

# The status of diffeomorphism superselection in Euclidean 2 + 1 gravity

Donald Marolf

*Department of Physics, Syracuse University, Syracuse, New York 13244*

José Mourão

*Departamento de Física, Instituto Superior, Técnico, Av. Rovisco Pais, 1096 Lisboa Codex, Portugal*

Thomas Thiemann<sup>a)</sup>

*Physics Department, Harvard University, Cambridge, Massachusetts 02138*

(Received 3 March 1997; accepted for publication 14 April 1997)

This work addresses a specific technical question of relevance to canonical quantization of gravity using the so-called new variables and loop-based techniques of Ashtekar, Rovelli, and Smolin. In particular, certain “superselection laws” that arise in current applications of these techniques to solving the diffeomorphism constraint are considered. Their status is elucidated by studying an analogous system: 2 + 1 Euclidean gravity. For that system, these superselection laws are shown to be spurious. This, however, is only a technical difficulty. The usual quantum theory may still be obtained from a loop representation and the technique known as “Refined Algebraic Quantization.” © 1997 American Institute of Physics.  
[S0022-2488(97)01409-6]

## I. INTRODUCTION

A recent advance in canonical quantization techniques was the introduction<sup>1</sup> of refined algebraic quantization (RAQ) (and related techniques<sup>2,3</sup>) for solving quantum constraints and for inducing physical inner products. As shown in Ref. 1 the use of such techniques often results in “superselection rules.” While such superselection rules can correspond to important properties of the physical system<sup>1,4</sup> which are present even at the classical level, when RAQ is used to solve the diffeomorphism constraints of a quantum theory of connections as in Ref. 1, the interpretation of the superselection rules is less clear.

In particular, when 2 + 1 gravity is expressed as a theory of connections,<sup>5-7</sup> the simplest observables appear to violate these rules. This is because, in a loop representation, these rules select between states associated with different topological types of graphs or loops, while important observables in the 2 + 1 theory are traces of holonomies of connections around noncontractable curves, which mix the above states.

In a loop representation, such operators change not only the topology but also the homotopy type of a loop-state. The goal of this paper is to determine the status of these superselection rules in Euclidean 2 + 1 gravity and to determine whether their presence prevents the quantization scheme described in Ref. 1 from succeeding. This will help to clarify the standing of such methods in the loop-based approach to 3 + 1 gravity.

We will proceed in two stages. It will first be shown that methods based on a loop representation and the “refined algebraic quantization scheme” (RAQ) of Ref. 1 do yield the usual results<sup>5,8,9</sup> for Euclidean 2 + 1 gravity when they are properly applied. In this case, the most straightforward treatment differs from the particular approach suggested in Ref. 1. However, we also show that the solution may be recast in the form advocated in Ref. 1 in which the diffeomor-

---

<sup>a)</sup>Present address: Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, Schlaatzweg 1, 14473 Potsdam, Germany, Electronic mail: thiemann@aei-potsdam.mpg.de

phism constraints are solved first and the Hamiltonian constraints are then solved as a second stage. Regarding the “diffeomorphism superselection rules” mentioned above, we will see that they disappear in the final solution of this system. Some concluding comments about expectations for the 3 + 1 theory are given in Sec. IV and we draw on supporting material from the Appendix. This discussion suggests that the intermediate presence of the superselection rules in the 2 + 1 theory is due to the singular nature of our description of the theory, but that a similar singularity may be present in the loop approach to the 3 + 1 case.<sup>1</sup>

This work will make use of a loop representation along the lines of Ref. 1 as well as the refined algebraic techniques discussed there, in Ref. 4, and elsewhere. As a result, what follows is best considered a technical addendum to Ref. 1 and the review of that material will be kept to a minimum. We use the same structures and definitions as in Ref. 1, except as where noted below. We will, however, briefly discuss the formulation of Euclidean 2 + 1 gravity as a canonical theory of connections since that was not discussed in Ref. 1.

As described by Witten,<sup>5</sup> Euclidean 2 + 1 vacuum gravity may be considered as a theory of cotriads  $\bar{e}_{aI}$  and SU(2) valued connections  $\bar{A}_a^I$ . Here  $a, b$  are space–time indices on a three manifold  $M$ . The system is governed by the action

$$S(\bar{e}_{aI}, \bar{A}_a^I) = \frac{1}{2} \int_M d^3x \bar{\epsilon}^{abc} \bar{e}_{aI} \bar{F}_{bc}^I, \tag{I.1}$$

where  $\bar{\epsilon}^{abc}$  is the Levi-Civita density on  $M$  and  $\bar{F}_{bc}^I$  is the curvature of the connection  $\bar{A}_a^I$ . This is just the 2 + 1 Einstein–Hilbert action written in terms of the triad and spin connection. For later convenience we have taken the action to differ from that of Ref. 5 by a factor of 1/2.

If we now take  $M$  to be of the form  $\mathbf{R} \times \Sigma$  (for a closed orientable two-manifold  $\Sigma$ ), we may make a 2 + 1 decomposition of the above action. The result is a system where the Hamiltonian is simply a sum of constraints. We shall take  $i, j, k$  to be abstract indices associated with the manifold  $\Sigma$ . The canonical variables are a connection  $A_i^I$  which is the pull back of the connection  $\bar{A}_a^I$  to  $\Sigma$  and a vector density  $\tilde{E}_I^i = \tilde{\epsilon}^{ij} e_{jI}$ , where  $\epsilon^{ij}$  is the Levi-Civita density on  $\Sigma$  and  $e_{jI}$  is the pull back of  $e_{aI}$  to  $\Sigma$ . These satisfy the canonical commutation relations

$$\{A_i^I(x), \tilde{E}_I^j(x')\} = \delta_i^j \delta_j^I \delta^2(x, x') \tag{I.2}$$

and, in terms of  $A_i^I, \tilde{E}_I^j$ , the constraints are

$$F_{ij}^I = 0, \quad D_j \tilde{E}_I^j = 0, \tag{I.3}$$

where  $F_{ij}^I$  and  $D_j$  are the curvature and covariant derivative associated with  $A_i^I$ , respectively.

