PHYSICAL REVIEW D **80,** 084014 (2009)

Bootstrapping gravity: A consistent approach to energy-momentum self-coupling

Luke M. Butcher,* Michael Hobson, and Anthony Lasenby
Astrophysics Group, Cavendish Laboratory, J J Thomson Avenue, Cambridge CB3 0HE, UK
(Received 4 June 2009; published 13 October 2009)

It is generally believed that coupling the graviton (a classical Fierz-Pauli massless spin-2 field) to its own energy-momentum tensor successfully recreates the dynamics of the Einstein field equations order by order; however the validity of this idea has recently been brought into doubt [T. Padmanabhan, Int. J. Mod. Phys. D **17**, 367 (2008).]. Motivated by this, we present a graviton action for which energy-momentum self-coupling is indeed consistent with the Einstein field equations. The Hilbert energy-momentum tensor for this graviton is calculated explicitly and shown to supply the correct second-order term in the field equations; in contrast, the Fierz-Pauli action fails to supply the correct term. A formalism for perturbative expansions of metric-based gravitational theories is then developed, and these techniques employed to demonstrate that our graviton action is a starting point for a straightforward energy-momentum self-coupling procedure that, order by order, generates the Einstein-Hilbert action (up to a classically irrelevant surface term). The perturbative formalism is extended to include matter and a cosmological constant, and interactions between perturbations of a free matter field and the gravitational field are studied in a vacuum background. Finally, the effect of a nonvacuum background is examined, and the graviton is found to develop a nonvanishing ''mass-term'' in the action.

DOI: 10.1103/PhysRevD.80.084014 PACS numbers: 04.20.Cv

I. INTRODUCTION

It is a standard view in particle physics that the nonlinearity of a field theory, such as those of Yang and Mills, can be equated with the notion that the field in question carries the charge of the very interaction it mediates. This idea has been brought to bear on gravity many times, and various arguments [1–7] aim to derive general relativity from a linear starting point by coupling gravity to the energy and momentum of all fields, including the gravitational field itself. Despite the conventional wisdom that this self-coupling process is already well understood, Padmanabhan has uncovered a number of serious problems with the standard arguments [8]. Although we postpone an examination of Padmanabhan's analysis to Appendix A, it suffices to express here what is, in our view, his most pertinent observation: one cannot start with *linear gravity*, the Fierz-Pauli massless spin-2 action [8,9], and generate the higher-order corrections of general relativity by coupling the gravitational field to its own Hilbert energy-momentum tensor. More succinctly: one cannot derive the Einstein equations by bootstrapping gravitons[1] to their own energy and momentum.

To clarify the content of this observation, consider a perturbative expansion of the Einstein field equations $G_{\alpha\beta} = \kappa T^{\text{matter}}_{\alpha\beta}$ about a Minkowski background: $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$. Working to second-order in $h_{\alpha\beta}$, we obtain

$$G^{(1)}_{\alpha\beta} = -G^{(2)}_{\alpha\beta} + \kappa T^{\text{matter}}_{\alpha\beta}, \tag{1}$$

where the numbers in parentheses denote the powers of $h_{\alpha\beta}$ the term contains. Because $G^{(1)}_{\alpha\beta} = 0$ is the equation of motion for a massless spin-2 field $h_{\alpha\beta}$, the right-hand side of (1) can be interpreted as this field's source. Thus a satisfying physical picture suggests itself: the gravitational field $h_{\alpha\beta}$ is induced by the energy-momentum tensor of *all* fields $T_{\alpha\beta} = T^{\text{matter}}_{\alpha\beta} + t_{\alpha\beta}$, where $t_{\alpha\beta}$ is gravity's own energy-momentum tensor, identified as $-G^{(2)}_{\alpha\beta}/\kappa$. In actuality, however, this description cannot be formulated in a straightforward manner. Although the Fierz-Pauli action S_{FP} is typically used to prescribe the dynamics of a massless spin-2 field, its Hilbert energy-momentum tensor[2]

$$t_{\alpha\beta} \equiv \frac{-1}{\sqrt{-\gamma}} \frac{\delta S_{\text{FP}}}{\delta \gamma^{\alpha\beta}}, \tag{2}$$

*l.butcher@mrao.cam.ac.uk

[1]In discussions of this nature, the word *graviton* is often used as a shorthand for the classical massless spin-2 field. We follow this convention to cohere with the literature, but stress that this graviton is in no way quantum mechanical. What is actually being referred to is a *gravitational wave*, a classical fluctuation in the geometry of spacetime.

