \gamma K_m^a + \frac{i}{4\sqrt{\tilde{E}}} e_m^a \bar{\chi} \chi. \quad (4.40)$$

This leads to a new reality problem, similar to that already encountered in [76], and not different from the one which eventually led LQG to abandon the 'old' Ashtekar variables with $\gamma = \pm i$. The second question is how to express all variables in terms of the LQG variables $h_e[A]$ and $F_S^a[\tilde{E}]$. The answer to this question can be obtained by employing the identities (3.26) and (3.27). A similar procedure will be spelled out in the next section in order to construct the Hamiltonian constraint, so we will refrain here from giving more details.

When implementing the matter Hamiltonians in terms of spin networks, the description of (Abelian or non-Abelian) gauge fields and matter fields is somewhat similar to the description of standard lattice gauge theories: matter fields (scalars or fermions) are attached to the vertices of the spin network, while gauge fields are associated with the edges (links), see e.g. [121–123] for a treatment of electromagnetism and [124] for scalar field theories in the LQG formalism. There are, however, two main differences between lattice gauge theory and LQG. In the former, the space or spacetime lattice is given, and hence there are only three (or four) possible directions for the edges. The coupling between two fermions χ_1 and χ_2 on adjacent vertices 1 and 2 connected by an edge $e \equiv (12)$, is uniquely given by $\bar{\chi}_1 \gamma_{12} h_{12}[A] \chi_2$, where γ_{12} is the γ-matrix 'along' the edge (12). In LQG (and in general relativity), this simple formula fails because the direction of the link is arbitrary, and a vielbein factor must be inserted to 'align' the γ-matrix with the holonomy. When rewriting the dreibein by means of the formulae of section 3.2, this usually leads to rather awkward expressions in terms of the spin network

variables $h_e[A]$ and $F_S^a[\tilde{E}]$. This problem is further compounded by a second difficulty which goes back to the issue of what the correct value is for the Barbero–Immirzi parameter. As we mentioned already, for $\gamma \neq \pm i$, the connection A is not the pullback of the (chiral) spin connection. Therefore, the holonomy $h_{12}[A]$ does not effect the correct parallel transport between χ_1 and χ_2, and must accordingly be amended by further (and again rather awkward) modifications.

We should stress that, *even* when one has obtained an expression for the matter action in terms of spin network operators, one does not know whether the resulting quantum theory has any relation to the usual Fock-space quantized theory. Showing that such a relation exists requires the field-theory analogues of the 'shadow states' for point particles [18]. These states are complicated linear combinations of elementary spin network states, approximating Fock space coherent states. In flat space some results have been obtained in the case of Maxwell [121–123] and scalar fields [124]. Recent attempts to construct coherent states in the presence of gravity include [19–24, 101, 102].

A completely different question concerns the role of the structure group $SU(2)$, and the possibility of generalizing Ashtekar's variables to higher dimensions, to achieve a Kaluza–Klein-type unification of matter and gravity. The closure of the classical constraint algebra in the Ashtekar reformulation of gravity requires use of the identity $\epsilon_{abe}\epsilon^{cde} = 2\delta_{ab}^{cd}$, which is only valid for the structure constants of the group $SU(2)$ and its non-compact form $SO(1, 2)$. These properties no longer hold for $SU(N)$ when $N > 2$. For this reason, the LQG reformulation of gravity works only in three and four spacetime dimensions. Nevertheless, one can try to generalize the formalism to $SU(N)$ [126] and thereby arrive at a novel type of coupling between gravity and Yang–Mills fields [127, 128]. A very different attempt to generalize the formalism to include matter couplings is based on the observation that an inverse densitized vielbein also appears naturally in certain reformulations of $D = 11$ supergravity [129]. It is not known whether any of these *ansätze* can be implemented in a spin network formulation (although replacing $SU(2)$ holonomies by $SU(N)$ holonomies seems straightforward enough).

5. Quantum constraints

As we have seen in section 3.1, the Hamiltonian framework leads to three classical constraints: the Gauss, diffeomorphism and Hamiltonian constraints. To implement these constraints at the quantum level, one must first properly *define* them, i.e. express them in terms of the elementary variables, the holonomies and fluxes, and then investigate their properties. We will first discuss the kinematical constraints, which are 'easy' (relatively speaking), whence we can be brief. We will then turn to the truly difficult part of the story—the Hamiltonian constraint.

5.1. Kinematical constraints

The Gauss constraint is already expressed in canonical variables $h_e[A]$ and $F_S[\tilde{E}, f]$, and one can solve it exactly. This is simply because fulfilment of the Gauss constraint is equivalent to the gauge invariance of the spin network wavefunction: invariance is simply achieved by contracting the $SU(2)$ representation indices entering a given vertex in an $SU(2)$ invariant manner. Equivalently, one says that *angular momentum is conserved at each vertex*, namely

$$\sum_e \hat{J}_{(e,v)}^a = 0 \tag{5.1}$$

for each $v \in \Gamma$. Recall (see the discussion around (4.17)) that the operator $\hat{J}_a^{(e,v)}$ is *defined* to act by inserting a matrix $\rho_{j_e}(\tau^a)$ at the vertex v in the representation belonging to the

(oriented) edge e, with appropriately contracted $SU(2)$ representation indices, and appropriate orientation. For the following discussion, we will tacitly assume that the wavefunction satisfies the Gauss constraint, whenever appropriate.

The diffeomorphism constraint is more difficult to impose—unlike in geometrodynamics, one cannot immediately write down formal expressions which are manifestly diffeomorphism invariant, because the spin network functions are not supported on all of Σ, but only on one-dimensional links. Nevertheless, it would seem most natural at first sight to try to solve the diffeomorphism constraint on the kinematical Hilbert space $\mathcal{H}_{\text{kin}}$. However, this is impossible for two reasons. First of all, the diffeomorphism generator does not even exist as an operator: there is no way to differentiate finite diffeomorphisms so as to obtain 'infinitesimal' ones due to the lack of weak continuity (and *a fortiori*, differentiability). Secondly, with the exception of the trivial state $\Psi = \mathbf{1}$ (the empty spin network), there is just *no* diffeomorphism invariant state in $\mathcal{H}_{\text{kin}}$. This is intuitively clear because every spin network state changes under a diffeomorphism which moves the associated graph around[15]. For these reasons, LQG invokes a variant of the so-called group averaging method, but in a way that is different from the standard group averaging procedure. Namely, application of the latter would require knowing an integration measure $\mathrm{d}\mu(\phi)$ on the infinite-dimensional group $\text{Diff}(\Sigma)$. The formally diffeomorphism-invariant state would then be

$$\Psi_{\Gamma,\psi}^{\text{inv}}[A] \stackrel{?}{=} \int_{\text{Diff}(\Sigma)} \mathrm{d}\mu(\phi)\Psi_{\Gamma,\psi}[A \circ \phi], \tag{5.2}$$

and if the measure existed, this state would legitimately belong to $\mathcal{H}_{\text{kin}}$. For the infinite-dimensional group $\text{Diff}(\Sigma)$ no such measure is known (the 'counting measure' (5.6) introduced below does not 'live' on $\mathcal{H}_{\text{kin}}$, but leads out of the kinematical Hilbert space), hence the above expression is devoid of meaning—confirming the conclusion above that $\mathcal{H}_{\text{kin}}$ does not admit non-trivial solutions. LQG circumvents this difficulty by implementing 'group averaging' in a way that is more adapted to the non-separable Hilbert space and the properties of the scalar product (4.7). The key observation here is that the constraint can be formally solved on a much larger space, namely *the algebraic dual $\mathcal{D}^*$ of some dense subspace $\mathcal{D}$ of $\mathcal{H}_{\text{kin}}$*, such that

$$\mathcal{D} \subset \mathcal{H}_{\text{kin}} \subset \mathcal{D}^*. \tag{5.3}$$

Typically, the dual space $\mathcal{D}^*$ is thus a space of distributions. This is not as outlandish as it may seem at first, because for constrained systems, the solutions of the constraints are generically distributions, see e.g. [130][16]. Given any (densely defined and closable) operator $\mathcal{O}$ and its adjoint $\mathcal{O}^\dagger$ with domain $\mathcal{D} \subset \mathcal{H}_{\text{kin}}$, the action of the *dual operator $\mathcal{O}^*$* on the algebraic dual $\mathcal{D}^*$ of $\mathcal{D}$ is defined for any particular element $\mathcal{X} \in \mathcal{D}^*$ by requiring

$$\langle \mathcal{O}^* \mathcal{X} \mid \Psi \rangle = \langle \mathcal{X} \mid \mathcal{O}^\dagger \Psi \rangle \tag{5.4}$$

to hold for all $\Psi \in \mathcal{D}$. The space $\mathcal{D}^*$ is very large, and in practice one restricts attention to a physically motivated vector subspace thereof, the so-called habitat $\mathcal{V}^* \subset \mathcal{D}^*$, whose specification is closely related to the definition of the Hamiltonian constraint. Observe also that the above definition does not require the operator $\mathcal{O}$ to be Hermitian.

The obvious ('canonical') choice for $\mathcal{D}$ is

$$\mathcal{D} = \mathcal{S} \tag{5.5}$$

---

[15] This problem disappears altogether in a *topological theory*, such as $(2 + 1)$-dimensional gravity, where the connection is flat and diffeomorphism invariance is manifest from the outset.

[16] A simple example is the solution of the Hamiltonian constraint for a relativistic point particle—the Klein–Gordon equation—which is $\tilde{\phi}(p) = \delta(p^2 - m^2)$. As is well known, a scalar product and a Hilbert space can be defined on this constraint hypersurface, but the scalar product is not positive definite. These features can also be made completely explicit in more complicated exactly solvable theories, such as pure gravity [131, 132] and pure supergravity [133] in three dimensions.

that is, the space of finite linear combinations of spin network states. However, there may exist other viable choices for $\mathcal{D}$. The dual space $\mathcal{S}^*$ is much larger than $\mathcal{H}_{\text{kin}}$, and actually too large for our purposes. What is important here is that $\mathcal{S}^*$ contains a certain subspace of diffeomorphism-invariant states, and its completion w.r.t. to the norm (5.8) below. This subspace is contained in any of the proposed 'habitats' $\mathcal{V}^*$, and is obtained by averaging spin network states. More specifically, for any $\Psi_\Gamma \in \mathcal{S}$ associated with a particular graph Γ we define the 'group average' by the formal sum

$$\eta(\Psi_\Gamma) \equiv \overline{\Psi}_\Gamma := \sum_{\phi \in \text{Diff}(\Sigma|\Gamma)} \phi^* \circ \Psi_\Gamma \qquad (5.6)$$

where $(\phi^* \circ \Psi)[A] := \Psi[A \circ \phi]$, and where $\text{Diff}(\Sigma|\Gamma)$ consists of the subgroup of $\text{Diff}(\Sigma)$ obtained by dividing out the diffeomorphisms leaving invariant the graph Γ on which Ψ_Γ lives. While $\Psi_\Gamma \in \mathcal{S}$ by assumption, averaging the state in this way throws it out of $\mathcal{S}$ and also $\mathcal{H}_{\text{kin}}$. At first sight, this seems to make matters worse: we have replaced the ill-defined integral over $\text{Diff}(\Sigma)$ by a highly formal continuous sum. However, the day is now saved by two facts: (i) unlike the integral, the sum no longer requires unavailable detailed knowledge of the properties of $\text{Diff}(\Sigma)$, and (ii) when applying an averaged spin network state $\eta(\Psi_{\Gamma'})$ (considered as a bra-vector) to any other spin network state $\Psi_\Gamma[A]$, the sum

$$\langle \eta(\Psi_{\Gamma'}) \mid \Psi_\Gamma \rangle = \sum_{\phi \in \text{Diff}(\Sigma|\Gamma')} \langle \phi^* \circ \Psi_{\Gamma'} \mid \Psi_\Gamma \rangle \qquad (5.7)$$

consists at most of a *finite number* of terms, and it is this fact which ensures that $\eta(\Psi_\Gamma)$ is indeed well defined as an element of the dual space $\mathcal{S}^*$. Namely, there is no contribution at all if Γ and Γ' are not diffeomorphic. For Γ and Γ' diffeomorphic, only that term contributes for which Γ and $\phi^* \circ \Gamma'$ coincide. In the above equation, the bracket on the rhs is just the scalar product (4.7), while the bracket on the lhs denotes the dual pairing between $\mathcal{S}$ and $\mathcal{S}^*$.