The second constraint is known as the Gauss constraint and generates SU(2) gauge transformations. The first constraint is more complicated, but clearly generates transformations that do not change the connection. The reader will, at this point, notice the distinct lack of a constraint that generates diffeomorphisms. Such a constraint would have the form  $\tilde{E}_I^i F_{ij}^I = 0$ . Although it is not one of the constraints (I.3), this function clearly vanishes on the constraint surface; the result is that any function invariant under the transformations generated by the  $F = 0$  constraint also becomes invariant under diffeomorphisms once it has been restricted to the constraint surface. In this sense then, the Witten constraints are in fact weakly equivalent (for nondegenerate triads) to the set of constraints<sup>10</sup>

$$D_i \tilde{E}_I^i = 0, \quad \tilde{E}_I^i F_{ij}^I = 0, \quad \epsilon_K^{IJ} \tilde{E}_I^i \tilde{E}_J^j F_{ij}^K = 0, \tag{I.4}$$

but (I.3) and (I.4) are not strongly equivalent.

We are therefore left with the question of which set of constraints to use here. On the one hand, the well-understood description of 2 + 1 gravity refers to the constraints (I.3). On the other,

we are most interested in gaining insight into 3+1 gravity, for which only a description of the form (I.4) is available. Furthermore, the question of the “diffeomorphism superselection sectors” which we wish to study does not arise unless there is in fact a diffeomorphism constraint. However, the densitized “Hamiltonian constraint”  $EEF=0$  of (I.4) is as difficult to define here as in the 3+1 case.

One approach might be to apply techniques such as those introduced by Thiemann for the 3+1 Hamiltonian constraints in Ref. 11. However, because of conceptual and technical complications involved, we leave direct investigation of the constraints (I.4) for future work<sup>12</sup> and content ourselves here with following a hybrid approach. After briefly reviewing the refined algebraic techniques in Sec. II, we define our system using the Witten constraints (I.3) and show that the combination of a loop representation with refined algebraic techniques generates the usual physical Hilbert space in a straightforward manner. In Sec. III, we show that the physical states generated in this way *are* annihilated by the diffeomorphism constraint  $\tilde{E}_i^l F_{ij}^l=0$  in the sense described in Ref. 1 and that our physical Hilbert space could have been constructed by following the procedure outlined in Ref. 1, in which the diffeomorphism constraint  $\tilde{E}_i^l F_{ij}^l$  is solved first (through RAQ) and the “remaining parts” of the constraints (I.3) are solved later by RAQ-like techniques. Section IV discusses the implications for the 3+1 theory, drawing on the Appendix for support.

## II. QUANTIZATION

We now proceed to quantize the system described in Sec. I and to impose the constraints (I.3) using a loop representation and the techniques of refined algebraic quantization. That is to say, we will follow Ref. 1 in considering an auxiliary kinematical Hilbert space  $H_{\text{aux}}=L^2(\overline{\mathcal{AS}}, d\mu_0)$  where  $\overline{\mathcal{AS}}$  is the Ashtekar–Isham “quantum configuration space of gauge equivalent connections,”<sup>13</sup> appropriate to the connections discussed above. This space contains not only connections but suitably generalized “distributional” connections as might be expected to be required in the configuration space of a quantum field theory. Note that  $d\mu_0$  is the corresponding Ashtekar–Lewandowski measure.<sup>14</sup> States in this space are gauge invariant, so there is no need to impose the Gauss constraints, they are considered to be identically satisfied on this space. [If one wishes, one may<sup>1</sup> begin with a larger Hilbert space  $L_2(\overline{\mathcal{A}}, d\hat{\mu}_0)$  whose states are not gauge invariant, introduce the Gauss constraints as operators on this space, and solve them by RAQ to arrive at  $H_{\text{aux}}=L^2(\overline{\mathcal{AS}}, d\mu_0)$  as above.]

We must, however, define and solve the  $F=0$  constraints. This involves a slight complication as the generalized connections of  $\overline{\mathcal{AS}}$  are not in general differentiable. Thus the curvature  $F$  is strictly speaking not well-defined on this space. What *is* well-defined though is the holonomy of a generalized connection, this is in fact the very definition of  $\overline{\mathcal{AS}}$ . We therefore proceed as follows: The statement  $F=0$  for a smooth connection  $A$  is equivalent to the statement that the holonomy  $h_\alpha(A)$  of  $A$  around each contractable loop  $\alpha$  in  $\Sigma$  is trivial, that is, gives just the unit element of  $SU(2)$ . A manifestly gauge invariant formulation of the constraints is thus

$$C'_\alpha := 2 - T_\alpha(A) = 0, \quad (\text{II.1})$$

where  $T_\alpha(A) := \text{tr}(h_\alpha(A))$ , for all contractable loops. The virtue of writing the  $F=0$  constraint in the form (II.1) is that we can extend it to  $\overline{\mathcal{AS}}$ . The disadvantage is that this constraint classically does not generate gauge transformations on the constraint surface as  $T_\alpha - 2$  is quadratic in  $F$ . We will actually use a constraint of the form

$$C_\alpha = |2 - T_\alpha(A)|^{3/2} \quad (\text{II.2})$$

for all contractable loops  $\alpha$ . While, classically, this is even worse than (II.1), we shall see that a certain ‘‘cancellation of singularities’’ occurs, and that this is in fact a preferred form of the constraints.