[2]Although other definitions of the energy-momentum tensor exist (see Sec. II C) we must define $t_{\alpha\beta}$ according to the Hilbert's prescription (2) in order to maintain the analogy with $T^{\text{matter}}_{\alpha\beta}$. This definition requires that S_{FP} be ''covariantized'' (represented in arbitrary coordinates using a *flat* metric $\gamma_{\alpha\beta}$) and a functional derivative taken with respect to the metric. It is important to realize that even though $\gamma_{\alpha\beta}$ is flat, the arbitrary variations $\delta\gamma_{\alpha\beta}$ required to construct the functional derivative inevitably explore *curved* metrics in a neighborhood of $\gamma_{\alpha\beta}$. Thus ''covariantization'' is not really sufficient: the action must be generalized to a curved background spacetime. One of the key aims of this paper is to generalize S_{FP} to curved spacetime in such a way that energy-momentum self-coupling is consistent with general relativity.

is *not* proportional to $G^{(2)}_{\alpha\beta}$, and thus cannot be used as the source-term for the second-order field equations. As an alternative approach, one could introduce energy-momentum self-coupling at the level of the action: because $t_{\alpha\beta}$ is a function of $h_{\alpha\beta}$, adding the self-coupling term $t_{\alpha\beta}h^{\alpha\beta}$ to the Lagrangian yields a different result from adding $t_{\alpha\beta}$ directly to the equations of motion. Unfortunately, this procedure also fails to generate $-G^{(2)}_{\alpha\beta}/\kappa$ in the field equations.

Padmanabhan claims that these realizations bring to light a previously neglected object $S^{\alpha\beta}$ (see Appendix A) which appears to codify the self-coupling of the gravitational field. Unfortunately, this object has many undesirable features: it is not a tensor under general coordinate transformations, has no clear physical interpretation, and fails to reveal any equivalence between the coupling of gravity to matter, and gravity to itself.

We propose an alternative solution to this apparent inconsistency: the action for the graviton is not the Fierz-Pauli action but is instead S_2 given by (4), possessing a nonminimally coupled term that vanishes when the (vacuum) background equations are enforced.[3] We shall demonstrate that the energy-momentum tensor of this action is the correct second-order contribution to the equation of motion, and furthermore, that this action provides the starting point for a straightforward energy-momentum self-coupling procedure that generates the Einstein-Hilbert action (modulo surface terms) to *arbitrary order*. We conclude the discussion by extending our formalism to nonvacuum spacetimes.

Throughout the article we employ the abstract index notation [10], with lower-case Roman indices indicating a tensor's ''slots,'' and Greek indices serving to enumerate its components in a particular coordinate system. The metric has signature $(-,+,+,+)$, $\kappa \equiv 8\pi G/c^4$, and the Riemann and Ricci tensor are defined with the following conventions: $R^a{}_{bcd}v^b \equiv 2\nabla_{[c}\nabla_{d]}v^a$, $R_{ab} \equiv R^c{}_{acb}$.

II. THE GRAVITON ACTION

Contrary to the standard approach, we represent the gravitational field as a perturbation h^{ab} of the *inverse* physical metric g^{ab} from the background $\bar{g}^{ab}$:

$$g^{ab} = \bar{g}^{ab} + h^{ab}. \tag{3}$$

This expression is *exact* in that we have not neglected terms $O(h^2)$; in contrast, the physical metric $g_{ab} = \bar{g}_{ab} - h^{cd}\bar{g}_{ca}\bar{g}_{db} + O(h^2)$. Following this convention, we use the contravariant field h^{ab}, rather than h_{ab}, as the fundamental dynamical variable of the action.[4] In general we will write bars over tensors derived solely from the background geometry, and adopt the usual notational convenience of raising and lowering indices with $\bar{g}^{ab}$ and $\bar{g}_{ab}$.[5]

We posit that the dynamics, energy and momentum of the gravitational field h^{ab}, propagating in a background spacetime with metric $\bar{g}_{ab}$, are all determined (to lowest-order) by the following action:

$$S_2[\bar{g}^{ab}, h^{ab}] \equiv \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}(\hat{G}_{abcd} + \bar{H}_{abcd})h^{cd}, \tag{4}$$

where

$$\hat{G}_{abcd} \equiv \frac{1}{2}(\bar{g}_{a(c}\bar{g}_{d)b} - \bar{g}_{ab}\bar{g}_{cd})\bar{\nabla}^2 - \bar{\nabla}_{(c}\bar{g}_{d)(a}\bar{\nabla}_{b)} + \frac{1}{2}\bar{g}_{ab}\bar{\nabla}_{(c}\bar{\nabla}_{d)} + \frac{1}{2}\bar{g}_{cd}\bar{\nabla}_{(a}\bar{\nabla}_{b)} \tag{5}$$

is a differential operator representing the linearized Einstein tensor (see Appendix B) and

$$\bar{H}_{abcd} \equiv \frac{1}{2}\bar{R}\left(\bar{g}_{ac}\bar{g}_{db} + \frac{1}{2}\bar{g}_{ab}\bar{g}_{cd}\right) - \bar{R}_{ab}\bar{g}_{cd}. \tag{6}$$