We can now promote the space of averaged diffeomorphism invariant spin network states to a pre-Hilbert space by 'dividing out the volume of the gauge group' ($= \text{Diff}(\Sigma)$). More precisely, we define the scalar product between two such states as

$$\langle\!\langle \eta(\Psi_{\Gamma',\psi'}) \mid \eta(\Psi_{\Gamma,\psi}) \rangle\!\rangle := \langle \eta(\Psi_{\Gamma',\psi}) \mid \Psi_{\Gamma,\psi} \rangle. \qquad (5.8)$$

Here the bracket on the rhs is the dual pairing between $\mathcal{S}$ and $\mathcal{S}^*$, and the bracket $\langle\!\langle . | . \rangle\!\rangle$ on the lhs denotes the new scalar product. The new Hilbert space $\mathcal{H}_{\text{diff}}$ of diffeomorphism averaged states is then obtained by completion w.r.t. this norm, and can be thought of to consist of the averaged versions of the states (4.8) (but let us emphasize again that it is *not* a subspace of $\mathcal{H}_{\text{kin}}$). Observe that the Hilbert space $\mathcal{H}_{\text{diff}}$ is still a 'small' subspace of $\mathcal{S}^*$ because the vast majority of elements in $\mathcal{S}^*$ does not correspond to diffeomorphism averaged spin network states, and cannot be obtained as limits of such states w.r.t. to the above norm. Moreover, $\mathcal{H}_{\text{diff}}$ is 'smaller' than the original kinematical Hilbert space $\mathcal{H}_{\text{kin}}$, because distinct elements of the latter which are related by a diffeomorphism get mapped to the *same* element of $\mathcal{S}^*$; equivalently, the averaging map has a very large kernel. There is no way to extend the above norm from $\mathcal{H}_{\text{diff}}$ to all of $\mathcal{S}^*$, in accordance with the fact that distribution spaces cannot be made into Hilbert spaces, but at best into topological vector spaces; however, there appears to be no consensus as to which of the many possible topologies is the right one for $\mathcal{S}^*$. While $\mathcal{H}_{\text{kin}}$ is non-separable, it is expected that $\mathcal{H}_{\text{diff}}$ *is* separable [134, 135] provided one averages over *homeomorphisms* rather than diffeomorphisms (if one divides out only by analytic or C^∞ diffeomorphisms, $\mathcal{H}_{\text{diff}}$ remains non-separable, see also [12] for a recent discussion of this point).

Which, then, is the arena, or 'habitat', in which quantum gravity takes place according to LQG, and where one must ultimately consider the action of the Hamiltonian constraint? The foregoing discussion shows that there is no easy answer to this question, and apparently also no complete consensus in the community as to which space is 'right'. The Hilbert space $\mathcal{H}_{\text{diff}}$ of diffeomorphism averaged states certainly appears to be a good starting point for implementing and solving the Hamiltonian constraint, but it may not be good enough. The reason is that the action of the Hamiltonian constraint in general does not preserve $\mathcal{H}_{\text{diff}}$; in other words, it maps diffeomorphism invariant states into ones that are not invariant. Consequently, when checking the commutator of two Hamiltonian constraints, one must again pass to a larger space, which is preserved by the action of the Hamiltonian, but which is no longer a Hilbert space. An explicit example are the 'vertex-smooth' states introduced in [29], obeying $\langle \mathcal{X} | \Psi_1 \rangle = \langle \mathcal{X} | \Psi_2 \rangle$ for all Ψ_1, Ψ_2 which are related by a diffeomorphism that leaves the vertices of the spin network fixed. Alternatively, it has been proposed to implement the Hamiltonian constraint on the kinematical Hilbert space $\mathcal{H}_{\text{kin}}$ via a weak* limit [27], see also the following section. At any rate, the challenge here—as with any other constrained quantum mechanical system—is to find a subspace that is annihilated by *all* the constraints, and to define a *physical inner product* which yields the 'final' *physical Hilbert space*.

5.2. Hamiltonian constraint

As we explained, the diffeomorphism constraint is imposed not in terms of an operator constraint, but through the formal group averaging method. By contrast, the Hamiltonian constraint is imposed as an *operator* constraint. Hence, as a zeroth step, one has to construct this Hamiltonian operator. As we explain in more detail below, in this derivation one is forced, at a certain stage, to make choices concerning the action of the Hamiltonian operator on a generic spin network state. Thus the resulting *definition* is indeed motivated by the classical expression for the Hamiltonian (expressed in loop variables), but it is plagued by a very large number of ambiguities[17]. To test the proposed prescription and compare it with other possibilities, one would need to find states which are annihilated by the Hamiltonian, and check that in the semi-classical limit these states produce sensible results. This is, however, currently out of reach: due to the great complexity of the proposed Hamiltonian, not a single physically interpretable eigenstate is known. While this should not come as a surprise (after all, we have no reason to expect to be able to solve the theory exactly), it is remarkable that even the far more modest goal of working out in complete detail the action of the (regulated) Hamiltonian on a general spin network wavefunction appears difficult to attain.

Even if we can work out the action of the Hamiltonian on any given element of $\mathcal{S}$, we are not done. As explained in the previous subsection, in order to ensure satisfaction of the diffeomorphism constraint, we must transfer the action of the Hamiltonian to the dual space $\mathcal{S}^*$, or at least to some physically motivated subspace $\mathcal{V}^*$ thereof. Below, we will be mostly dealing with a *regulated Hamiltonian* $\hat{H}[N, \epsilon]$ (where N is the lapse function). Here, the regularization parameter $\epsilon > 0$ enters via a plaquette $P(\epsilon)$, which is attached to a vertex and which must be shrunk to zero at the end of the calculation. However, for a given Ψ, the limit $\epsilon \to 0$ of $\hat{H}[N, \epsilon]\Psi$ does *not* exist on $\mathcal{S}$, because wave functionals supported on the same network, but with an extra loop $\partial P(\epsilon)$ attached to one of the vertices, are orthogonal to one another for different values of ϵ by (4.7). For this reason, one must resort to a weaker notion

---

[17] We should caution readers that it is a matter of debate as to what constitutes an 'ambiguity', or whether a particular choice narrowing down the number of possibilities should be viewed as 'natural' or even mandatory in order to obtain sensible expressions (one example being the 'translation invariant' choice of measure factor in the definition of the volume operator). At any rate, LQG practitioners may well disagree with our terminology here.

of limit by transferring the action of the Hamiltonian to the dual space. More specifically, we define the limit $\epsilon \to 0$ by demanding

$$\langle \hat{H}^*[N]\mathcal{X} \mid \Psi \rangle = \lim_{\epsilon \to 0} \langle \mathcal{X} \mid \hat{H}[N, \epsilon]\Psi \rangle \tag{5.9}$$

to hold for all $\Psi \in \mathcal{S}$ and $\mathcal{X} \in \mathcal{V}^*$. Here $\mathcal{V}^* \subset \mathcal{S}^*$ is the 'habitat' already mentioned in the previous subsection, a vector subspace of $\mathcal{S}^*$. The choice of $\mathcal{V}^*$ should be physically motivated. Natural choices are $\mathcal{V}^* = \mathcal{H}_{\text{diff}}$ or the space of vertex smooth states [29], but there may be others. An evident requirement here is, of course, that the choice of $\mathcal{V}^* \subset \mathcal{S}^*$ should be such that the limit (5.9) exists. Although it is not entirely clear (to us) what precisely the conditions on $\mathcal{V}^*$ should be for this limit to exist, and to what extent they would determine $\hat{H}^*[N]\mathcal{X} \in \mathcal{V}^*$ uniquely, it seems obvious that the limit (5.9) will *not* exist if $\mathcal{V}^*$ is 'much larger' than $\mathcal{H}_{\text{diff}}$. There is an alternative point of view, which does not require 'habitats': according to [27, 74], the choice $\mathcal{V}^* = \mathcal{H}_{\text{diff}}$ in fact *is* singled out, and formula (5.9) should not be viewed as defining a dual Hamiltonian on $\mathcal{V}^*$, but rather as defining the limiting Hamiltonian as a weak*limit on the *kinematical* Hilbert space $\mathcal{H}_{\text{kin}}$ w.r.t. the weak*topology defined by $\mathcal{V}^* = \mathcal{H}_{\text{diff}}$.[18] In that interpretation, only the rhs of (5.9) is meaningful—but let us repeat that there is *no* state in $\mathcal{H}_{\text{kin}}$ that could be interpreted as '$\lim_{\epsilon \to 0} \hat{H}[N, \epsilon]\Psi$'.[19]

Independently of whether or not one adopts the 'habitat' point of view, an essential ingredient in existing constructions is the 'diffeomorphism covariance' of the Hamiltonian. This means that by transferring the action of $\hat{H}$ to the space of diffeomorphism invariant states, the existence of the above limit can be ensured by the fact that the habitat $\mathcal{V}^*$ is sensitive only to the diffeomorphism invariant characteristics of the regulator. This is also borne out by the construction of [29] which takes $\mathcal{V}^*$ to be the (larger) space of 'vertex smooth states'. (But note that from the point of view of the Hamiltonian action, going to the slightly larger habitat of 'vertex-smooth' states does not make much of a difference; namely, when acting on a spin network state, the Hamiltonian does not shift nodes around. Hence its action on $\mathcal{H}_{\text{diff}}$ or on the larger habitat of 'vertex-smooth' states is essentially the same.) In addition to restricting the choice of $\mathcal{V}^*$, requiring the limit (5.9) to exist eliminates at least some of the ambiguities inherent in the definition of the Hamiltonian.

To sum up, the Hamiltonian $\hat{H}[N]$ can only be defined, via (5.9), as a *weak limit*. This is reminiscent of the LSZ formalism of ordinary quantum field theory, where one also defines the asymptotic field operators as weak rather than strong limits. It is remarkable that the known constructions make essential use of the diffeomorphism covariance, and thus implicit use of the constraints, already in the very definition of the Hamiltonian. However, for all we can tell, the question as to which $\mathcal{V}^*$ is 'best' is not settled. On a more practical note, and independently of the choice of $\mathcal{V}^*$, we note that the concrete evaluation of the limit requires simultaneous knowledge of the action of the original regulated Hamiltonian $\hat{H}[N, \epsilon]$ on *all* elements of $\mathcal{S}$. Furthermore, *explicit* expressions for the dual Hamiltonian $\hat{H}^*[N]$ that could be used for any practical application of the formalism are even more difficult to obtain than for the Hamiltonian acting on spin network states, and certainly beyond the scope of our modest efforts in this section. Nevertheless, we should stress once more that the dual Hamiltonian $\hat{H}^*$ is our true object of interest, and that all further considerations (constraint algebra, semi-classical states, and so on) should be based on it, rather than on the regulated Hamiltonian acting on $\mathcal{S}$. With these caveats in mind, let us now proceed.

---

[18] There are actually infinitely many such operator topologies w.r.t. the dual pair $\langle \mathcal{S}^*, \mathcal{S} \rangle$, depending on the choice of $\mathcal{V}^*$ (see, e.g., [136], p 88). The bigger the subspace $\mathcal{V}^*$, the stronger the topology.

[19] A more recent proposal to circumvent these difficulties is the so-called Master Constraint Programme of [26], where the Hamiltonian constraint is replaced by another constraint ('master constraint') $\propto \int \mathrm{d}^3 x (\tilde{E})^{-1/2} (H_0(x))^2$ obtained by squaring the original Hamiltonian. This operator can be implemented as a positive self-adjoint operator directly on $\mathcal{H}_{\text{diff}}$ (after suitable regularizations), so questions about 'habitats' are sidestepped.

To motivate the form of the *quantum Hamiltonian* one starts with the classical expression, written in loop variables. Despite the simplifications brought about by equation (3.21), the Hamiltonian constraint still looks formidable,

$$H[N] = \int_\Sigma d^3x \, N \frac{\tilde{E}_a^m \tilde{E}_b^n}{\sqrt{\det \tilde{E}}} \left(\epsilon^{abc} F_{mnc} - \frac{1}{2}(1+\gamma^2) K_{[m}{}^a K_{n]}{}^b \right). \tag{5.10}$$

In order to write the constraint in terms of only holonomies and fluxes, one has to eliminate the inverse square root as well as the extrinsic curvature factors. This can be done [74] using the relations (3.26)–(3.28). Inserting these into the Hamiltonian constraint one obtains the expression

$$H[N] = \int_\Sigma d^3x \, N \epsilon^{mnp} \, \text{Tr}\big(F_{mn}\{A_p, V\} - \frac{1}{2}(1+\gamma^2)\{A_m, \bar{K}\}\{A_n, \bar{K}\}\{A_p, V\} \big). \tag{5.11}$$

This expression is the starting point for the construction of the quantum constraint operator.