We now wish to solve these constraints using the refined algebraic quantization procedure. Recall that this involves introducing a preferred dense subspace  $\Phi$  of the Hilbert space  $H_{\text{aux}}$  and defining an (antilinear) map  $\eta$  from  $\Phi$  to its dual  $\Phi'$  which ‘‘projects a state onto the constraint surface’’ in the sense that the image  $\eta\phi$  of any  $\phi \in \Phi$  is a solution of the constraints:  $C(\eta\phi) = 0$ , where  $C$  denotes the constraints. Note that the action of  $C$  on  $\Phi'$  is defined to be the dual of its action on  $\Phi$ . The details are given in Ref. 1, but we remind the reader that the solutions  $\eta\phi$  in the image of  $\eta$  are given a Hilbert space structure through

$$\langle \eta\alpha | \eta\beta \rangle \equiv (\eta\beta)[\alpha]. \quad (\text{II.3})$$

The map  $\eta$  must be real and positive in the sense that, for all  $\phi_1, \phi_2 \in \Phi$ ,

$$(\eta\phi_1)[\phi_2] = ((\eta\phi_2)[\phi_1])^* \quad \text{and} \quad (\eta\phi_1)[\phi_1] \geq 0 \quad (\text{II.4})$$

and  $\eta$  must commute with every strong observable  $A$ . That is, for any operator  $A$  which commutes with all gauge transformations, we must have

$$(\eta\phi_1)[A\phi_2] = ((\eta A^\dagger \phi_1))[\phi_2]. \quad (\text{II.5})$$

In this case (II.3) defines an inner product which may be used to complete the set of states  $\eta\phi$  in the image of  $\eta$  to a ‘‘gauge invariant’’ Hilbert space  $H_{\text{inv}}$ . Moreover, this inner product has the property that any strong observable  $A$  on  $H_{\text{aux}}$  induces on operator  $A_{\text{inv}}$  on the physical Hilbert space satisfying the same reality conditions; i.e.,  $A_{\text{inv}}^\dagger = (A^\dagger)_{\text{inv}}$ . The operator  $A_{\text{inv}}$  is defined by

$$A_{\text{inv}}(\eta\phi) = \eta(A\phi). \quad (\text{II.6})$$

Note that the invariant Hilbert space was referred to as the ‘‘physical’’ Hilbert space in Refs. 1 and 4. The terminology we use here is more appropriate for the current setting, in which we allow the possibility that this procedure be applied more than once, solving only some of the constraints at each step.

A nice idea for constructing the map  $\eta$  is through ‘‘group averaging.’’<sup>1,3,4</sup> Under appropriate conditions, an expression of the form

$$(\eta\phi_1)[\phi_2] = \int_G dg \langle \phi_1 | U(g) | \phi_2 \rangle, \quad (\text{II.7})$$

with  $dg$  the Haar measure on the gauge group  $G$ , gives a well-defined map  $\eta$  with the required properties. This heuristic idea is often quite useful in applying RAQ, although it will not be of direct use for our case.

The constraint  $F=0$  is a pure configuration constraint: It does not involve the canonical momenta. This situation is reminiscent of solving the relativistic free particle constraint  $p^2 + m^2 = 0$  in the momentum representation. Let us recall how this works as it will clarify our case.

In the relativistic particle case we choose  $H_{\text{aux}} := L_2(\mathbb{R}^2, d^4p)$  and  $\Phi := C_0^\infty(\mathbb{R}^4)$ , say, the smooth test functions of compact support. The constraint  $C = p^2 + m^2 = 0$  is easy to solve: each solution can be written in the form  $\psi_f(p) = \delta(C)f(\mathbf{p})$  where  $f \in \Phi$ . The point is that  $\psi_f \in \Phi'$  is not an element of  $H_{\text{aux}}$ . But why can we claim that the constraint was solved by group averaging? This is because  $\hat{C} := \hat{p}^2 + m^2$  is an essentially self-adjoint operator on  $H_{\text{aux}}$  with core  $\Phi$  whose unique self-adjoint extension we may exponentiate to obtain a one-parameter unitary group  $\hat{U}(a) := \exp(ia\hat{C})$  with  $a \in \mathbb{R}$ . The Haar measure on  $\mathbb{R}$  is the Lebesgue measure and so for each  $\phi \in \Phi$  we obtain the following group average map

$$(\eta f)(\phi) := \int_{\mathbb{R}} \frac{da}{2\pi} \langle f, \hat{U}(a)\phi \rangle = \psi_f(\phi) \tag{II.8}$$

in analogy with (II.7).

Why does this lead to the desired result? The answer is that one way of looking at the group average procedure is that one wishes to solve the exponentiated constraint  $\hat{U}(a) = 1 \forall a$  and the average over all the  $\hat{U}(a)$  has to be done in such a way that we get the  $\delta(\hat{C})$  back, or, in other words, such that the translation of the parameter  $a$  in  $\hat{U}(a)\hat{U}(b) = \hat{U}(a+b)$  is irrelevant because we are using a translation invariant measure (the Haar measure) on the parameter space.

Another feature of the relativistic particle shared by our model is that the solutions to  $C=0$  are not unique. For the free particle they are twofold,  $p^0 = \pm \sqrt{\mathbf{p}^2 + m^2} =: \pm \omega(\mathbf{p})$  which we may encode in the following way  $C = C_+ C_-$  where  $C_{\pm} = p^0 \pm \sqrt{\mathbf{p}^2 + m^2}$ . Also, in the sense of distributions  $\delta(C) = (1/2\omega)[\delta(C_+) + \delta(C_-)] =: \int_{\mathcal{M}} d\nu(\omega) \delta(p^0, \omega)$  where  $\mathcal{M} = \{\pm \omega(\mathbf{p})\}$  is the solution space and  $\nu$  is proportional to a counting measure on  $\mathcal{M}$ . We will encounter precisely the same structure in our model. This concludes the discussion of the relativistic particle.