While $\bar{H}_{abcd}$ has no obvious geometric interpretation, we intend to show that its contribution to the action is necessary for the consistency of energy-momentum self-coupling with general relativity. Further motivation for this ansatz is given in Sec. III.

[3]More precisely, S_2 is the action for the graviton in a background spacetime with metric in some small neighborhood of the solutions of the vacuum field equations. We use the term *vacuum* to signify a region without matter; this does not necessarily imply the absence of spacetime curvature.

[4]Any metric theory of gravity will have an ambiguity as to which variable $g \in \{g^{ab}, g_{ab}, \sqrt{-g}g^{ab}, \ldots\}$ should be identified as the true ''gravitational field.'' Such a distinction is of no physical consequence and is largely unnecessary for a nonperturbative calculation; however for the present discussion we are forced to single out a particular field variable for the expansion $g = \bar{g} + h$. Our aim is to connect gravity to the particle physics notion of a spin-2 field and elucidate a simple energy-momentum self-coupling scheme that generates general relativity; to this end we are required to pick $g \in \{g^{ab}, g_{ab}\}$ as it is only for these that h is a genuine spin-2 field, i.e. a symmetric tensor (not a tensor density) with (lowest-order) infinitesimal gauge transformation $\delta h^{ab} = 2\bar{\nabla}^{(a}\varepsilon^{b)}$. Fortunately, it is precisely for $g \in \{g^{ab}, g_{ab}\}$ that the necessary energy-momentum self-coupling is its most simple: $h^{ab}t_{ab}$ (see Sec. III). These considerations provide no criteria for choosing the metric over its inverse as our expansion variable, and while this choice only trivially alters the perturbation theory at first-order ($h^{ab} \leftrightarrow -h_{ab}$) to second-order (the relevant order for S_2, t_{ab}, and $G^{(2)}_{ab}$) the two definitions of the h-field differ by a term of the form $h^{ac}h^b{}_c$. Our choice of $g = g^{ab}$ is preferable for this article because it simplifies the mathematics of the action and energy-momentum tensor. The reason for this is explored in Sec. III E, and stems from the fact that any Lagrangian for pure gravity must contain more factors of g^{ab} than g_{ab} in order that all the derivatives ∂_a be contracted; thus an expansion in $g = g^{ab}$ will be algebraically simpler. Indeed, this observation still holds when coupling gravity to a scalar field ϕ or a 1-form A_a, and thus taking $g = g^{ab}$ simplifies many of the calculations of the nonvacuum case also (see Sec. IV).

[5]The only exception to this rule is the physical metric and its inverse, for which $g^{ab} \neq g_{cd}\bar{g}^{ac}\bar{g}^{db}$, but rather $g^{ab}g_{bc} = \delta^a_c$.

Naturally, if we are to obtain general relativity without at first assuming it, we must begin by considering the graviton in a *flat* background spacetime. Nevertheless, we will see from the formalism of Sec. III that (provided we use S_2 to describe the graviton) energy-momentum self-coupling generates the Einstein-Hilbert action even when the background is not flat; $\bar{g}^{ab}$ need only satisfy the weaker condition

$$\bar{G}_{ab} \equiv \bar{R}_{ab} - \frac{1}{2}\bar{g}_{ab}\bar{R} = 0. \tag{7}$$

While this equation expresses the generality of the analysis that is to follow, it should be stressed that no knowledge of (7) will be required to assemble the Einstein-Hilbert action order by order: a flat background will serve as a perfectly satisfactory starting point.[6] No matter which background we use, however, it is absolutely crucial that we refrain from inserting this particular metric [or even Eq. (7)] into the action, thereby reducing S_2 to $\frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}\hat{G}_{abcd}h^{cd}$. This is because we will need to be able to perform arbitrary variations of $\bar{g}^{ab}$, not just those consistent with $\bar{R}_{abcd} = 0$ or $\bar{R}_{ab} = 0$, to construct the energy-momentum tensor for h^{ab}. That said, it will be instructive to temporarily ignore this advice so that we may relate S_2 to the Fierz-Pauli action.