The next step is to express the Ashtekar connection A_m as well as field strength F_{mn} and the extrinsic curvature $\bar{K}$ in terms of holonomies. As in lattice gauge theory, this requires the introduction of a cell structure. In the *classical* expression one splits the target space into cells of size $\sim \epsilon^3$ in *coordinate* space. In order to express connections and curvatures in terms of holonomies, one now has to choose edges e_m within each cell, along which the holonomies are going to be evaluated. Once the edges e_m are chosen, one uses that the connection along a line of *coordinate* distance ϵ can be approximated by

$$\left\{ \int_e A, V \right\} = -h_e[A]^{-1}\{h_e[A], V\} + o(\epsilon),$$

$$\left\{ \int_e A, \bar{K} \right\} = -h_e[A]^{-1}\{h_e[A], \bar{K}\} + o(\epsilon). \tag{5.12}$$

For the field strength one has the standard lattice expression

$$\int_{P(\epsilon)} F[A] = \frac{1}{2}\big(h_{\partial P(\epsilon)}^{-1}[A] - h_{\partial P(\epsilon)}[A] \big) + o(\epsilon^3), \tag{5.13}$$

where $P(\epsilon)$ is a two-dimensional infinitesimal plaquette with boundary $\partial P(\epsilon)$, which shrinks to zero as $\epsilon \to 0$. The extrinsic curvature is expressed using the relation (3.28), into which one substitutes (5.13). Finally, as for the volume operator, one replaces integrals with Riemann sums over the cells, which we will again label with α.

In order to quantize the classical Hamiltonian (5.11) with the above regularization, we next elevate all classical objects to quantum operators as described in the foregoing sections, and replace the Poisson brackets in (5.11) by quantum commutators. The resulting *regulated Hamiltonian* then reduces to a sum over the vertices v_α of the spin network with lapses $N(v_\alpha)$

$$\hat{H}[N, \epsilon] = \sum_\alpha N(v_\alpha) \epsilon^{mnp} \, \text{Tr}\big((h_{\partial P_{mn}(\epsilon)} - h_{\partial P_{mn}(\epsilon)}^{-1}) h_p^{-1}[h_p, \hat{V}] \big)$$

$$- \frac{1}{2}(1+\gamma^2) \sum_\alpha N(v_\alpha) \epsilon^{mnp} \, \text{Tr}\big(h_m^{-1}[h_m, \bar{K}] h_n^{-1}[h_n, \bar{K}] h_p^{-1}[h_p, \hat{V}] \big), \tag{5.14}$$

where we have already assumed a specific ordering of the operators. The fact that the regularization parameter ϵ drops out 'miraculously' from this expression, and the integral (5.11) can be replaced by the sum (5.14), requires the Hamiltonian density to have the correct weight, and would not work with weight two density (3.21), cf footnote 6. If one does not want to make reference to any particular spin network, the Hamiltonian can also be expressed more abstractly as a continuous sum

$$\hat{H}[N, \epsilon] = \sum_{\mathbf{x} \in \Sigma} N(\mathbf{x}) \hat{H}(\mathbf{x}; \epsilon) \tag{5.15}$$

which, on any given spin network, reduces to a finite sum as in (5.14). In the remainder, we will keep the regularization parameter $\epsilon > 0$ fixed, and not always indicate it explicitly in the formulae below.

A first stumbling block here is the presence of the volume operator $\hat{V}$ in (5.14). Although its square has been expressed in terms of fluxes in section 4.4.2, the spectrum of this operator is only partially known, see, however, [41, 115, 117] for recent progress (recall that $\hat{q}$ is not necessarily positive). The further evaluation of (5.14) on a given spin network wavefunction would now, in particular, require a diagonal basis of spin network states, on which we could determine the square root of $|\hat{q}|$ for each node (cf (4.30). To treat all the relevant terms, this diagonalization would have to be repeated several times, along the lines of [117], with intermittent application of other operators. Although all the steps required in this calculation are thus clearly laid out, presently available technology does not, as far as we are aware, allow a complete evaluation of $\hat{H}[N, \epsilon]$ on a given, but arbitrary spin network state. For this reason, we will only be able to exploit certain qualitative properties of the volume operator in our analysis of the Hamiltonian constraint below. We will thus adopt a pragmatical approach in the remainder, performing all computations with the local trilinear operator $\hat{q}$ instead of the volume operator itself, and assume that we can divide by the relevant power of the volume at the end (thereby postponing the most difficult part of the calculation)[20]. Accordingly, the arguments in this section should be viewed as no more than a *Gedanken calculation*, and there is no claim that this is the way the calculation should ultimately be done.

Apart from obvious ordering ambiguities, and the difficulties with volume operator just described, the action of the quantum expression (5.14) on a generic spin network state also depends strongly on the choice of the coordinate system on Σ. To be more precise, the action depends on whether or not the basis vectors (i.e. the directions m, n and p) are aligned with the spin network edges at a given node. There exist a large number of possibilities for these alignments, which are *not* diffeomorphically equivalent, and one might thus simply define the action of the Hamiltonian by employing a particular alignment, as LQG does. Some of these possibilities are explicitly listed below. We regard these choices as genuine ambiguities in the quantization procedure, in the sense that observables quantities may depend on them, if they can be shown to exist at all. This potential prescription dependence, and the abundance of possible outcomes for *physical quantities* are most difficult to digest from the perspective of conventional quantum field theory, especially because, the result for the corresponding classical quantity does not depend on such choices when the limit $\epsilon \to 0$ is taken.

To get an idea of the complications, let us consider the action of the first term in the regulated Hamiltonian constraint, which is occasionally called the 'Euclidean Hamiltonian'[21] (because with a Euclidean signature and $\gamma = \pm 1$, the bothersome second term in (5.14) would be absent, as it is for $\gamma = \pm i$ with Lorentzian signature). We will act with this operator on a 3-valent vertex v in a spin network, for given $\epsilon > 0$. The spin network wavefunction on which we will act is given by

$$\psi = \left(h_{e_1}^{j_1}\right)_{\alpha_1\beta_1} \left(h_{e_2}^{j_2}\right)_{\alpha_2\beta_2} \left(h_{e_3}^{j_3}\right)_{\alpha_3\beta_3} C_{\beta_1\beta_2\beta_3}^{j_1 j_2 j_3} \cdots, \tag{5.16}$$

where the dots denote other parts of the spin network. Since the Hamiltonian acts node by node, we can restrict our attention to this local structure at v. Let us also write out all indices on the Hamiltonian operator, restricted to that part which acts at v,

$$\hat{H} = \text{sgn}(\hat{q})|\hat{q}|^{-1/2} \epsilon^{mnp} \left(F_{mn}^{j}\right)_{\lambda_1\lambda_2} \left(h_p^{j}\right)_{\lambda_2\lambda_3}^{-1} \left[\hat{q}, \left(h_p^{j}\right)_{\lambda_3\lambda_1}\right] + \cdots, \tag{5.17}$$

---

[20] We note that replacements such as $[\ldots, \hat{V}] \longrightarrow \sum_v \text{sgn}(\hat{q})|\hat{q}|^{-1/2}[\ldots, \hat{q}]$ introduce further operator ordering ambiguities into the definition of the Hamiltonian operator.
[21] And thus often designated as $\mathcal{C}^{\text{Eucl}}$.

where the dots now denote terms which act at other vertices. The choice of the spin j is a new ambiguity. In the first step, the operator $h_p^{-1}[h_p, \hat{q}] = \hat{q} - h_p^{-1}\hat{q}h_p$ acts at the vertex of the network. When $\hat{q}$ in the second term acts on three holonomies already present in the spin network, the result cancels against the first term. (Note that the result only vanishes because we are using $\hat{q}$ instead of $\hat{V}$, so that the local volume operator acts tri-linearly in holonomy derivatives. This is no longer true when one uses the square-root expression for $\hat{V}$ in the Hamiltonian constraint).

When $\hat{q}$ acts on two existing holonomies and on the h_p factor, however, the result is non-trivial. In order to work out this non-trivial action of $\hat{q}$, it is necessary to fix the orientation of the coordinate system on Σ. We will here choose the m, n and p directions to be aligned with the three edges e_1, e_2 and e_3 of the spin network respectively, as is usually done in the LQG literature. Note, however, that this is not the only possibility; we will return to this arbitrariness shortly. In effect, this means that the quantum Hamiltonian is *postulated* to be

$$\hat{H} = \sum_{\alpha} \sum_{I,J,K} N(v_\alpha) L^{IJK} \mathrm{Tr}\big((h_{\partial P_{IJ}} - h_{\partial P_{IJ}}^{-1}) h_{s_K}^{-1}[h_{s_K}, \hat{V}]\big)$$
$$- \frac{1}{2}(1 + \gamma^2) \sum_{\alpha} \sum_{I,J,K} N(v_\alpha) L^{IJK} \mathrm{Tr}\big(h_{s_I}^{-1}[h_{s_I}, \bar{K}] h_{s_J}^{-1}[h_{s_J}, \bar{K}] h_{s_K}^{-1}[h_{s_K}, \hat{V}]\big).$$
$$(5.18)$$

The second sum is a sum over all triplets of edges emanating from the vertex v_α in the centre of the cell α. The boundary of a plaquette in the plane spanned by the edges s_I and s_J is denoted by ∂P_{IJ}. Finally, the value of the constant, fully anti-symmetric tensor L^{IJK} depends on the particular vertex α.

Having made this coordinate choice, we can now return to our calculation. The situation is as in the first graph of figure 10. Because the small edge p is extending in the e_3 direction, we can use the tensor product decomposition rule to first 'merge' the h_p holonomy with (a segment of) the h_{e_3} holonomy of the spin network. For two arbitrary holonomies along identical edges, this decomposition rule reads

$$\big(h_e^{j_1}\big)_{\alpha_1 \beta_1} \big(h_e^{j_2}\big)_{\alpha_2 \beta_2} = \sum_k C^{j_1 j_2 k}_{\alpha_1 \alpha_2 \gamma_1} C^{j_1 j_2 k}_{\beta_1 \beta_2 \gamma_2} \big(h_e^k\big)_{\gamma_1 \gamma_2}. \tag{5.19}$$

By splitting the edge e_3 into a 'small' piece, overlapping with the p edge, and a 'larger'[22] remainder, we can write

$$\big(h_{e_3}^{j_3}\big)_{\alpha_3 \beta_3} = \big(h_{e_3'}^{j_3}\big)_{\alpha_3 \gamma_3} \big(h_{e_3''}^{j_3}\big)_{\gamma_3 \beta_3}, \tag{5.20}$$

where the edge e_3'' is the small piece of the edge, equal to p. Applying the tensor product rule to the holonomies h_p and $h_{e_3''}$ one now obtains (recalling that also m and n are aligned with e_1 and e_2 respectively)

$$\hat{H}\psi = \mathrm{sgn}(\hat{q}) |\hat{q}|^{-1/2} \big(F_{e_1 e_2}^j\big)_{\lambda_1 \lambda_2} \big(h_{e_3'}^j\big)_{\lambda_2 \lambda_3}^{-1} \big(h_{e_3'}^{j_3}\big)_{\alpha_3 \gamma_3} \hat{q}$$
$$\times \left(\sum_{k_3} C^{j j_3 k_3}_{\lambda_3 \gamma_3 \rho_3} C^{j j_3 k_3}_{\lambda_1 \beta_3 \sigma_3} \big(h_{e_3''}^{k_3}\big)_{\rho_3 \sigma_3} \right) \big(h_{e_1}^{j_1}\big)_{\alpha_1 \beta_1} \big(h_{e_2}^{j_2}\big)_{\alpha_2 \beta_2} C^{j_1 j_2 j_3}_{\beta_1 \beta_2 \beta_3} \dots . \tag{5.21}$$

The factor in brackets corresponds to the two merged holonomies. We can now work out the action of the first line of this expression.

---

[22] We put quotation marks here, because there is no *a priori* reference metric w.r.t. which the lengths of these subedges could be measured.