Let us now turn to our case. The solutions to  $F=0$  are the flat connections, and, since we are interested only in gauge-invariant information, we have the space  $\mathcal{M}$ , the moduli space of flat connections modulo gauge transformations as our solution space. Therefore we write the distribution  $\delta(F)$  as

$$\delta(F) = \int_{\mathcal{M}} d\nu(A_0) \delta(A, A_0), \tag{II.9}$$

where  $\nu$  is some (real-valued) measure on  $\mathcal{M}$ . We will derive a preferred measure  $d\nu$  below which agrees with the one give by Witten.<sup>5</sup> This is in direct analogy with writing the  $\delta(p^2 + m^2)$  as a sum of two  $\delta$  distributions, the discrete measure there was replaced by the measure  $\nu$  accounting for the fact that  $\mathcal{M}$  is a manifold.

The next step is nontrivial: We have to write  $\delta(A_0, A)$  as a well-defined distribution on a suitable  $\Phi$ . Let us choose, as in the 3+1 case,  $H_{\text{aux}} := L_2(\mathcal{AS}, d\mu_0)$  where  $\mu_0$  is the Ashtekar–Lewandowski measure and let  $\Phi := \Phi_{\text{cyl}}$  be the cylindrical functions on  $\mathcal{AS}$ . It turns out that  $H_{\text{aux}}$  has an orthonormal basis, the so-called spin-network states  $T_{\gamma, \mathbf{j}, \mathbf{c}}$  (see Ref. 1). Here  $\gamma$  stands for a piecewise analytic closed graph,  $\mathbf{j} = (j_1, \dots, j_E)$  is a labeling of its edges  $e_1, \dots, e_E$  with spin quantum numbers and  $\mathbf{c} = (c_1, \dots, c_V)$  is a labeling of its vertices with certain  $SU(2)$  invariant matrices. The state  $T_{\gamma, \mathbf{j}, \mathbf{c}}$  is built from  $\mathbf{c}$  and  $\otimes_{k=1}^E \pi_{j_k}(h_{e_k}(A))$ , where  $\pi_j$  is the  $j$ th irreducible representation of  $SU(2)$ , by contraction of all group indices in such a way that it is gauge invariant. We may use such states to represent  $\delta(A, A_0)$  as

$$\delta(A_0, A) = \sum_{\gamma, \mathbf{j}, \mathbf{c}} \overline{T_{\gamma, \mathbf{j}, \mathbf{c}}(A)} T_{\gamma, \mathbf{j}, \mathbf{c}}(A_0) \tag{II.10}$$

since, by the orthonormality of spin networks, this satisfies  $\int d\mu_0(A) \phi(A) = \phi(A_0)$  for all  $\phi \in \Phi_{\text{cyl}}$ . The associated rigging map  $\eta_F: \Phi_{\text{cyl}} \rightarrow \Phi'_{\text{cyl}}$  is given by

$$(\eta_F \psi)(\phi) = \int_{\mathcal{AS}} d\mu_0(A) \overline{\psi(A)} \delta(F) \phi(A). \tag{II.11}$$

Notice that, although the sum (II.10) ranges over a complete set of piecewise analytic graphs (an uncountable set), the result  $\eta_F \psi$  is still a well-defined element of  $\Phi'_{\text{cyl}}$ .

Can the result (II.11) also be obtained by explicitly averaging the constraints (II.1) in analogy with (II.8)? At least at a heuristic level, the answer is in the affirmative. (At a more technical level,

there is a subtlety in that the group generated by the full set of constraints (II.2) does not fit well with the projective structure of  $\mathcal{H}_{\text{aux}}$ .) To see this, notice first that one can write the delta distribution on  $SU(2)$  with respect to the Haar measure  $\mu_H$  as follows

$$\delta(g,1) = \int_{\mathbb{R}} \frac{dt}{2\pi} \exp(it[1 - \text{tr}(g)/2]^{3/2}) \tag{II.12}$$

as the reader can check himself by explicitly writing  $\mu_H$  in terms of local coordinates on  $S^3$ . Note that the power 3/2 is important here as it cancels certain singularities (actually, degeneracies) in the measure. This observation motivates us to construct a cylindrical definition of  $\eta_F$  which we sketch below.

For each graph  $\gamma$  choose a set of generators  $\alpha_1(\gamma), \dots, \alpha_{n(\gamma)}(\gamma)$  of the subgroup of the homotopy group of  $\gamma$  corresponding to contractable loops on  $\Sigma$ . Let now

$$U_\gamma(t_1, \dots, t_{n(\gamma)}) := \prod_{i=1}^{n(\gamma)} U_{\alpha_i(\gamma)}(t_i) \quad \text{where} \quad U_\alpha(t) := \exp(itC_\alpha). \tag{II.13}$$

We are now in the position to define  $\eta_F$  cylindrically: Since each  $\phi, \psi \in \Phi_{\text{cyl}}$  are just finite linear combinations of spin-network states it will be sufficient to define  $\eta_F$  on spin-network states  $\psi = T_{\gamma,j,c}$  through (II.11) for each  $\phi = T_{\gamma',j',c'}$ . It turns out that the proper definition, precisely in analogy to (II.8), is given by

$$(\eta_F T_{\gamma,j,c})(T_{\gamma',j',c'}) := \int_{\mathbb{R}^n} \frac{d^n t}{(2\pi)^n} \langle T_{\gamma,j,c}, U_{\gamma \cup \gamma'}(t_1, \dots, t_n) T_{\gamma',j',c'} \rangle, \tag{II.14}$$

where  $n = n(\gamma \cup \gamma')$ . Namely, using the definition of  $\mu_0$  which assigns to each holonomically independent loop one independent integration variable with respect to the Haar measure on  $SU(2)$  we explicitly compute that (II.14) equals

$$\int d\mu_H(g_1) \cdots d\mu_H(g_m) [\overline{T_{\gamma,j,c} T_{\gamma',j',c'}}](g_1, \dots, g_m) =: \int_{\mathcal{M}} d\nu(A_0) (\overline{T_{\gamma,j,c} T_{\gamma',j',c'}})(A_0), \tag{II.15}$$

where the square brackets on the left-hand side mean that the function is to be evaluated on the trivial holonomy for the contractable loops which thus leaves only an integration over holonomies  $g_1, \dots, g_m$  along loops that generate the homotopy group of  $\Sigma$ . The right-hand side defines the measure  $d\nu$  on  $\mathcal{M}$  and agrees with the measure given by Witten.<sup>5</sup> It is easy to see that (II.14) coincides with (II.11). Note that, even though we must make a choice of generators of  $\pi_1(\gamma \cup \gamma')$  to even write down the integral (II.14), the resulting definition of  $\eta_F$  is independent of this choice. In addition, note that we have seen no sign of the superselection rules that arose in Ref. 1. We shall return to this issue in the next section.