A. The Fierz-Pauli action

For a flat background, $\bar{H}_{abcd}$ vanishes, and we can choose coordinates $\{x^\alpha\}$ such that $\bar{g}^{\alpha\beta} = \eta^{\alpha\beta}$ and evaluate S_2 as a functional of the components $h^{\alpha\beta}$. Integrating by parts and discarding surface terms, we find that S_2 reduces to $\frac{-1}{2\kappa}\int d^4x\mathcal{L}_{\mathrm{FP}}$, where

$$\mathcal{L}_{\mathrm{FP}} = \frac{1}{2}\partial_\lambda h_{\alpha\beta}\partial^\lambda h^{\alpha\beta} - \frac{1}{2}\partial_\lambda h\partial^\lambda h - \partial_\lambda h^{\alpha\beta}\partial_\alpha h_\beta{}^\lambda + \partial_\alpha h\partial_\beta h^{\alpha\beta} \tag{8}$$

is the Fierz-Pauli Lagrangian [8].[7] Modulo surface terms and an overall rescaling, $\mathcal{L}_{\mathrm{FP}}$ is the unique specially relativistic Lagrangian for a symmetric tensor field $h^{\alpha\beta}$ that is invariant under the infinitesimal gauge transformation $\delta h^{\alpha\beta} = 2\partial^{(\alpha}\varepsilon^{\beta)}$ (see [8] for proof); hence it is the Lagrangian for the graviton (massless spin-2 field) in flat spacetime.

[6]Of course, once the self-coupling procedure is complete, and the Einstein-Hilbert action has been assembled starting from the graviton on a flat background, we will be in an excellent position to justify (7), as this is precisely the field equation (applied to the background) that we will have derived. With hindsight, then, we can see there was nothing special about our flat-space starting point: we may begin with any *one* solution to (7) and use energy-momentum self-coupling to derive the action (and field equation) that defines *all* the others.

[7]Here and elsewhere we use the customary shorthand $h \equiv h^a{}_a \equiv h^{ab}\bar{g}_{ab}$.

Starting from (8), we can covariantize $\mathcal{L}_{\mathrm{FP}}$ by making the replacements $\eta_{\alpha\beta} \to \bar{g}_{\alpha\beta}$, $\partial_\alpha \to \bar{\nabla}_\alpha$ and multiplying by $\sqrt{-\bar{g}}$. This process obviously generates a unique manifestly covariant Lagrangian density if $\bar{g}^{ab}$ is flat, as in this case the procedure is equivalent to representing the same Lagrangian in arbitrary coordinates. However, for the purposes of calculating the energy-momentum tensor (via arbitrary variations of $\bar{g}^{ab}$) it will be necessary to generalize $\mathcal{L}_{\mathrm{FP}}$ to arbitrary backgrounds, and for a curved metric the covariantization procedure is ambiguous. To see this, observe that we can transmute the third term of (8) by twice integrating by parts:

$$\partial_\lambda h^{\alpha\beta}\partial_\alpha h_\beta{}^\lambda \leftrightarrow \partial_\alpha h^{\alpha\beta}\partial_\lambda h_\beta{}^\lambda. \tag{9}$$

However this equivalence relies on the commutativity of partial derivatives, and does not occur for the covariant derivatives of a curved background; instead, integration by parts yields

$$\bar{\nabla}_c h^{ab}\bar{\nabla}_a h_b{}^c \leftrightarrow \bar{\nabla}_a h^{ab}\bar{\nabla}_c h_b{}^c - h^{ca}h^b{}_c\bar{R}_{ab} - h^{ab}h^{cd}\bar{R}_{acdb}. \tag{10}$$

Thus we are forced to make a seemingly arbitrary choice: do we to covariantize (8) as written, or should we do so after performing (9)? These two possibilities determine Lagrangians which differ by $h^{ca}h^b{}_c\bar{R}_{ab} + h^{ab}h^{cd}\bar{R}_{acdb}$; they lead to different (first-order) equations of motion if the background is curved,[8] and determine different energy-momentum tensors even if the background is flat.[9] This last problem is discussed by Padmanabhan [8], and is one of his many nontrivial objections to the conventional wisdom that general relativity is the unique energy-momentum self-coupled limit of the flat-space massless spin-2 field.