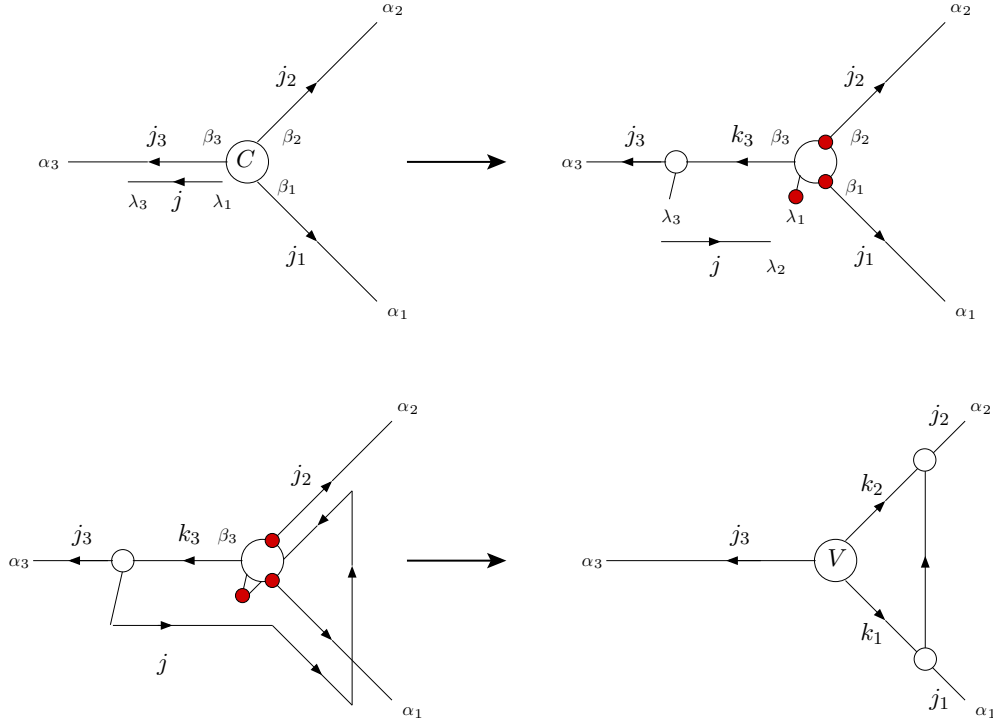**Figure 10.** The action of the first term in the Hamiltonian constraint on a spin network, in the special case that the m, n and p directions are aligned with three existing edges of the spin network. The three edges are drawn in the plane for convenience, but should span a tetrahedron in order for the volume operator to act non-trivially. Filled dots denote insertions of τ matrices. In the last two steps, we have only drawn one term in the plaquette (5.23).

As explained in section 4.4.2, the square of the local volume operator inserts a τ matrix at the end of the three holonomies on which it acts. In our case, this means that the matrices are connected to the indices λ_1, β_1 and β_2, as these indices correspond to the end-points of the three holonomies which are located at the vertex v (remember that in expression (5.17), the only non-zero contribution arises when $\hat{q}$ acts on two holonomies of the spin network and on the holonomy $(h_p)_{\lambda_3\lambda_1}$ inside the Hamiltonian constraint; the λ_1 index is the one located at the vertex). Up to a normalization factor, this leads to

$$\hat{H}\psi = \text{sgn}(\hat{q})|\hat{q}|^{-1/2}\epsilon^{abc}\left(F_{e_1e_2}^j\right)_{\lambda_1\lambda_2}\left(h_{e_3''}^j\right)_{\lambda_2\lambda_3}^{-1}\left(h_{e_3'}^{j_3}\right)_{\alpha_3\gamma_3}$$

$$\times (\tau^a)_{\lambda_1\tau_3}\left(\sum_{k_3} C_{\lambda_3\gamma_3\rho_3}^{jj_3k_3} C_{\tau_3\beta_3\sigma_3}^{jj_3k_3}\left(h_{e_3''}^{k_3}\right)_{\rho_3\sigma_3}\right)\left(h_{e_1}^{j_1}\tau^b\right)_{\alpha_1\beta_1}\left(h_{e_2}^{j_2}\tau^c\right)_{\alpha_2\beta_2} C_{\beta_1\beta_2\beta_3}^{j_1j_2j_3}\cdots.$$

$$(5.22)$$

What is left to do is to insert the plaquette. Because of our choice of coordinate system on Σ, the plaquette extends in the plane spanned by the e_1 and e_2 edges. In the limit $\epsilon \to 0$, the field strength can be expressed in terms of elementary holonomies as

$$\left(F_{e_1e_2}^j\right)_{\lambda_1\lambda_2} \simeq \left(h_{e_1}^j\right)_{\lambda_1\kappa_1}\left(h_{e_1e_2}^j\right)_{\kappa_1\kappa_2}\left(h_{e_2}^j\right)_{\kappa_2\lambda_2} - \text{(inverted loop)}, \qquad (5.23)$$

where $h_{e_1e_2}^j$ denotes a holonomy in the direction which connects the edge e_1 with the edge e_2. When we insert this expression in (5.22), we can again split the edges into a small and
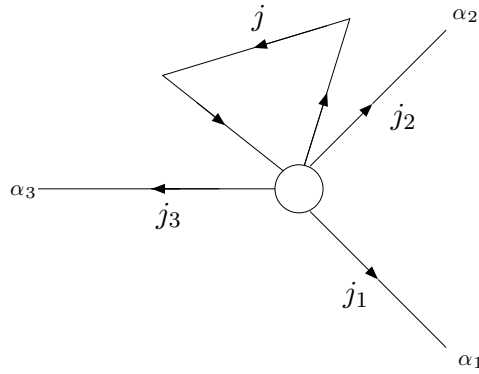
Figure 11. The action of the first term in the Hamiltonian constraint, for the case in which the m and n directions on Σ are not aligned with any of the edges emanating from the vertex.

large piece as in (5.20) and then use the tensor product decomposition (5.19) to merge the two holonomies in the e_1 direction and the two holonomies in the e_2 direction. This introduces four new Clebsch–Gordan coefficients, with the result

$$\hat{H}\psi = \text{sgn}(\hat{q})|\hat{q}|^{-1/2}\epsilon^{abc}\left(h^{j_1}_{e'_1}\right)_{\alpha_1\gamma_1}\left(h^{j_2}_{e'_2}\right)_{\alpha_2\gamma_2}\left(h^{j_3}_{e'_3}\right)_{\alpha_3\gamma_3}\left(h^{j}_{e''_3}\right)^{-1}_{\lambda_2\lambda_3}\left(h^{j}_{e_1e_2}\right)_{\kappa_1\kappa_2}C^{j_1j_2j_3}_{\tau_1\tau_2\beta_3}$$

$$\times (\tau^c)_{\beta_1\tau_1}(\tau^b)_{\beta_2\tau_2}(\tau^a)_{\lambda_1\tau_3}\left(\sum_{k_3}C^{jj_3k_3}_{\lambda_3\gamma_3\rho_3}C^{jj_3k_3}_{\tau_3\beta_3\sigma_3}\left(h^{k_3}_{e''_3}\right)_{\rho_3\sigma_3}\right)$$

$$\times\left(\sum_{k_1}C^{jj_1k_1}_{\lambda_1\beta_1\sigma_1}C^{jj_1k_1}_{\kappa_1\gamma_1\rho_1}\left(h^{k_1}_{e''_1}\right)_{\rho_1\sigma_1}\right)\left(\sum_{k_2}C^{jj_2k_2}_{\lambda_2\beta_2\sigma_2}C^{jj_2k_2}_{\kappa_2\gamma_2\rho_2}\left(h^{k_2}_{e''_2}\right)_{\rho_2\sigma_2}\right)\dots. \quad (5.24)$$

Altogether, these manipulations have introduced three new vertices, at locations ϵ away along each edge, and they have also modified the original vertex. The three new vertices are given by

$$(V^{(1)})^{k_1j_1j}_{\rho_1\gamma_1\kappa_1} = C^{jj_1k_1}_{\kappa_1\gamma_1\rho_1}, \qquad (V^{(2)})^{k_2j_2j}_{\rho_2\gamma_2\kappa_2} = C^{jj_2k_2}_{\kappa_2\gamma_2\rho_2}, \qquad (V^{(3)})^{k_3j_3j}_{\rho_3\gamma_3\lambda_3} = C^{jj_3k_3}_{\lambda_3\gamma_3\rho_3}, \quad (5.25)$$

while the old vertex has been changed to

$$V^{k_1k_2k_3j}_{\sigma_1\sigma_2\sigma_3\lambda_2} = C^{jj_1k_1}_{\lambda_1\beta_1\sigma_1}C^{jj_2k_2}_{\lambda_2\beta_2\sigma_2}C^{jj_3k_3}_{\tau_3\beta_3\sigma_3}\times\epsilon^{abc}(\tau^b)_{\beta_1\tau_1}(\tau^c)_{\beta_2\tau_2}(\tau^a)_{\lambda_1\tau_3}\times C^{j_1j_2j_3}_{\tau_1\tau_2\beta_3}. \quad (5.26)$$

One can still merge $\left(h^{j}_{e''_3}\right)^{-1}_{\lambda_2\lambda_3}$ and $\left(h^{k_3}_{e''_3}\right)_{\rho_3\sigma_3}$ using the tensor product decomposition rule, which will reduce the vertex $V^{(3)}$ to an overall factor times a Kronecker delta. We will refrain from spelling out these details.

Let us now comment on other possible coordinate systems and alignments. The use of the tensor product decomposition (5.19) depends crucially on the edges of the spin network being aligned with the edges of the plaquette (and, through (5.14), with the coordinate system on Σ). If this is not the case for the m, n and e_1, e_2 edges, the plaquette will not get glued into the spin network, but instead remains freely floating, only connected at the original vertex (see figure 11). Similarly, if the p edge is not aligned with e_3, the plaquette will only be connected to the spin network through the τ^a matrix in the construction above. In the LQG literature, such cases are excluded by hand, by always choosing a local coordinate system around each vertex which is completely aligned with (three of the) edges emanating from the vertex. This choice is usually justified by invoking background independence, which would be violated otherwise [11].

We have so far not discussed the second term in the constraint (5.14), which arose from the product of extrinsic curvatures in (5.10). This term is hardly ever discussed in detail, and with good reason: it is far more complicated than the first one. In fact, the technical complications which we have just described pile up in such a way as to make any concrete calculation almost unfeasible. The complexity of this term stems from the complexity of the intermediate expression

$$[A_I, \bar{K}] = h_{e_I}^{-1}\big[h_{e_I}, \epsilon^{JKL}\big[\big(h_{\alpha_{JK}} - h_{\alpha_{JK}}^{-1}\big)h_{e_L}^{-1}\big[h_{e_L}, \hat{V}\big], \hat{V}\big]\big], \qquad (5.27)$$

which is obtained using (3.26)–(3.28). One recognizes the combination of operators which opens up the vertex and inserts a plaquette, but there is an additional volume operator which applies further changes to the node. If the coordinate system on Σ is aligned in such a way that all three directions coincide with an existing edge, the action of the second term in the constraint is roughly to glue in three new edges, introducing six new vertices. But as the discussion above shows, there is a plethora of other possibilities.

Let us, for clarity, summarize the problems which we see with the Hamiltonian constraint.

- *Ambiguities.* The construction of the Hamiltonian outlined here is plagued with a variety of ambiguities. Firstly, a choice has to be made for the coordinate system on Σ. The action of the Hamiltonian depends in a qualitative way on this choice. Secondly, there is an ambiguity in choosing the representation in which the trace is evaluated (i.e. the representation in which the field strength F_{mn} and the holonomy h_p are taken). Thirdly, there are operator ordering ambiguities. Finally, and less conspicuously, there may exist $\hbar$-dependent 'higher order' corrections to the Hamiltonian (5.14), which are still compatible with diffeomorphism covariance, but would not affect the classical limit. Not only will physical predictions depend on the choices made to fix these ambiguities, but these choices are important in order to determine whether or not the limit $\epsilon \to 0$ of (5.9) exists at all.

- *Ad hoc choices.* The quantum constraint should be well-defined and unambiguous on the full space of spin-network states. If this is not the case, it should be shown that the choices do not matter for the physical predictions, or alternatively that a particular choice is singled out by physical consistency. By focussing only on the operator (5.18), it becomes impossible to tell whether the constraint operator is indeed independent of the coordinate choice on Σ.