### III. A SOLUTION IN TWO STAGES

Recall that one of the main objectives of the present paper is to solve the theory using the space of diffeomorphism invariant states (from Ref. 1) as an intermediate step. That is, we use a rigging map  $\eta_{\text{Diff}}$  from Ref. 1 to define a Hilbert space  $H_{\text{Diff}}$  of diffeomorphism invariant states and then solve the Hamiltonian constraint using a second topological vector subspace  $\Phi'_{\text{Diff}}$  of  $H_{\text{Diff}}$  and a rigging map  $\eta_{\text{Ham}}: \Phi'_{\text{Diff}} \rightarrow \Phi'_{\text{Diff}}$ . (In the 3 + 1 case there are some additional difficulties with such an approach due to the fact that the corresponding Hamiltonian constraint does not

commute with diffeomorphisms. In the present model this problem does not occur because the set of  $F=0$  constraints is invariant under diffeomorphisms.)

That is, roughly speaking, we wish to write

$$\eta_F = \eta_{\text{Ham}} \circ \eta_{\text{Diff}}. \tag{III.1}$$

In contrast, in Sec. III we solved all of the constraints in one step. As outlined in the Introduction, the diffeomorphism and Hamiltonian constraint are included in the  $F=0$  constraint. What we would like to see now is how the  $F=0$  constraint can be split into two parts, the diffeomorphism part and a remainder. This remainder will, in some sense, define our ‘‘Hamiltonian constraint.’’

There are, however, two immediate problems with (III.1). The first is that each rigging map is antilinear, so that the left-hand side is antilinear while the right is linear. The other is that the left-hand side is a map from  $\Phi_{\text{Cyl}}$  to  $\Phi'_{\text{Cyl}}$ , while the right is a map from  $\Phi_{\text{Cyl}}$  to  $\Phi'_{\text{Diff}}$  (through  $\Phi_{\text{Diff}}$ ). Clearly then, we will need a natural antilinear map  $\sigma: \Phi'_{\text{Diff}} \rightarrow \Phi'_{\text{Cyl}}$ . This map will be an extension of an adjoint map, and will be discussed below in the course of our argument.

We do this as follows. Recall that each diffeomorphism invariant distribution in the space  $\Phi_{\text{Diff}}$  (constructed in Ref. 1) is a linear combination of spin-network states associated with a finite number of graphs. To be more precise, collect the triple  $\gamma, \mathbf{j}, \mathbf{c}$  into a single index  $c$  and let  $T_{[c]} \times(A)$  be the distribution defined by

$$T_{[c]}(A) := \sum_{c' \in [c]} T_{c'}(A), \tag{III.2}$$

where  $[c]$  is the set of labels of the spin-network states that one obtains by acting on  $T_c$  with all possible analytic diffeomorphisms. Our objective is now to write a solution  $\delta(F)T_c$  to the  $F=0$  constraint in terms of  $\eta_{\text{Diff}}$  and a remaining operation  $\eta_{\text{Ham}}$  to be obtained. To that end we write (II.9) explicitly as

$$\delta(F) = \sum_c \overline{T_c(A)} \int_{\mathcal{M}} d\nu(A_0) T_c(A_0) =: \sum_c \overline{T_c(A)} k_c, \tag{III.3}$$

where the sum is over *all* labels  $c$ . What helps us now is that since  $T_c(A_0)$  is diffeomorphism invariant for  $A_0 \in \mathcal{M}$ , it follows that the integrals  $k_c$  do not depend on  $c$  but only on the diffeomorphism equivalence class  $[c]$ . We also note that  $T_c(A)$  is real, so that we may drop the overline.

We will therefore relabel  $k_c$  as  $k_{[c]}$  and so write (III.3) in the form

$$\delta(F) = \sum_{[c]} k_{[c]} T_{[c]}(A), \tag{III.4}$$

which is already a sum of diffeomorphism invariant distributions only.

If we introduce the notation  $T_{[c]}$  for the linear functional on  $\Phi_{\text{Cyl}}$  given by  $T_{[c]}(\phi) = \int_{\mathcal{M}} \overline{T_{[c]}(A)} \phi(A)$ , then we may write

$$\eta_F(1) = \sum_{[c]} k_{[c]} T_{[c]}. \tag{III.5}$$

In order to connect with Ref. 1, recall that, due to the superselection rules,  $\nu_{\text{Diff}}$  was not uniquely defined in Ref. 1. In fact, the possible rigging maps were labeled by uncountably many real parameters. However, all of these maps were of a similar form. Let us simply choose one of these maps and refer to it as  $\eta_{\text{Diff}}$ . We will see that nothing will depend on which map was chosen. Note that  $\eta_{\text{Diff}}$  then has the form

$$\eta_{\text{Diff}} T_c = \alpha_{[c]} T_{[c]} \tag{III.6}$$

for some  $\alpha_{[c]} \in \mathcal{R}^+$ .

Let  $c_0$  be a particular label. We wish to place our rigging map in the form

$$\eta_F T_{c_0} = (\sigma \circ \eta_{\text{Ham}} \circ \eta_{\text{Diff}}) T_{c_0} = (\sigma \circ \eta_{\text{Ham}}) T_{[c_0]} \alpha_{[c_0]}. \tag{III.7}$$

The map  $\eta_{\text{Ham}}$  will act on  $\Phi_{\text{Diff}}$ , the group averaged cylindrical functions on  $\overline{\mathcal{M}\mathcal{G}}$ . Notice that the space  $\Phi_{\text{Diff}}$  is a space of distributions on  $\Phi_{\text{Cyl}}$  but a space of test functions for the space  $\Phi'_{\text{Diff}}$ , the dual of  $\Phi_{\text{Diff}}$ .