A greater problem than this ambiguity, however, is that neither choice (nor an admixture) leads to general relativity after coupling it to its own energy-momentum. As we shall see in Sec. III, the contribution from $h^{ab}\bar{H}_{abcd}h^{cd}$ is necessary to achieve this, and it is impossible to use the covariantizing ambiguity to produce this tensor because it does not contain $h^{ab}h^{cd}\bar{R}_{acdb}$. Instead, the presence of $\bar{H}_{abcd}$ represents a rather different coupling ambiguity faced when moving from a flat background to a curved one. Typically we would invoke the Einstein equivalence principal to banish from the action terms coupling matter fields and Ricci tensors; we would argue that, working in locally inertial coordinates about a point p, the Lagrangian at p should have the same form as the Lagrangian in flat spacetime. This amounts to a minimal coupling procedure:

[8]The first-order field equation only describes the spacetime perturbations of general relativity if the ambiguous term is covariantized to become $\bar{\nabla}_c h^{ab}\bar{\nabla}_a h_b{}^c$; see Sec. II B and Appendix B.

[9]Note that all other terms of $\mathcal{L}_{\mathrm{FP}}$ are invariant under the operation that generated (9) so do not introduce further ambiguity.

once we have covariantized a specially relativistic Lagrangian, the job of coupling the field to the gravity is complete. However, while this rule may make sense to curve the background spacetime of a spin-2 field that is "just another matter-field" and has nothing to do with gravitation, it is far from clear that the principal should hold for the graviton, for which it was only ever a convenient fiction to think of as a tensor field propagating over a background geometry.

In summary, the Fierz-Pauli action is insufficient to determine S_2 for an arbitrary background geometry; the principal of equivalence fails to give a unique solution, and cannot justify all the contributions necessary for an energy-momentum self-coupling procedure consistent with general relativity. However, it was never our aim to construct general relativity from $\mathcal{L}_{\mathrm{FP}}$, and we do not pretend to be able to derive a curved spacetime theory of gravity from purely specially relativistic concepts. S_2 will serve as our starting point, and the only significance we shall ascribe $\mathcal{L}_{\mathrm{FP}}$ is that of a special case.

B. Field equations

Leaving the Fierz-Pauli action behind, we retrain our attention on S_2 and begin the process of deriving its advertised connection to general relativity. First, we shall calculate the associated field equations. As usual, the equations of motion are derived from the condition that their solutions be stationary configurations of S_2 with respect to variations in the dynamical field h^{ab}. As we will have no cause to vary $\bar{g}^{ab}$ in the derivation, we can enforce the background equations (7) immediately and discard $\bar{H}_{abcd}$. Next, observe that $\hat{G}_{abcd}$ is "self-conjugate": for any tensor fields A^{ab} and B^{ab}

$$\int d^4x\sqrt{-\bar{g}}A^{ab}\hat{G}_{abcd}B^{cd} = \int d^4x\sqrt{-\bar{g}}B^{ab}\hat{G}_{abcd}A^{cd}, \tag{11}$$

provided either A^{ab} or B^{ab} has compact support. Therefore, holding $\bar{g}^{ab}$ constant and performing a variation δh^{ab} (a symmetric tensor field with compact support) gives rise to a variation in the action

$$\delta S_2 = \frac{1}{\kappa}\int d^4x\sqrt{-\bar{g}}\delta h^{ab}\hat{G}_{abcd}h^{cd}. \tag{12}$$

As $\hat{G}_{abcd}$ is already symmetric in its first two indices, we can conclude that the equation of motion is

$$\frac{1}{\sqrt{-\bar{g}}}\frac{\delta S_2}{\delta h^{ab}} = \kappa^{-1}\hat{G}_{abcd}h^{cd} = 0. \tag{13}$$

The centrally important feature of this equation is that $\hat{G}_{abcd}h^{cd} = G^{(1)}_{ab}$, the linear approximation to the Einstein tensor under the inverse metric expansion (3). This is particularly easy to verify for the special case of a flat background in Lorentzian coordinates, but is shown to hold more generally for vacuum backgrounds in Appendix B. Thus S_2 prescribes the correct first-order equation of motion for the graviton. In the next section we show that by adding the energy-momentum tensor t_{ab} of h^{ab} (determined by S_2) to the right-hand side of (13) we successfully generate the Einstein field equations correct to *second-order.*[10]

C. Energy-momentum tensor

We will now calculate the energy-momentum tensor of the graviton and relate it to the second-order contribution to the Einstein field equations. We follow Hilbert's prescription and define the energy-momentum tensor as a functional derivative of the action with respect to the (background) metric:

$$t_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S_2}{\delta \bar{g}^{ab}}, \tag{14}$$

where h^{ab} (rather than h_{ab} or $h^a{}_b$) is to be held constant when taking this derivative, as this is the field we have taken to be the fundamental dynamical variable.[11]