- *Ultralocality.* From the explicit discussion of the Hamiltonian constraint it should be clear that its action is always 'ultralocal', in the sense that all changes to the spin network are made in an $\epsilon \to 0$ neighbourhood of a given vertex, while the spin network graph is kept fixed [137–139]. More specifically, it has been argued [29] that the Hamiltonian acts at a particular vertex only by changing the intertwiners at that vertex, namely

$$\hat{H}(v)|\Gamma, \{j\}, \{C\}\rangle = \sum_{I,J} U^a |\Gamma, \{j\}, C_a(\Gamma, \{j\}, I, J)\rangle \qquad (5.28)$$

where U^a is the traceless part of the holonomy associated with the plaquette P_{IJ} and C_a a triplet of new intertwiners at v, and the sum is over pairs of edges I, J entering the vertex. This is in contrast to what happens in lattice field theories. There the action of the Hamiltonian always non-trivially links two existing nodes, the plaquettes are by construction always spanned between existing nodes, and the continuum limit involves the lattice as a whole, not only certain sub-plaquettes.

6. Constraint algebra

As we already emphasized in the introduction, the proper closure of the quantum constraint algebra is perhaps *the* crucial consistency requirement that must be met by any candidate theory of quantum gravity. Of course, the devil hides in the meaning of the word 'proper'. To find out where it might possibly hide, we now return to the points raised at the end of section 5.1. We will not bother here with the commutators involving the Gauss constraints, which are straightforward to take care of. The canonical (Poisson) brackets involving the diffeomorphism generators read

$$\{D_m[\vec{M}], D_n[\vec{N}]\}_{\text{PB}} = \int_\Sigma \mathrm{d}^3x\,(M^m\partial_m N^n - N^m\partial_m M^n)D_n, \tag{6.1}$$

$$\{D_m[\vec{M}], H_0[N]\}_{\text{PB}} = \int_\Sigma \mathrm{d}^3x\,(M^m\partial_m N - N\partial_m M^m)H_0. \tag{6.2}$$

Here $D_m \equiv e_m{}^a D_a$ is the diffeomorphism generator, while the smeared Hamiltonian and smeared diffeomorphism generators are defined as

$$H_0[M] := \int_\Sigma \mathrm{d}^3x\, H_0(\mathbf{x})M(\mathbf{x}), \qquad D_m[\vec{M}] := \int_\Sigma \mathrm{d}^3x\, D_m(\mathbf{x})M^m(\mathbf{x}). \tag{6.3}$$

Because infinitesimal diffeomorphisms do not actually exist on $\mathcal{H}_{\text{kin}}$, only the exponentiated version of (6.1) makes sense in LQG. For the same reason, relation (6.2) is implemented by considering only finite (exponentiated) diffeomorphisms φ. Given a realization of φ on the Hilbert space through a unitary operator $\hat{U}[\varphi]$, the bracket (6.2) can be exponentiated formally to[23]

$$\hat{U}[\varphi]\hat{H}_0[N]\hat{U}[\varphi]^{-1} = \hat{H}_0[\varphi^*N]. \tag{6.5}$$

These relations constitute what we would call the 'kinematical part' of the constraint algebra.

The crucial, remaining relation is the commutator of two Hamiltonian constraints, whose classical analogue reads

$$\{H_0[M], H_0[N]\}_{\text{PB}} = \int_\Sigma \mathrm{d}^3x\,(M\partial_m N - N\partial_m M)g^{mn}D_n. \tag{6.6}$$

Together with the 'kinematical relations' above, this relation encapsulates the closure of the classical algebra of spacetime diffeomorphisms in the Hamiltonian formulation (we should note, however, that the relation between timelike diffeomorphisms on the one hand, and the 'motion' generated by the Hamiltonian constraint on the other, is quite subtle [141]). To be sure, the proper closure of this part of the quantum constraint algebra, i.e. the issue of promoting (6.6) to a relation between operators acting on a suitable Hilbert space of states, is also one of the major unsolved problems of the conventional geometrodynamics approach. There it has not been possible to resolve the ordering and other ambiguities, and to get rid of the infinities resulting from the clash of functional differential operators (the situation is partly better, though, in certain supergravity models whose constraint superalgebra always contains the bosonic constraint algebra as a subalgebra [76]).

---

[23] When this relation is understood to act on a state in $\mathcal{H}_{\text{diff}}$, one can use invariance of the state under diffeomorphisms, $\hat{U}[\varphi]|\psi\rangle = |\psi\rangle$, to rewrite equation (6.5) as [140]

$$[\hat{U}[\varphi], \hat{H}_0[N]] = \hat{H}_0[\varphi^*N - N]. \tag{6.4}$$

6.1. The bosonic string as an example

Before continuing our discussion of the quantum constraint algebra in canonical gravity, we would like to recall that there does exist a model which can be solved exactly, and all the way. Hence it can serve as an example of how quantization should work. This model is the bosonic string (in a Polyakov-type formulation), when viewed as a *bona fide* model of matter coupled quantum gravity in two dimensions (see also [142] for an earlier discussion of the bosonic string in this context). For this purpose, we only need to forget about the target space interpretation and its ramifications, such as string scattering, or higher order corrections in g_s, involving splitting and joining of strings (in the present context, the latter processes would amount to changes of the spatial topology, 'baby universes' and the like, hence, 'third quantization'). The resulting theory is manifestly independent of the two-dimensional spacetime background, free of divergences and other pathologies, and also possesses nice observables (vertex operators). More specifically, the diffeomorphism and scalar constraints D and H are given by (see, e.g., [143] or [144])

$$
D[f] = \frac{1}{2} \int_0^{2\pi} d\sigma \ f(\sigma) P_\mu(\sigma) \partial_\sigma X^\mu(\sigma),
$$

$$
H[N] = \frac{1}{4} \int_0^{2\pi} d\sigma \ N(\sigma)(P_\mu(\sigma) P^\mu(\sigma) + \partial_\sigma X_\mu(\sigma) \partial_\sigma X^\mu(\sigma))
$$

(6.7)

with shift $f(\sigma)$ and lapse $N(\sigma)$. The classical constraint algebra, given by (6.1), (6.2) and (6.6) in four dimensions, reduces to

$$
\{D[f_1], D[f_2]\}_{\text{PB}} = \tfrac{1}{2} D[f_1 \partial f_2 - f_2 \partial f_1],
$$

$$
\{D[f], H[N]\}_{\text{PB}} = \tfrac{1}{2} H[f \partial N - N \partial f],
$$

$$
\{H[N_1], H[N_2]\}_{\text{PB}} = \tfrac{1}{2} D[N_1 \partial N_2 - N_2 \partial N_1].
$$

(6.8)

Canonical quantization is straightforward in the conformal gauge. The state space of the theory $\mathcal{F}$ is simply the tensor product of the space of one-particle wavefunctions and an infinite number of harmonic oscillator Hilbert spaces associated with the excited string modes. Composite operators become well defined, i.e. have finite matrix elements between arbitrary states in $\mathcal{F}$, through normal ordering. Both the kinematical and the Hamiltonian constraint are Hermitian operators. With the normal ordered energy momentum tensors $T_{\pm\pm}(\sigma) \equiv: (H(\sigma) \pm D(\sigma)):$, the classical constraint algebra is modified by the anomaly (the central term), which is responsible for almost everything that is non-trivial about string theory,

$$
[T_{\pm\pm}(\sigma), T_{\pm\pm}(\sigma')] = \delta'(\sigma, \sigma')(T_{\pm\pm}(\sigma) + T_{\pm\pm}(\sigma')) + \hbar c \delta'''(\sigma, \sigma').
$$

(6.9)

Our main point in repeating this well-known story here is that this algebra holds on the full space of states, i.e. prior to the imposition of any constraints, hence closes *off-shell*. Only at this point can we impose and solve the quantum constraints to obtain the subspace $\mathcal{F}_0 \subset \mathcal{F}$ of *physical states*—with the well-known result ('No Ghost Theorem') that negative norm states are absent if the number of scalar fields is $\leqslant 26$. Because the Hilbert spaces $\mathcal{F}_0$ and $\mathcal{F}$ of physical and unphysical states, respectively, are well defined, there is no need to discuss 'habitats' here. Of course, one key simplifying feature of the string model must be pointed out here, which is not shared by (matter coupled) gravity in higher dimensions: (6.8) is an honest (albeit infinite-dimensional) Lie algebra (and (6.9) closes in Dirac's sense), whereas (6.1), (6.2), and (6.6) are not. This Lie algebra furthermore possesses a triangular decomposition, allowing for a systematic study of its representations, and this fact is a crucial ingredient in establishing the above results.

Very recently, LQG methods have been applied to string theory, but have been shown to lead to results which are at variance with those obtained using Fock-space quantization

[145, 146]. It therefore does not appear that LQG can recover conventional results; instead it gives rise to an inequivalent quantization of the string model, in which the quantum constraint algebra is realized without a central term, and without any restriction on the target space dimension. The meaning of these results is currently subject of an intense debate, see [147, 148].

Another simple model which can be quantized all the way is dilaton gravity in $1 + 1$ dimensions. For this model the constraint algebra has been worked out for various inequivalent representations of the operators, and the physical consequences have been analysed in detail. See for instance [149–151]. See also [152] for a review of dilaton gravity with many further references.

6.2. On-shell versus off-shell closure

We have seen that LQG treats the diffeomorphism and Hamiltonian constraints in a very different manner. Diffeomorphism invariance is implemented by an averaging procedure that makes the states invariant under *finite* (i.e. 'exponentiated') diffeomorphisms. In LQG a canonical generator of diffeomorphisms simply does not exist, since the lack of weak continuity that goes with the scalar product (4.7) prevents us from differentiating finite diffeomorphisms so as to obtain 'infinitesimal' ones. By contrast, at least so far, group averaging is of no use, even at the formal level and allowing for whatever approximation scheme, when it comes to imposing the Hamiltonian constraint. Therefore the latter, whose cumbersome form we have exhibited in the foregoing section, must be analysed 'the traditional way'[24]. This restriction applies also to the commutator of two Hamiltonian constraint operators. A further (and well-known) difficulty is that, even for the classical theory, this commutator does not generate an ordinary Lie algebra, in the sense that there appear *field-dependent* structure 'constants'; as is well known, these lead to ordering problems and ambiguities in the quantum constraint algebra.

While there is general agreement as to what one means when one speaks of 'closure of the constraint algebra' in classical gravity (or any other classical constrained system), this notion is more subtle in the quantized theory. Let us therefore clarify first the various notions of closure that can arise: we see at least three different possibilities. As we explained in the foregoing section, the 'true' Hamiltonian $\hat{H}^*$ is defined as a weak limit on some habitat $\mathcal{V}^* \subset \mathcal{S}^*$ by means of (5.9). Assuming that this subspace is preserved by the action of $\hat{H}^*$, the strongest notion is 'off-shell closure' (or 'strong closure'), where one seeks to calculate

$$[\hat{H}^*[N_1], \hat{H}^*[N_2]] = \hat{O}^*(N_1; N_2) \tag{6.10}$$

without further restrictions on the states, on which this relation is supposed to hold—in other words on a sufficiently large 'habitat' $\mathcal{V}^*$ that does not only contain diffeomorphism invariant states. The operator $\hat{O}^*(N_1; N_2)$ here is the quantum analogue of the rhs of (6.6), and would be proportional to the diffeomorphism generator (but this commutator might also differ from the classical one by certain quantum modifications). Together with the 'kinematical part' of the quantum constraint algebra, the above relation would express the *quantum spacetime covariance* of the theory. Strong closure is realized for the string model, with the well-known result (6.9). If one could resolve the ordering ambiguities and eliminate all singularities, this is also the type of closure expected to work in conventional geometrodynamics. On the other hand, although the calculation in principle does make sense in LQG, too, the rhs of (6.10) may

---

[24] To overcome the difficulties, it has even been suggested that in analogy with the diffeomorphism generator, the Hamiltonian itself may actually not exist as an infinitesimal generator, but that only 'finite' translations in time might be well defined [53, 153].

not exist as an operator in LQG, as we explained above. Of course, even if the diffeomorphism generator does not make sense as an operator, it is still conceivable that a combination like $g^{mn} H_n$ does exist as a 'composite operator'. In this case, however, this operator would not generate a closed algebra: repeated commutation would generate an infinite tower of new operators, and one would therefore no longer be discussing the algebra of constraints.