We now address the map  $\sigma$ . It is to be an antilinear map from  $\Phi'_{\text{Diff}}$  to  $\Phi'_{\text{Cyl}}$ . We will construct this map by (anti)linearly extending the adjoint map on  $\mathcal{H}_{\text{Diff}}$ . Recall that,  $\mathcal{H}_{\text{Diff}}$  is defined through the following inner product on  $\Phi_{\text{Diff}}$ :

$$\langle \eta_{\text{Diff}} \phi, \eta_{\text{Diff}} \phi' \rangle_{\text{Diff}} := [\eta_{\text{Diff}} \phi](\phi') \quad \text{for all } \phi, \phi' \in \Phi_{\text{Cyl}}. \tag{III.8}$$

Thus (III.8) defines an antilinear (adjoint) map  $\dagger: \Phi_{\text{Diff}} \rightarrow \Phi'_{\text{Diff}}$ . On the image  $\dagger \Phi_{\text{Diff}} \subset \Phi'_{\text{Diff}}$ , this map is invertible and the inverse  $\dagger^{-1}$  is also antilinear. We note that  $\dagger \Phi_{\text{Diff}}$  in fact provides a basis for  $\Phi'_{\text{Diff}}$  and that  $\Phi_{\text{Diff}} \subset \Phi'_{\text{Cyl}}$ . Using antilinearity then, we may attempt to extend  $\dagger^{-1}$  to map from all of  $\Phi'_{\text{Diff}}$  into  $\Phi'_{\text{Cyl}}$ . The result is in fact well defined and gives the desired map  $\sigma: \Phi'_{\text{Diff}} \rightarrow \Phi'_{\text{Cyl}}$ .

Let us now define  $\eta_{\text{Ham}}$  to be of the form

$$\eta_{\text{Ham}} \cdot T_{[c]} = \sum_{[c']} a([c],[c']) T_{[c']}^\dagger, \tag{III.9}$$

where  $\dagger: \Phi_{\text{Diff}} \rightarrow \Phi'_{\text{Diff}}$  is the map given above. Then

$$(\sigma \circ \eta_{\text{Ham}}) T_{[c]} = \sum_{[c']} a^*([c],[c']) T_{[c']}. \tag{III.10}$$

The coefficients  $a$  will be chosen so that (III.7) is satisfied. The equality  $\alpha_{[c_0]}(\sigma \circ \eta_{\text{Ham}}) T_{[c_0]} = \delta(F) T_{c_0}$  is to be understood in the sense of distributions on  $\Phi_{\text{Cyl}}$  and so can be checked by evaluating both sides on all possible  $T_c$ . In order to do that we need the Clebsh–Gordon formula

$$T_{c_0} T_c = \sum_{c'} b(c_0, c; c') T_{c'}, \tag{III.11}$$

which is a finite sum thanks to the piecewise analyticity of the graphs involved. Notice that the coefficients  $b(c_0, c; c'(c_0, c))$  are invariant under simultaneous diffeomorphic mappings of  $c_0$  and  $c$ .

Finally, using  $T_{[c]}(T_{c'}) = \chi_{[c]}(c')$  [where  $\chi_{[c]}(c')$  is the characteristic function given by 1 for  $c' \in [c]$  and 0 otherwise] together with (III.11), we find

$$[\eta_F T_{c_0}](T_c) = [\eta_F(1)](T_{c_0} T_c) = \sum_{c'} b(c_0, c; c') [\eta_F(1)](T_{c'}) = \sum_{c'} b(c_0, c; c') k_{[c']}. \tag{III.12}$$

The first equality in (III.12) uses the fact that  $T_c$  is real valued. Notice that despite the appearance of the Clebsh–Gordon coefficients  $b(c_0, c; c')$  (which seem to depend on  $c_0, c$ ), the corresponding sum actually depends only on the equivalence classes  $[c_0], [c]$  since we have  $[\varphi \cdot T_{c_0}][\varphi' \cdot T_c] = T_{c_0} T_c$  on the space that of flat connections, for arbitrary  $\varphi, \varphi' \in \text{Diff}(\Sigma)$ .

Thus if we define  $a([c_0],[c]) = \alpha_{[c_0]}^{-1} \sum_{c'} b(c_0, c; c') k_{[c']}$ , then, since the Clebsh–Gordon coefficients are real we have

$$[(\sigma^\circ \eta_{\text{Ham}} \circ \eta_{\text{Diff}})(T_{[c_0]})](T_c) = \sum_{[c']} b([c_0],[c]; c') k_{[c']} = [\eta_{\text{Ham}} T_{[c_0]}](T_c). \quad (\text{III.13})$$

That is, we have constructed a map  $\eta_{\text{Ham}} : \Phi_{\text{Diff}} \rightarrow \Phi'_{\text{Diff}} \subset \Phi'_{\text{Cyl}}$  such that  $\eta_F = \sigma^\circ \eta_{\text{Ham}} \circ \eta_{\text{Diff}}$ . Note that the composition  $\eta_{\text{Ham}} \circ \eta_{\text{Diff}}$  is independent of  $\alpha_{[c]}$ , and thus independent of the particular choice of the map  $\eta_{\text{Diff}}$ .