As an aside, it is worth contrasting the variational definition (14) with Noether's (canonical) energy-momentum tensor:

$$t^{\mu\nu}_{\mathrm{can}} \equiv \frac{\partial\mathcal{L}}{\partial(\partial_\mu h^{\alpha\beta})}\partial^\nu h^{\alpha\beta} - \eta^{\mu\nu}\mathcal{L}, \tag{15}$$

comprising the four conserved currents associated with the invariance of the Lagrangian $\mathcal{L}$ under rigid spacetime translations. The canonical tensor cannot be used in the present discussion for a number of reasons. Firstly, it is not uniquely determined by the action for h^{ab}: as it depends directly on the Lagrangian, we are free to alter $t^{\mu\nu}_{\mathrm{can}}$ by adding a four-divergence to $\mathcal{L}$, without changing either the dynamics of h^{ab} or S_2. Secondly, we require a *symmetric* tensor to act as the source for the first-order field equation (13), but the canonical tensor need not have this property.[12] Lastly, Noether's definition does not naturally generalize to curved spacetime in such a way that $t^{\mu\nu}_{\mathrm{can}}$ inherits a *covariant* conservation law [11]. None of these

[10]Of course, the resulting field equation will no longer be a stationary configuration of the action S_2. In order that this self-coupled equation of motion can be derived from the principle of stationary action it will be necessary to introduce a third-order correction to the action S_3. Naturally, S_3 will alter the energy-momentum tensor of h^{ab} by a term $O(h^3)$; however, seemingly by miracle, this will be precisely the *third-order* part of the Einstein field equations. This process continues indefinitely and is explained systematically in Sec. III. For the moment we content ourselves with exploring the theory to second-order only.

[11]In later sections, the tensor written here as t_{ab} will be notated t^2_{ab} to indicate that it is the energy-momentum contribution from the second-order action S_2 only. Here we need not make this distinction.

[12]It is true that the canonical tensor can be *made* symmetric by adding to it an identically conserved "correction" $\partial_\alpha\phi^{\mu[\nu\alpha]}$, a function of h^{ab} that cancels the antisymmetric part of $t^{\mu\nu}_{\mathrm{can}}$. However, if we allow this sort of *ad hoc* adjustment of the energy-momentum tensor, we only exacerbate the problem of nonuniqueness.

issues arise with t_{ab}, and in any case our aim has been to connect the coupling between matter and gravity found in general relativity with a perturbative coupling of gravity to itself; it is the Hilbert energy-momentum tensor of matter, not the canonical tensor, that appears in the full Einstein field equations as the gravitational source. For these reasons we discard the canonical tensor and henceforth refer to t_{ab}, following Hilbert's prescription (14), as the energy-momentum tensor of h^{ab}.

To begin the calculation of t_{ab}, we divide the action into two pieces $S_2 = S_{2G} + S_{2H}$:

$$S_{2G} \equiv \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}\hat{G}_{abcd}h^{cd}, \tag{16}$$

$$S_{2H} \equiv \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}\bar{H}_{abcd}h^{cd}. \tag{17}$$

It will be convenient to perform the functional derivative (14) on these two components separately. Focusing first on S_{2G}, we integrate by parts[13] so as to remove the second derivatives from the integrand:

$$S_{2G} = \frac{-1}{2\kappa}\int d^4x\sqrt{-\bar{g}}\bar{\nabla}_c h^{ab}\bar{\nabla}_d h^{ef}K_{ab\ ef}^{\ \ c\ \ d}, \tag{18}$$

for which we have introduced the abbreviation

$$\begin{aligned}K_{ab\ ef}^{\ \ c\ \ d} &\equiv \frac{1}{2}(\bar{g}^{cd}\bar{g}_{a(e}\bar{g}_{f)b} - \bar{g}^{cd}\bar{g}_{ab}\bar{g}_{ef} - 2\delta^c_{(e}\bar{g}_{f)(a}\delta^d_{b)} \\ &\quad + \delta^c_{(e}\delta^d_{f)}\bar{g}_{ab} + \delta^d_{(a}\delta^c_{b)}\bar{g}_{ef}) \\ &= K_{ba\ ef}^{\ \ c\ \ d} = K_{ab\ fe}^{\ \ c\ \ d} = K_{ef\ ab}^{\ \ d\ \ c}.\end{aligned} \tag{19}$$