At any rate, it appears that the goal of determining $\hat{O}^*(N_1; N_2)$ as a *bona fide* 'off-shell' operator on $\mathcal{V}^*$, and prior to the imposition of any constraints, is unattainable within the current framework of LQG. For this reason, LQG must resort to weaker notions of closure, by making partial use of the constraints. More specifically, equation (6.10) can be relaxed substantially by demanding only

$$[\hat{H}^*[N_1], \hat{H}^*[N_2]]\mathcal{X} = 0. \tag{6.11}$$

This 'weak closure' should hold for all states $\mathcal{X}$ in a restricted habitat $\mathcal{V}^*$ of states that are 'naturally' expected to be annihilated by the rhs of (6.10), and that are subject to the further requirement that the Hamiltonian can be applied twice without leaving the 'habitat'. The latter condition is, for instance, met by the 'vertex smooth' states of [29]. As shown in [28, 29], the commutator of two Hamiltonians indeed vanishes on this 'habitat', and one is therefore led to conclude that the full constraint algebra closes 'without anomalies'.

The same conclusion was already arrived at in an earlier computation of the constraint algebra in [74, 27], which was done from a different perspective (no 'habitats'), and is based on the choice $\mathcal{V}^* = \mathcal{H}_{\text{diff}}$, the 'natural' kernel of the rhs of (6.10). Here the first step consists in calculating the commutator of two regulated Hamiltonians (see section 5.2)

$$\hat{O}(N_1; N_2; \epsilon) := [\hat{H}[N_1, \epsilon], \hat{H}[N_2, \epsilon]], \tag{6.12}$$

for fixed but arbitrary $\epsilon > 0$ on the space $\mathcal{S}$ of spin network states. As for the Hamiltonian itself, letting $\epsilon \to 0$ in this commutator produces an uncountable sequence $\hat{O}(N_1, N_2; \epsilon)\Psi$ of mutually orthogonal states w.r.t. the scalar product (4.7); consequently, $\lim_{\epsilon \to 0} \hat{O}(N_1, N_2; \epsilon)$ again does not exist in the usual sense, but only as a weak* limit. More specifically, one shows that [27, 74]

$$\lim_{\epsilon \to 0} \langle \mathcal{X} \mid \hat{O}(N_1; N_2; \epsilon)\Psi \rangle = 0 \tag{6.13}$$

for all $\mathcal{X} \in \mathcal{H}_{\text{diff}}$, and for all $\Psi \in \mathcal{S}$ (the 'diffeomorphism covariance' of the Hamiltonian is again essential for this result); in this sense, the constraint algebra is free of anomalies. Let us stress that (6.11) and (6.13) are by no means the same: in (6.11) one uses the dual Hamiltonian defined in (5.9) (where the limit $\epsilon \to 0$ has already been taken) and performs the calculation on a subspace $\mathcal{V}^* \subset \mathcal{S}^*$, whereas the calculation of the commutator in (6.13) takes place on the space $\mathcal{S}$, that is, inside $\mathcal{H}_{\text{kin}}$, and the limit $\epsilon \to 0$ is taken only *after* computing the commutator of two regulated Hamiltonians on $\mathcal{S}$. In other words, these two operations (taking the limit $\epsilon \to 0$, and calculating the commutator) need not commute. Because with both (6.11) and (6.13), one forgoes the aim of finding an operatorial expression for the commutator $[\hat{H}^*[N_1], \hat{H}^*[N_2]]$ as an operator on $\mathcal{V}^*$, or the commutator (6.12) on the kinematical Hilbert space, making partial use of the constraints, we say (in a partly supergravity inspired terminology) that the algebra closes 'on-shell'.

At this point we should like to remark that computations done so far make only little use of the detailed structure of the terms in the Hamiltonian. In particular, (on-shell) closure is achieved independently of whether or not one takes into account the second term in (5.14), and there may exist (diffeomorphism covariant) modifications of the type discussed at the end of this section, or 'higher order' modifications of the Hamiltonian (5.14), which do not alter this result. The analysis of [29] uses only schematic formulae like (5.28) rather than *explicit*

operatorial expressions for the Hamiltonian. All this is contrary to naive expectations, because computations of more complicated gravity-matter systems (such as the supergravity constraint algebra studied in [75]) show that the detailed structure and the interplay of all terms in the constraints *is* indispensable already for the proper closure of the classical constraint algebra. This may indicate that crucial information is lost, and that, in a sense, the on-shell closure is already 'built into' the formalism via the diffeomorphism covariance of the Hamiltonian, and by restricting the calculation to special states (or matrix elements) only. Still, one would hope to be able in principle to extract from this computation some kind of explicit operatorial expression for the commutator. However, (at least our) attempts to actually *evaluate* (6.12) on a given spin network wavefunction Ψ along the lines of section 5.2, and before transferring the result to $\mathcal{V}^*$, get stuck quickly: the resulting state would be substantially more complicated than the (incomplete) expressions displayed there.

The procedure of making partial use of the constraints *prior* to the evaluation of the constraint algebra, as well as in the very construction of the Hamiltonian constraint, is very different from the conventional treatment of quantized canonical systems [60], as exemplified by the bosonic string in the previous subsection. There, one first studies the constraint operators and tries to make them well defined as quantum operators without constraining the allowed states, and *before* checking the closure of the constraint algebra. When one then in a second step computes the algebra, one usually encounters subtle quantum (i.e. $\hbar$-dependent) modifications of this algebra, whose determination requires great care. Because these modifications are subject to certain constraints (Wess–Zumino-type consistency conditions and their offspring in BRST cohomology), they can often be determined in a representation and gauge-independent way, as is the case for the group $\mathrm{Diff}(S^1)$. If one succeeds in working out this algebra and demonstrating its consistency, the quantum constraint algebra closes 'off-shell'. Only *after* ensuring the closure of the quantum constraint algebra does one proceed to impose and solve the quantum constraints. Let us also remind readers that, at least according to conventional (quantum field theory) wisdom, 'on-shell' calculations of the algebra might be 'empty': in general it is not correct to use a symmetry before it has been shown that it can be implemented without anomalies [154].

Why do we emphasize off-shell closure, despite the difficulties (ordering ambiguities, field dependent structure constants, and the like), which likewise plague other approaches to canonical quantum gravity? Our main reason for emphasizing this requirement is two-fold: first and most importantly, the off-shell closure of the quantum constraint algebra (which, as we pointed out, need not coincide with the classical algebra) encapsulates the key property of quantum gravity, namely *quantum spacetime covariance*. Secondly, this property furnishes an excellent means to distinguish between the 'correct' theory and a mere regularization: consider for instance replacing the kinetic part of the WDW operator by a smeared point-split operator. With this modification, all constraint operators become formally well defined, so one might contemplate 'defining' the quantized theory in terms of them, and simply proceed to study the solutions of these modified constraints. This procedure would, of course, suffer from the same kind of ambiguities encountered in section 5.2 and before, but it might still take us a long way if we did not pay due attention to the constraint algebra. If, however, we do subject the modified WDW Hamiltonian to this acid test, it is immediately clear that the resulting operator (or any other obtained by such a regularization) is 'wrong': it does not satisfy the correct algebra, whence the quantized theory would not satisfy the basic requirement of quantum spacetime covariance.

As another example, consider modifying the Hamiltonian constraint of string theory by multiplying it with an operator which commutes with all Virasoro generators,

$$(\hat{T}_{++} + \hat{T}_{--}) \longrightarrow (\hat{T}_{++} + \hat{T}_{--})\hat{C}, \tag{6.14}$$

where $[\hat{C}, \hat{T}_{++}] = 0$ and $[\hat{C}, \hat{T}_{--}] = 0$, while keeping the diffeomorphism constraint $(\hat{T}_{++} - \hat{T}_{--})$ unchanged. There are many such operators $\hat{C}$ in string theory. The simplest is to take $\hat{C}$ to be equal to the mass operator $P^\mu P_\mu$ minus some arbitrary positive integer, where P^μ is the conserved global momentum of the string (which indeed commutes with all Virasoro operators). In this way, we arrive at a realization of the constraint operators which is very similar to that used in LQG: the algebra of spatial diffeomorphisms is realized via a (projective) unitary representation, and the Hamiltonian constraint transforms covariantly as in (6.5) (the extra factor $\hat{C}$ does not matter, because it commutes with all constraints). In a first step, one can restrict attention to the subspace of states annihilated by the diffeomorphism constraint, the analogue of the space $\mathcal{H}_{\text{diff}}$. Imposing $(\hat{T}_{++} + \hat{T}_{--})\hat{C}|\text{phys}\rangle = 0$ on this subspace would now produce a 'non-standard' spectrum by allowing extra diffeomorphism invariant states of a certain prescribed mass, but the algebra would still close on-shell, i.e. on the 'habitat' of states annihilated by the diffeomorphism constraint. The point here is not so much whether this new spectrum is 'right' or 'wrong', but rather that in allowing such modifications which are compatible with on-shell closure of the constraint algebra, we introduce an infinite ambiguity and arbitrariness into the definition of the physical states. In other words, if we only demand on-shell closure as in LQG, there is no way of telling whether or not the vanishing of a commutator is merely accidental, that is, not really due to the diffeomorphism invariance of the state, but caused by some other circumstance.

This, then, is our main point: by weakening the requirements on the constraint algebra and by no longer insisting on off-shell closure, crucial information gets lost. This loss of information is reflected in the ambiguities inherent in the construction of the LQG Hamiltonian. It is quite possible that the LQG Hamiltonian admits many further modifications on top of the ones we have already discussed, for which the commutator continues to vanish on a suitably restricted habitat of states—in which case neither (6.11) nor (6.13) would amount to much of a consistency test.

7. Conclusions

String theory and LQG pursue the same goal, a consistent theory of quantum gravity, though with very different means. Both approaches address core issues of quantum gravity, but concentrate on complementary aspects of the problem, and have led to valuable insights. For this reason, the opinion has been voiced that string theory and LQG may ultimately merge together in a grand synthesis, or that LQG might become part of string theory [155, 156][25]. On the basis of the available evidence, this does not appear a likely outcome to us: the basic hypotheses underlying string theory and LQG seem impossible to reconcile without major modifications in either approach. There *is* a basic clash here, as reflected in such questions as to whether or not supersymmetry is necessary for a consistent quantization of gravity. One might thus say that the differences between the two approaches are such that, in the end, at most one 'can be right'.

We have reported here on the status of loop quantum gravity from a somewhat uncommon perspective. Our general conclusion is that, despite the optimism prevalent in many other reviews, more attention should be paid to basic aspects and unresolved problems of the theory. As many LQG practitioners may find this assessment too harsh, let us therefore summarize once more the basic point we have been trying to make in this review. The main issue here is

---

[25] Finding a connection between *classical* two-form gravity in four dimensions and topological string theory [157] can hardly be considered evidence for the connection between LQG and string theory. String theory provides a large variety of topological theories, of which the ones that appear in LQG are only special examples. More importantly, the quantization methods employed there are entirely different from those of LQG.

not so much specific details of the LQG approach, where there has been a lot of serious work and considerable progress, but rather the question: *what does one mean when one speaks of a consistent theory of quantum gravity?* And what are the basic properties that such a theory should satisfy? It is here that opinions differ. In the introduction, we have spelled out some of the criteria that *we* believe such a theory should meet. Apart from the question of the semi-classical limit (which is generally acknowledged to be a main outstanding problem of the LQG approach), and our insistence on the key role of the perturbative two-loop divergence (1.1), we have emphasized the importance of finding a good criterion for ensuring the *spacetime* covariance of the theory at the quantum level—after all, this, rather than merely covariance under spatial diffeomorphisms, is the essence of Einstein's theory. We have recalled the well-known fact that, in the canonical approach, it is the constraint algebra that encodes the information about spacetime covariance, and that everything hinges on how this algebra is to be implemented in the quantum theory. Whereas LQG proponents generally seem to be content with (6.11) or (6.13) as an expression of covariance, as well as the procedure of solving constraints 'in steps', using the diffeomorphism constraint already in the very definition of the Hamiltonian, we have tried to explain why we consider both (6.11) and (6.13) as too weak, hence insufficient, in this regard. Instead, we have proposed the *off-shell closure* of the quantum constraint algebra as a criterion of spacetime covariance. We have also argued that imposing off-shell closure may further reduce the large number of ambiguities present in the formalism, and in the definition of the Hamiltonian constraint, in particular. As we have stressed repeatedly, it is this constraint which lies at the core of canonical quantum gravity.

In summary, we feel that there are still important problems at a basic level that need to be addressed and resolved before one can tell whether or not the loop quantum gravity programme can succeed. We hope that the present review will provide an incentive to re-focus attention to these basic issues, and contribute to the debate between the different approaches.