Let us now examine the status of the “superselection rules” described in Ref. 1, associated with the averaging over diffeomorphisms. According to these rules, the states  $T_{[c]}, T_{[c']} \in H_{\text{Diff}}$  were superselected whenever  $c$  and  $c'$  were associated with graphs  $\gamma, \gamma'$  in distinct diffeomorphism classes. Note, however, that in this case we may still have  $[\eta_{\text{Ham}}(T_{[c]})](T_{[c']}) \neq 0$  or, equivalently,

$$\langle T_{[[c]]}, T_{[[c']]} \rangle_{\text{phys}} \neq 0, \quad (\text{III.14})$$

so that the corresponding states are not superselected. In fact, whenever  $c$  and  $c'$  are associated with homotopically equivalent triples  $c = (\gamma, \mathbf{j}, \mathbf{e})$  and  $c' = (\gamma', \mathbf{j}', \mathbf{e})$ , the states  $T_{[[c]]}$  and  $T_{[[c']]} are proportional. Furthermore, the operator  $\hat{T}_{\alpha, \text{phys}} : \hat{T}_{\alpha, \text{phys}} \eta_F f = \eta_F [T_{\alpha} f]$  is well-defined and mixes even homotopically distinct graphs. As a result, no sign of the superselection rules remains in the physical Hilbert space.$

#### IV. CONCLUSIONS

We have seen in Sec. II that the “superselection rules” among the diffeomorphism in variant states in no way carry over to the physical space. This result was not unexpected, as we took the “correct” description of 2+1 Euclidean gravity to be that given by Witten<sup>5</sup> in which no superselected sectors arise. However, the appearance of such spurious superselection rules in an intermediate stage was in no way an obstacle to the solution of the theory using loop representation and refined algebraic techniques, or even to solving first the diffeomorphism constraint and then implementing the remaining constraints. We recall that any of the possible maps  $\eta_{\text{Diff}}$  can be used and that they all lead to the same physical Hilbert space in the end. (In our presentation this was true by construction. However, the map  $\eta_F$  is unique at least up to the choice of measure  $d\nu$  while any choice of  $\eta_{\text{Diff}}$  is compatible with any choice of  $d\nu$ . Therefore, given any  $\eta_{\text{Diff}}$ , the same array of physical Hilbert spaces may be constructed through maps of the form  $\eta_{\text{Ham}} \circ \eta_{\text{Diff}}$ .) This can be taken as an encouraging sign for a similar approach in the 3+1 case. On the other hand, we have used the Witten constraints and the fact that they are well-defined on  $H_{\text{aux}}$  to achieve our goals. Such techniques are not available in the 3+1 case; it remains to be seen if this difference is crucial.

Let us now address the question of whether the diffeomorphism superselection rules will be spurious in 3+1 gravity. We first note that, as described in Refs. 1 and 4, there are many examples for which superselection laws arising from RAQ *are* of physical relevance, as they have analogues even at the classical level. What accounts for the difference between these systems? Several answers may be given. For example, in Ref. 4 it was found that spurious superselection rules can arise through a poor choice of the subspace  $\Phi$ . In general though, it appears that such spurious superselection rules are associated with singular structures in either the system or in our description of it.

To illustrate this point, recall the RAQ deals directly with only the strong observables of the system. Now, at least classically, this is no problem for any sufficiently smooth system. Let  $\Gamma$  be the phase space of a classical system with  $C$  the corresponding constraint surface and  $G$  the group of gauge transformations. When  $C/G$  is a smooth submanifold of  $\Gamma/G$ , all of the physics is indeed

captured by strong observables. Any observable (that is, any function on  $C/G$ ) may be extended to  $\Gamma/G$  and pulled back to  $\Gamma$ , where it defines a strong observable. Thus the strong observables capture all of the physics of the system.

In our quantum case, however, there were interesting observables (the  $\hat{T}_{\alpha, \text{phys}}$ ) which were not strongly diffeomorphism invariant. It would be interesting to understand whether this was due to some sort of singularity in the classical phase space or simply due to our quantum description. In any case, something analogous happens for 3+1 systems. It is shown in the Appendix that, at least in the representation based on the Ashtekar–Lewandowski Hilbert space, there are many quantum operators which are weak observables (with respect to the diffeomorphism constraint) but which do not become equivalent to any strongly diffeomorphism invariant observable when the diffeomorphism constraints are imposed.

There are of course several possible interpretations here. Note that the Appendix considers only the diffeomorphism (and gauge) constraints. It is therefore possible that, once the full algebra including the Hamiltonian constraints are considered, no analogue of these operators will remain. Another possibility is that these observables are simply spurious results of the quantization method and have no physical meaning. A third, however, is that such observables are important for a proper treatment of the system and that we must expand our techniques to take them into account. In any case, when we consider that the Hamiltonian constraint of 3+1 gravity<sup>11</sup> mixes the “superselected sectors” much as  $\eta_{\text{Ham}}$  does in the 2+1 case, it appears likely that the diffeomorphism superselection laws are spurious in 3+1 gravity as well.

## ACKNOWLEDGMENTS

Many thanks to the ESI, where much of this work was accomplished, for their hospitality and partial support. D.M. was supported in part by NSF Grant No. PHY95-07065 and funds provided by Syracuse University. J.M. was supported in part by CERN Grant No. CERN/S?FAE/1030/95 and in part by CENTRA/Algarve. T.T. was supported in part by DOE Grant No. DE-FG02-94ER25228 to Harvard University. We would like to thank Abhay Ashtekar for suggesting this project and Abhay Ashtekar, John Baez, and Jerzy Lewandowski for useful discussions.

## APPENDIX: STRONG QUANTUM OBSERVABLES IN 3+1 GRAVITY

In this appendix we show that the quantization scheme for 3+1 gravity considered in Ref. 1 contains operators that are weakly invariant under diffeomorphisms but which are not weakly equivalent to any operator which is *strongly* invariant under diffeomorphisms. Before proceeding, we should recall certain subtleties of the Ashtekar–Lewandowski Hilbert space and carefully define what we mean by weak observables. Recall that the diffeomorphism constraints themselves are not actually defined as operators on this space.<sup>1</sup> Instead, it is the finite diffeomorphisms (which may be interpreted as exponentiated versions of the constraints) which are defined on  $H_{\text{aux}}$ . These operators are, however, sufficient to define a space  $\Phi_{\text{Diff}}$  of diffeomorphism invariant states which are naturally thought of as the quantum analogue of the classical space of solutions to the diffeomorphism constraints. By weak equivalence of two operators  $B$  and  $C$ , we therefore mean that  $B$  and  $C$  coincide when acting on  $\Phi_{\text{Diff}}$ . Furthermore, a weak observable is naturally defined to be one which maps  $\Phi_{\text{Diff}}$  into itself.