An infinitesimal variation in the inverse background metric $\delta\bar{g}^{ab}$, vanishing on the boundary of the integral, induces a variation in the action

$$\begin{aligned}\delta S_{2G} = \frac{-1}{2\kappa}\int d^4x\sqrt{-\bar{g}}\Big[\delta\bar{g}^{pq}\bar{\nabla}_c h^{ab}\bar{\nabla}_d h^{ef}\Big(\frac{\partial K_{ab\ ef}^{\ \ c\ \ d}}{\partial\bar{g}^{pq}} \\ - \frac{1}{2}\bar{g}_{pq}K_{ab\ ef}^{\ \ c\ \ d}\Big) + 4\bar{\nabla}_c h^{ab}C^{(e}_{\ \ sd}h^{f)s}K_{ab\ ef}^{\ \ c\ \ d}\Big],\end{aligned}$$

where

$$\begin{aligned}C^a_{\ bc} &\equiv \frac{1}{2}\bar{g}^{ad}(\bar{\nabla}_b\delta\bar{g}_{cd} + \bar{\nabla}_c\delta\bar{g}_{bd} - \bar{\nabla}_d\delta\bar{g}_{bc}) \\ &= -\frac{1}{2}(2\delta^a_p\delta^r_{(b}\bar{g}_{c)q} - \bar{g}^{ar}\bar{g}_{bp}\bar{g}_{qc})\bar{\nabla}_r\delta\bar{g}^{pq}\end{aligned} \tag{20}$$

is the connection that arises from the variation of the covariant derivative: $\nabla_{\bar{g}+\delta\bar{g}} = \bar{\nabla} + C$. We can move the covariant derivatives off $\delta\bar{g}^{pq}$ in the connection term using integration by parts, and arrive at an equation of the form $\delta S_{2G} = \int d^4x\delta\bar{g}^{pq}[\ldots]_{pq}$; the tensor density in square brackets is then the functional derivative we seek:

$$\begin{aligned}\frac{\kappa}{\sqrt{-\bar{g}}}\frac{\delta S_{2G}}{\delta\bar{g}^{pq}} &= \frac{-1}{2}\bar{\nabla}_c h^{ab}\bar{\nabla}_d h^{ef}\Big(\frac{\partial K_{ab\ ef}^{\ \ c\ \ d}}{\partial\bar{g}^{pq}} - \frac{1}{2}\bar{g}_{pq}K_{ab\ ef}^{\ \ c\ \ d}\Big) \\ &\quad - \bar{\nabla}_r(\bar{\nabla}_c h^{ab}(K_{ab\ (p|f|q)}^{\ \ c}h^{rf} + K_{ab\ f(p}^{\ \ c\ \ r}h_{q)}^{\ \ f} \\ &\quad - K_{ab}^{\ \ cr}{}_{f(p}h_{q)}^{\ \ f})).\end{aligned} \tag{21}$$

Meanwhile, S_{2H} varies by

$$\begin{aligned}\delta S_{2H} &= \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}\delta\bar{R}_{ab} \\ &\quad \times\Big(\frac{1}{2}\bar{g}^{ab}\Big(\frac{1}{2}h^2 + h_{cd}h^{cd}\Big) - h^{ab}h\Big),\end{aligned} \tag{22}$$

where we have used the background equation (7) (*after* the variation) to remove the terms proportional to $\bar{R}_{ab}$; these would only be significant if we intended to perform further variations in the metric. Now, because

$$\begin{aligned}\delta\bar{R}_{ab} &= 2\bar{\nabla}_{[c}C^c_{\ b]a} \\ &= \Big(\frac{1}{2}\bar{g}^{rs}\bar{g}_{ap}\bar{g}_{qb} + \frac{1}{2}\delta^r_{(a}\delta^s_{b)}\bar{g}_{pq} \\ &\quad - \delta^r_p\delta^s_b\bar{g}_{aq}\Big)\bar{\nabla}_r\bar{\nabla}_s\delta\bar{g}^{pq},\end{aligned}$$

when we (twice) integrate by parts to alleviate $\delta\bar{g}^{ab}$ of its covariant derivatives, we generate a second-order differential operator

$$\hat{R}_{\ pqab} \equiv \frac{1}{2}\bar{g}_{a(p}\bar{g}_{q)b}\bar{\nabla}^2 + \frac{1}{2}\bar{g}_{pq}\bar{\nabla}_{(a}\bar{\nabla}_{b)} - \bar{\nabla}_{(a}\bar{g}_{b)(p}\bar{\nabla}_{q)}, \tag{23}$$