Acknowledgments

References

[1] Goroff M H and Sagnotti A 1985 Quantum gravity at two loops *Phys. Lett.* B **160** 81
[2] Goroff M H and Sagnotti A 1986 The ultraviolet behavior of Einstein gravity *Nucl. Phys.* B **266** 709
[3] van de Ven A E M 1992 Two loop quantum gravity *Nucl. Phys.* B **378** 309–66
[4] Weinberg S 1979 Ultraviolet divergences in quantum gravity *An Einstein Centenary Survey* ed S Hawking and W Israel (Cambridge: Cambridge University Press) pp 790–832
[5] Weinberg S 1997 What is quantum field theory, and what did we think it was? *Preprint* hep-th/9702027
[6] Lauscher O and Reuter M 2002 Is quantum Einstein gravity nonperturbatively renormalizable? *Class. Quantum Grav.* **19** 483–92 (*Preprint* hep-th/0110021)

[7] Gibbons G W and Hawking S W 1977 Action integrals and partition functions in quantum gravity *Phys. Rev.* D **15** 2752–6

[8] Hawking S W 1979 The path-integral approach to quantum gravity *An Einstein Centenary Survey* ed S Hawking and W Israel (Cambridge: Cambridge University Press) pp 746–89

[9] Hartle J B and Hawking S W 1983 Wave function of the universe *Phys. Rev.* D **28** 2960–75

[10] Gambini R and Pullin J 1996 *Loops, Knots, Gauge Theories and Quantum Gravity* (Cambridge: Cambridge University Press)

[11] Thiemann T 2001 Introduction to modern canonical quantum general relativity *Preprint* gr-qc/0110034

[12] Ashtekar A and Lewandowski J 2004 Background independent quantum gravity: a status report *Class. Quantum Grav.* **21** R53 (*Preprint* gr-qc/0404018)

[13] Rovelli C 2004 *Quantum Gravity (Cambridge Monographs on Mathematical Physics)* (Cambridge: Cambridge University Press)

[14] Perez A 2004 Introduction to loop quantum gravity and spin foams *Preprint* gr-qc/0409061

[15] Horowitz G T 1991 Ashtekar's approach to quantum gravity *Strings and Symmetries* (Stony Brook: World Scientific) pp 154–67 (*Preprint* hep-th/9109002)

[16] Kuchar K V 1993 Canonical quantum gravity *General Relativity and Gravitation* ed R Gleiser, C Kozameh and O Moreschi (Bristol: Institute of Physics Publishing) pp 119–50 (*Preprint* gr-qc/9304012)

[17] Carlip S 2001 Quantum gravity: a progress report *Rep. Prog. Phys.* **64** 885 (*Preprint* gr-qc/0108040)

[18] Ashtekar A, Fairhurst S and Willis J L 2003 Quantum gravity, shadow states, and quantum mechanics *Class. Quantum Grav.* **20** 1031–62 (*Preprint* gr-qc/0207106)

[19] Thiemann T 2001 Gauge field theory coherent states (GCS): I. General properties *Class. Quantum Grav.* **18** 2025–64 (*Preprint* hep-th/0005233)

[20] Thiemann T and Winkler O 2001 Gauge field theory coherent states (GCS): II. Peakedness properties *Class. Quantum Grav.* **18** 2561–636 (*Preprint* hep-th/0005237)

[21] Thiemann T and Winkler O 2001 Gauge field theory coherent states (GCS) III: Ehrenfest theorems *Class. Quantum Grav.* **18** 4629–82 (*Preprint* hep-th/0005234)

[22] 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 (*Preprint* hep-th/0005235)

[23] Thiemann T 2002 Complexifier coherent states for quantum general relativity *Preprint* gr-qc/0206037

[24] 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 (*Preprint* gr-qc/0102038)

[25] Borissov R, De Pietri R and Rovelli C 1997 Matrix elements of Thiemann's Hamiltonian constraint in loop quantum gravity *Class. Quantum Grav.* **14** 2793–823 (*Preprint* gr-qc/9703090)

[26] Thiemann T 2003 The Phoenix project: master constraint programme for loop quantum gravity *Preprint* gr-qc/0305080

[27] Thiemann T 1996 Anomaly-free formulation of non-perturbative, four-dimensional Lorentzian quantum gravity *Phys. Lett.* B **380** 257–64 (*Preprint* gr-qc/9606088)

[28] 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 (*Preprint* gr-qc/9710018)

[29] Lewandowski J and Marolf D 1998 Loop constraints: a habitat and their algebra *Int. J. Mod. Phys.* D **7** 299–330 (*Preprint* gr-qc/9710016)

[30] Zeh H 1999 *The Physical Basis of the Direction of Time* 3rd edn (Berlin: Springer)

[31] Isham C 1991 Conceptual and geometrical problems in quantum gravity *Recent Aspects of Quantum Fields (Springer Lecture Notes in Physics* vol 396*)* ed H Mitter and H Gausterer (Berlin: Springer)

[32] Banks T 1985 T C P, quantum gravity, the cosmological constant and all that... *Nucl. Phys.* B **249** 332

[33] Hartle J B 1993 Space-time quantum mechanics and the quantum mechanics of spacetime *Preprint* gr-qc/9304006

[34] Kiefer C 2004 *Quantum Gravity* (Oxford: Clarendon)

[35] Rovelli C 1991 What is observable in classical and quantum gravity? *Class. Quantum Grav.* **8** 297–316

[36] Rovelli C 2002 GPS observables in general relativity *Phys. Rev.* D **65** 044017 (*Preprint* gr-qc/0110003)

[37] Rovelli C 2002 Partial observables *Phys. Rev.* D **65** 124013 (*Preprint* gr-qc/0110035)

[38] Bojowald M 2001 Absence of singularity in loop quantum cosmology *Phys. Rev. Lett.* **86** 5227–30 (*Preprint* gr-qc/0102069)

[39] Ashtekar A, Bojowald M and Lewandowski J 2003 Mathematical structure of loop quantum cosmology *Adv. Theor. Math. Phys.* **7** 233–68 (*Preprint* gr-qc/0304074)

[40] Russo J G and Townsend P K 2005 Cosmology as relativistic particle mechanics: from big crunch to big bang *Class. Quantum Grav.* **22** 737–52 (*Preprint* hep-th/0408220)

[41] Brunnemann J and Thiemann T 2005 On (cosmological) singularity avoidance in loop quantum gravity *Preprint* gr-qc/0505032

[42] Coule D H 2003 Contrasting quantum cosmologies *Preprint* gr-qc/0312045

[43] Bojowald M 2002 Inflation from quantum geometry *Phys. Rev. Lett.* **89** 261301 (*Preprint* gr-qc/0206054)

[44] Vandersloot K 2005 On the Hamiltonian constraint of loop quantum cosmology *Phys. Rev.* D **71** 103506 (*Preprint* gr-qc/0502082)

[45] Bekenstein J D 1973 Black holes and entropy *Phys. Rev.* D **7** 2333–46

[46] Hawking S W 1975 Particle creation by black holes *Commun. Math. Phys.* **43** 199–220

[47] Ashtekar A, Baez J, Corichi A and Krasnov K 1998 Quantum geometry and black hole entropy *Phys. Rev. Lett.* **80** 904–7 (*Preprint* gr-qc/9710007)

[48] Ashtekar A, Baez J C and Krasnov K 2000 Quantum geometry of isolated horizons and black hole entropy *Adv. Theor. Math. Phys.* **4** 1–94 (*Preprint* gr-qc/0005126)

[49] Strominger A and Vafa C 1996 Microscopic origin of the Bekenstein–Hawking entropy *Phys. Lett.* B **379** 99–104 (*Preprint* hep-th/9601029)

[50] de Wit B 2005 Introduction to black hole entropy and supersymmetry *Preprint* hep-th/0503211

[51] Ashtekar A and Krishnan B 2004 Isolated and dynamical horizons and their applications *Living Rev. Rel.* **7** 10 http://www.livingreviews.org/lrr-2004-10 (*Preprint* gr-qc/0407042)

[52] Reisenberger M P 1994 World sheet formulations of gauge theories and gravity *Preprint* gr-qc/9412035

[53] Reisenberger M P and Rovelli C 1997 'Sum over surfaces' form of loop quantum gravity *Phys. Rev.* D **56** 3490–508 (*Preprint* gr-qc/9612035)

[54] Baez J C 1998 Spin foam models *Class. Quantum Grav.* **15** 1827–58 (*Preprint* gr-qc/9709052)

[55] Barrett J W and Crane L 1998 Relativistic spin networks and quantum gravity *J. Math. Phys.* **39** 3296–302 (*Preprint* gr-qc/9709028)

[56] Rovelli C 1999 The projector on physical states in loop quantum gravity *Phys. Rev.* D **59** 104015 (*Preprint* gr-qc/9806121)

[57] Arnowitt R, Deser S and Misner C W 1962 The dynamics of general relativity *Gravitation: An Introduction to Current Research* ed L Witten (New York: Wiley) pp 227–65 (*Preprint* gr-qc/0405109)

[58] Misner C, Thorne K and Wheeler J 1973 *Gravitation* (San Francisco: Freeman)

[59] Dirac P A M 1975 *General Theory of Relativity* (Princeton, NJ: Princeton University Press)

[60] Henneaux M and Teitelboim C 1992 *Quantization of Gauge Systems* (Princeton, NJ: Princeton University Press)

[61] Nicolai H and Matschull H J 1993 Aspects of canonical gravity and supergravity *J. Geom. Phys.* **11** 15–62

[62] Geroch R 1970 Domain of dependence *J. Math. Phys.* **11** 437

[63] DeWitt B S 1967 Quantum theory of gravity: I. The canonical theory *Phys. Rev.* **160** 1113–48

[64] DeWitt B S 1967 Quantum theory of gravity: II. The manifestly covariant theory *Phys. Rev.* **162** 1195–239

[65] DeWitt B S 1967 Quantum theory of gravity: III. Applications of the covariant theory *Phys. Rev.* **162** 1239–56

[66] Wheeler J A 1964 Geometrodynamics and the issue of the final state *Relativity, Groups and Topology* ed C DeWitt and J Wheeler (London: Blackie)

[67] Ashtekar A 1987 New Hamiltonian formulation of general relativity *Phys. Rev.* D **36** 1587–602

[68] Holst S 1996 Barbero's Hamiltonian derived from a generalized Hilbert–Palatini action *Phys. Rev.* D **53** 5966–9 (*Preprint* gr-qc/9511026)

[69] Samuel J 2001 Comment on Holst's Lagrangian formulation *Phys. Rev.* D **63** 068501

[70] Barbero J F 1995 Real Ashtekar variables for Lorentzian signature space times *Phys. Rev.* D **51** 5507–10 (*Preprint* gr-qc/9410014)

[71] Immirzi G 1997 Real and complex connections for canonical gravity *Class. Quantum Grav.* **14** L177–81 (*Preprint* gr-qc/9612030)

[72] Gambini R, Obregon O and Pullin J 1999 Yang–Mills analogues of the Immirzi ambiguity *Phys. Rev.* D **59** 047505 (*Preprint* gr-qc/9801055)

[73] Bengtsson I 1989 Yang–Mills theory and general relativity in three and four dimensions *Phys. Lett.* B **220** 51

[74] Thiemann T 1998 Quantum spin dynamics (QSD) *Class. Quantum Grav.* **15** 839–73 (*Preprint* gr-qc/9606089)

[75] Nicolai H 1991 The canonical structure of maximally extended supergravity in three dimensions *Nucl. Phys.* B **353** 493–518

[76] Matschull H J and Nicolai H 1994 Canonical quantum supergravity in three dimensions *Nucl. Phys.* B **411** 609–46 (*Preprint* gr-qc/9306018)

[77] Samuel J 2000 Is Barbero's Hamiltonian formulation a gauge theory of Lorentzian gravity? *Class. Quantum Grav.* **17** L141–8 (*Preprint* gr-qc/0005095)

[78] Kodama H 1990 Holomorphic wave function of the universe *Phys. Rev.* D **42** 2548–65

[79] Jackiw R 1985 Topological investigations in quantized gauge theories *Current Algebra and Anomalies* ed S B Treiman (Singapore: World Scientific) p 258, exercise 3.7

[80] Witten E 2003 A note on the Chern–Simons and Kodama wavefunctions *Preprint* gr-qc/0306083

[81] Freidel L and Smolin L 2004 The linearization of the Kodama state *Class. Quantum Grav.* **21** 3831–44 (*Preprint* hep-th/0310224)

[82] Rovelli C and Smolin L 1990 Loop space representation of quantum general relativity *Nucl. Phys.* B **331** 80

[83] Simon B 1974 *The P(φ)₂ Euclidean (Quantum) Field Theory (Princeton Series in Physics)* (Princeton, NJ: Princeton University Press)