The construction of the observables is quite straightforward. Recall that  $\Phi_{\text{Diff}}$  is in fact a space of “dual states,” specifically, of linear functionals on the space  $\Phi_{\text{cyl}}$  of cylindrical functions. Thus any operator  $A$  whose adjoint  $A^\dagger$  acts on and preserves the space  $\Phi_{\text{cyl}}$  of cylindrical states has a natural “dual action” on  $\Phi_{\text{Diff}}$  given by

$$[A \psi_{\text{Diff}}](\phi) = \psi_{\text{Diff}}(A^\dagger \phi). \quad (\text{A1})$$

Now, simply choose any two nontrivial spin network states  $T_1, T_2$  and consider the sets  $S_{T_1}, S_{T_2}$  of all states that can be obtained from  $T_1, T_2$ , respectively, by diffeomorphisms. Since  $T_1$  and  $T_2$  are each associated with analytic graphs, the cardinality of both sets  $S_{T_1}$ , and  $S_{T_2}$  is the same, namely that of the power set  $\mathcal{P}(\mathbf{R})$  of all real numbers. As a result, there is bijection  $\alpha$  between  $S_{T_1}$  and  $S_{T_2}$  and our observable  $A$  may be defined by

$$AT = \alpha(T) \quad \text{for } T \in S_{T_1}, \quad (\text{A2})$$

$$\text{while } A\psi = 0 \quad \text{if } \langle \psi | T \rangle = 0 \quad \text{for all } T \in S_{T_1}.$$

Note that  $A$  is a bounded operator whose range lies in  $V_{T_2}$ , the space of states spanned by spin network states in  $S_{T_2}$ . The adjoint  $A^\dagger$  is of a similar form but is defined by the map  $\alpha^{-1}$ .

Both  $A$  and  $A^\dagger$  are in fact weak observables. To see this, we simply compute the action of a diffeomorphism on  $A\psi_{\text{Diff}}$  for a diffeomorphism invariant state  $\psi_{\text{Diff}}$ . Since  $\Phi_{\text{Diff}}$  is a space of linear functionals on  $\Phi_{\text{Cyl}}$ ,  $A\psi_{\text{Diff}}$  is entirely determined by its action on spin network states, which form a basis for  $\Phi_{\text{Cyl}}$ . For any diffeomorphism  $D$  and any spin network  $T$ ,  $[DA\psi_{\text{Diff}}] \times (T) = \psi_{\text{Diff}}(A^\dagger D^{-1}T)$ . If  $T$  is orthogonal to the space  $V_{T_2}$  (spanned by spin networks in  $S_{T_2}$ ) then this vanishes. Otherwise, we may take  $T$  to be  $D'T_2$  for some diffeomorphism  $D'$ . In either case we have

$$[DA\psi_{\text{Diff}}](T) = \psi_{\text{Diff}}(D''A^\dagger T) = [A\psi_{\text{Diff}}](T) \quad (\text{A3})$$

for some diffeomorphism  $D''$ . Thus  $A$  preserves  $\Phi_{\text{Diff}}$  and is a weak observable with respect to diffeomorphisms. However, since we are free to choose  $T_1$  and  $T_2$  from different superselected sectors (as defined by Ref. 1) for the algebra of strongly diffeomorphism invariant operators, it is clear that the action of  $A$  on  $\Phi_{\text{Diff}}$  does not preserve the superselection sectors. As a result,  $A$  cannot be weakly equivalent to any strongly diffeomorphism invariant operator.

<sup>1</sup>A. Ashtekar, J. Lewandowski, D. Marolf, J. Mourão, and T. Thiemann, *J. Math. Phys.* **36**, 6456 (1995).

<sup>2</sup>N. Landsman, *J. Geom. Phys.* **15**, 285 (1995).

<sup>3</sup>A. Higuchi, *Class. Quantum Grav.* **8**, 1983 (1991); **8**, 2023 (1991).

<sup>4</sup>D. Marolf, Banach Center Publication; (to appear) Report No. gr-qc/9508015.

<sup>5</sup>E. Witten, *Nucl. Phys. B* **311**, 46 (1988).

<sup>6</sup>V. Moncrief, *J. Math. Phys.* **30**, 2907 (1989).

<sup>7</sup>G. Mess, "Lorentz spacetimes of constant curvature," Institut des Hautes Etudes Scientifiques report, 1990 (unpublished).

<sup>8</sup>A. Ashtekar, *Non-perturbative Canonical Gravity*, lecture notes prepared in collaboration with R. S. Tate (World Scientific, Singapore, 1991).

<sup>9</sup>J. D. Romano, *Gen. Relativ. Gravit.* **25**, 759 (1993).

<sup>10</sup>I. Bengtsson, *Phys. Lett. B* **220**, 51 (1989).

<sup>11</sup>T. Thiemann, *Phys. Lett. B* **380**, 257 (1996); "Quantum Spin Dynamics (QSD)," Harvard University Report No. HUTMP-96/B-351, gr-qc/9606089 (to appear in *Class. Quantum Grav.*); "Quantum Spin Dynamics (QSD) II", Harvard University Report No. HUTMP-96/B-352, gr-qc/9606090 (to appear in *Class. Quantum Grav.*).

<sup>12</sup>A. Ashtekar and J. Lewandowski (in preparation).

<sup>13</sup>A. Ashtekar and C. Isham, *Class. Quantum Grav.* **9**, 1433 (1992).

<sup>14</sup>A. Ashtekar and J. Lewandowski, *J. Math. Phys.* **36**, 2170 (1995).