[84] 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 (*Preprint* gr-qc/9806041)

[85] Stone M H 1930 Linear transformations in Hilbert space: III. Operational methods and group theory *Proc. Natl. Acad. Sci.* **16** 172–5

[86] von Neumann J 1931 Die Eindeutigkeit der Schrödingerschen Operatoren *Math. Ann.* **104** 570

[87] Rosenberg J 2004 A selective history of the Stone–von Neumann theorem *Operator Algebras, Quantization, and Noncommutative Geometry: A Centennial Celebration Honoring John von Neumann and Marshall H Stone* ed R S Doran and R V Kadison *(Contemporary Mathematics* vol 365*)* (Providence, RI: American Mathematical Society)

[88] Perez A 2003 Spin foam models for quantum gravity *Class. Quantum Grav.* **20** R43 (*Preprint* gr-qc/0301113)

[89] Sahlmann H 2002 Some comments on the representation theory of the algebra underlying loop quantum gravity *Preprint* gr-qc/0207111

[90] Sahlmann H 2002 When do measures on the space of connections support the triad operators of loop quantum gravity? *Preprint* gr-qc/0207112

[91] Sahlmann H and Thiemann T 2003 On the superselection theory of the Weyl algebra for diffeomorphism invariant quantum gauge theories *Preprint* gr-qc/0302090

[92] Okolow A and Lewandowski J 2003 Diffeomorphism covariant representations of the holonomy-flux ∗-algebra *Class. Quantum Grav.* **20** 3543–68 (*Preprint* gr-qc/0302059)

[93] Sahlmann H and Thiemann T 2003 Irreducibility of the Ashtekar–Isham–Lewandowski representation *Preprint* gr-qc/0303074

[94] Okolow A and Lewandowski J 2005 Automorphism covariant representations of the holonomy-flux ∗-algebra *Class. Quantum Grav.* **22** 657–80 (*Preprint* gr-qc/0405119)

[95] Thirring W and Narnhofer H 1992 Covariant QED without indefinite metric *Rev. Math. Phys.* **SI1** 197–211

[96] Migdal A A 1983 Loop equations and $1/N$ expansion *Phys. Rev.* **102** 199–290

[97] Polyakov A M 1987 Gauge fields and strings *Contemporary Concepts in Physics* vol 3 (New York: Harwood Academic)

[98] Montvay I and Münster G 1994 *Quantum Fields on A Lattice* (Cambridge: Cambridge University Press)

[99] Rothe H J 1992 *Lattice Gauge Theories: An Introduction* (Singapore: World Scientific)

[100] Creutz M 1983 *Quarks, Gluons and Lattices* (Cambridge: Cambridge University Press)

[101] Sahlmann H and Thiemann T 2002 Towards the QFT on curved spacetime limit of QGR: I. A general scheme *Preprint* gr-qc/0207030

[102] Sahlmann H and Thiemann T 2002 Towards the QFT on curved spacetime limit of QGR: II. A concrete implementation *Preprint* gr-qc/0207031

[103] Brown J D and Kuchar K V 1995 Dust as a standard of space and time in canonical quantum gravity *Phys. Rev.* D **51** 5600–29 (*Preprint* gr-qc/9409001)

[104] Kuchar K V 1993 Matter time in canonical quantum gravity *Directions in General Relativity: An International Symposium in Honor of the 60th Birthdays of Dieter Brill and Charles Misner* pp 27–9

[105] Dittrich B 2004 Partial and complete observables for Hamiltonian constrained systems *Preprint* gr-qc/0411013

[106] Thiemann T 2004 Reduced phase space quantization and Dirac observables *Preprint* gr-qc/0411031

[107] Kuchar K V 1991 Time and interpretations of quantum gravity *Proc. 4th Canadian Conf. on General Relativity and Relativistic Astrophysics* ed G Kunstatter, D Vincent and J Williams (Singapore: World Scientific) pp 211–314

[108] Ashtekar A and Lewandowski J 1997 Quantum theory of geometry: I. Area operators *Class. Quantum Grav.* **14** A55–82 (*Preprint* gr-qc/9602046)

[109] Ashtekar A and Lewandowski J 1998 Quantum theory of geometry: II. Volume operators *Adv. Theor. Math. Phys.* **1** 388 (*Preprint* gr-qc/9711031)

[110] Thiemann T 1998 A length operator for canonical quantum gravity *J. Math. Phys.* **39** 3372–92 (*Preprint* gr-qc/9606092)

[111] Domagala M and Lewandowski J 2004 Black hole entropy from quantum geometry *Class. Quantum Grav.* **21** 5233–44 (*Preprint* gr-qc/0407051)

[112] Meissner K A 2004 Black hole entropy in loop quantum gravity *Class. Quantum Grav.* **21** 5245–52 (*Preprint* gr-qc/0407052)

[113] Ghosh A and Mitra P 2005 An improved lower bound on black hole entropy in the quantum geometry approach *Phys. Lett.* B **616** 114–7 (*Preprint* gr-qc/0411035)

[114] Khriplovich I B 2004 Holographic bound and spectrum of quantized black hole *Preprint* gr-qc/0411109

[115] Thiemann T 1998 Closed formula for the matrix elements of the volume operator in canonical quantum gravity *J. Math. Phys.* **39** 3347–71 (*Preprint* gr-qc/9606091)

[116] De Pietri R and Rovelli C 1996 Geometry eigenvalues and scalar product from recoupling theory in loop quantum gravity *Phys. Rev.* D **54** 2664–90 (*Preprint* gr-qc/9602023)

[117] Brunnemann J and Thiemann T 2004 Simplification of the spectral analysis of the volume operator in loop quantum gravity *Preprint* gr-qc/0405060

[118] Glimm J and Jaffe A 1987 *Quantum Physics—A Functional Integral Point of View* (New York: Springer)

[119] Baez J C and Krasnov K V 1998 Quantization of diffeomorphism-invariant theories with fermions *J. Math. Phys.* **39** 1251–71 (*Preprint* hep-th/9703112)

[120] Thiemann T 1998 QSD V: quantum gravity as the natural regulator of matter quantum field theories *Class. Quantum Grav.* **15** 1281–314 (*Preprint* gr-qc/9705019)

[121] Varadarajan M 2000 Fock representations from U(1) holonomy algebras *Phys. Rev.* D **61** 104001 (*Preprint* gr-qc/0001050)

[122] Varadarajan M 2001 Photons from quantized electric flux representations *Phys. Rev.* D **64** 104003 (*Preprint* gr-qc/0104051)

[123] Ashtekar A and Lewandowski J 2001 Relation between polymer and Fock excitations *Class. Quantum Grav.* **18** L117–28 (*Preprint* gr-qc/0107043)

[124] Ashtekar A, Lewandowski J and Sahlmann H 2003 Polymer and Fock representations for a scalar field *Class. Quantum Grav.* **20** L11–1 (*Preprint* gr-qc/0211012)

[125] Thiemann T 1998 Kinematical Hilbert spaces for fermionic and Higgs quantum field theories *Class. Quantum Grav.* **15** 1487–512 (*Preprint* gr-qc/9705021)

[126] Ashtekar A, Lewandowski J, Marolf D, Mourao J and Thiemann T 1995 Quantization of diffeomorphism invariant theories of connections with local degrees of freedom *J. Math. Phys.* **36** 6456–93 (*Preprint* gr-qc/9504018)

[127] Peldan P 1992 Ashtekar's variables for arbitrary gauge group *Phys. Rev.* D **46** 2279–82 (*Preprint* hep-th/9204069)

[128] Chakraborty S and Peldan P 1994 Towards a unification of gravity and Yang–Mills theory *Phys. Rev. Lett.* **73** 1195–98 (*Preprint* gr-qc/9401028)

[129] Melosch S and Nicolai H 1998 New canonical variables for $D = 11$ supergravity *Phys. Lett.* B **416** 91–100 (*Preprint* hep-th/9709227)

[130] Hajicek P and Kuchar K V 1990 Constraint quantization of parametrized relativistic gauge systems in curved spacetimes *Phys. Rev.* D **41** 1091–104

[131] Witten E 1988 (2 + 1)-dimensional gravity as an exactly soluble system *Nucl. Phys.* B **311** 46

[132] 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

[133] de Wit B, Matschull H J and Nicolai H 1993 Physical states in $d = 3$, $N = 2$ supergravity *Phys. Lett.* B **318** 115–21 (*Preprint* gr-qc/9309006)

[134] Zapata J A 1997 A combinatorial approach to diffeomorphism invariant quantum gauge theories *J. Math. Phys.* **38** 5663–81 (*Preprint* gr-qc/9703037)

[135] Fairbairn W and Rovelli C 2004 Separable Hilbert space in loop quantum gravity *J. Math. Phys.* **45** 2802–14 (*Preprint* gr-qc/0403047)

[136] Edwards R 1965 *Functional Analysis: Theory and Applications* (New York: Holt, Rinehart and Winston)

[137] Smolin L 1996 The classical limit and the form of the Hamiltonian constraint in non-perturbative quantum general relativity *Preprint* gr-qc/9609034

[138] Neville D E 1999 Long range correlations in quantum gravity *Phys. Rev.* D **59** 044032 (*Preprint* gr-qc/9803066)

[139] Loll R 1998 On the diffeomorphism-commutators of lattice quantum gravity *Class. Quantum Grav.* **15** 799–809 (*Preprint* gr-qc/9708025)

[140] Thiemann T 1998 QSD III: quantum constraint algebra and physical scalar product in quantum general relativity *Class. Quantum Grav.* **15** 1207–47 (*Preprint* gr-qc/9705017)

[141] Mukhanov V and Wipf A 1995 On the symmetries of Hamiltonian systems *Int. J. Mod. Phys.* A **10** 579–610 (*Preprint* hep-th/9401083)

[142] Kouletsis I and Kuchar K V 2002 Diffeomorphisms as symplectomorphisms in history phase space: bosonic string model *Phys. Rev.* D **65** 125026 (*Preprint* gr-qc/0108022)

[143] Green M B, Schwarz J H and Witten E 1987 *Superstring Theory* (Cambridge: Cambridge University Press)

[144] Polchinski J 1998 *String Theory* (Cambridge: Cambridge University Press)

[145] Starodubtsev A 2002 String theory in a vertex operator representation: a simple model for testing loop quantum gravity *Preprint* gr-qc/0201089
[146] Thiemann T 2004 The LQG string: loop quantum gravity quantization of string theory: I. Flat target space *Preprint* hep-th/0401172
[147] Schreiber U 2004 DDF and Pohlmeyer invariants of (super)string *J. High Energy Phys.* JHEP05(2004)027 (*Preprint* hep-th/0403260)
[148] Helling R C and Policastro G 2004 String quantization: Fock versus LQG representations *Preprint* hep-th/0409182
[149] Louis-Martinez D, Gegenberg J and Kunstatter G 1994 Exact Dirac quantization of all 2-d dilaton gravity theories *Phys. Lett.* B **321** 193–8 (*Preprint* gr-qc/9309018)
[150] Benedict E, Jackiw R and Lee H J 1996 Functional Schrödinger and BRST quantization of (1 + 1)-dimensional gravity *Phys. Rev.* D **54** 6213–25 (*Preprint* hep-th/9607062)
[151] Cangemi D, Jackiw R and Zwiebach B 1996 Physical states in matter coupled dilaton gravity *Ann. Phys.* **245** 408–44 (*Preprint* hep-th/9505161)
[152] Grumiller D, Kummer W and Vassilevich D V 2002 Dilaton gravity in two dimensions *Phys. Rep.* **369** 327–430 (*Preprint* hep-th/0204253)
[153] Markopoulou F and Smolin L 1998 Quantum geometry with intrinsic local causality *Phys. Rev.* D **58** 084032 (*Preprint* gr-qc/9712067)
[154] Jonghe F De, Siebelink R, Troost W, Vandoren S, van Nieuwenhuizen P and Van Proeyen A 1992 The regularized BRST Jacobian of pure Yang–Mills theory *Phys. Lett.* B **289** 354–60 (*Preprint* hep-th/9206097)
[155] Musser G 2003 The future of string theory—a conversation with Brian Greene *Sci. Am.* November 48–53
[156] Overbye D 2004 String theory, at 20, explains it all (or not), New York Times 7 December
[157] Dijkgraaf R, Gukov S, Neitzke A and Vafa C 2004 Topological M-theory as unification of form theories of gravity *Preprint* hep-th/0411073