with the property

$$\int d^4x\sqrt{-\bar{g}}\delta\bar{R}_{ab}A^{ab} = \int d^4x\sqrt{-\bar{g}}\delta\bar{g}^{pq}\hat{R}_{pqab}A^{ab} \tag{24}$$

for all A^{ab}. Therefore, we can conclude from (22) that

$$\frac{\kappa}{\sqrt{-\bar{g}}}\frac{\delta S_{2H}}{\delta\bar{g}^{pq}} = \frac{1}{2}\hat{R}_{pqab}\Big(\frac{1}{2}\bar{g}^{ab}\Big(\frac{1}{2}h^2 + h_{cd}h^{cd}\Big) - h^{ab}h\Big). \tag{25}$$

Finally, we have only to combine Eqs. (21) and (25), expand out all the products and derivatives, and assemble the outcome into a formula for t_{ab} as a function of $\bar{\nabla}_c h^{ab}$. This is a straightforward but arduous calculation, and as such we chose to complete it with a computer algebra package. The result is

[13]More precisely, one adds to the integrand a divergence of the form $\partial_a(\sqrt{-\bar{g}}[h\bar{\nabla}h]^a) = \sqrt{-\bar{g}}\bar{\nabla}_a[h\bar{\nabla}h]^a$ that alters S_2 only by a function of the fields on the boundary (or at infinity) and thus may be neglected for the purposes of functional variation.

$$\kappa t_{pq} = \frac{1}{4}\bar{g}_{pq}\Big(h\bar{\nabla}_a\bar{\nabla}_b h^{ab} + 2h^{ab}\bar{\nabla}_a\bar{\nabla}_b h - 2h_{ab}\bar{\nabla}^2 h^{ab} - h\bar{\nabla}^2 h - \frac{1}{2}\bar{\nabla}_a h\bar{\nabla}^a h - \frac{5}{2}\bar{\nabla}_c h_{ab}\bar{\nabla}^c h^{ab} + \bar{\nabla}_c h_a{}^b\bar{\nabla}_b h^{ac} + 2\bar{\nabla}_a h\bar{\nabla}_b h^{ab}\Big) + \frac{1}{4}h\bar{\nabla}_{(p}\bar{\nabla}_{q)}h - \frac{1}{2}h_{pq}\bar{\nabla}^2 h + \frac{1}{4}h\bar{\nabla}^2 h_{pq} + h_{a(p}\bar{\nabla}^2 h_{q)}{}^a - \frac{1}{2}h^{ab}\bar{\nabla}_a\bar{\nabla}_b h_{pq} + \frac{1}{2}h_{pq}\bar{\nabla}_a\bar{\nabla}_b h^{ab} - h_{a(p}\bar{\nabla}^b\bar{\nabla}_{q)}h^a{}_b + \frac{1}{2}h_{ab}\bar{\nabla}_{(p}\bar{\nabla}_{q)}h^{ab} - \frac{1}{2}h\bar{\nabla}_a\bar{\nabla}_{(p}h_{q)}{}^a + \frac{1}{4}\bar{\nabla}_a h\bar{\nabla}^a h_{pq} + \frac{1}{2}\bar{\nabla}_b h_{ap}\bar{\nabla}^b h^a{}_q - \frac{1}{2}\bar{\nabla}_a h_{pq}\bar{\nabla}_b h^{ab} + \frac{3}{4}\bar{\nabla}_p h_{ab}\bar{\nabla}_q h^{ab} - \bar{\nabla}_b h^a{}_{(p}\bar{\nabla}_{q)}h_a{}^b - \frac{1}{2}\bar{\nabla}_b h\bar{\nabla}_{(p}h_{q)}{}^b + \frac{1}{2}\bar{\nabla}_b h^a{}_p\bar{\nabla}_a h^b{}_q. \tag{26}$$

It is possible to render this formula rather more manageable by working in a gauge with $\bar{\nabla}_a h^{ab} = 0$, $h = 0$:

$$\kappa t_{pq} = \bar{g}_{pq}\Big(\frac{1}{4}\bar{\nabla}_c h_a{}^b\bar{\nabla}_b h^{ac} - \frac{5}{8}\bar{\nabla}_c h_{ab}\bar{\nabla}^c h^{ab} - \frac{1}{2}h_{ab}\bar{\nabla}^2 h^{ab}\Big) + h_{a(p}\bar{\nabla}^2 h_{q)}{}^a - \frac{1}{2}h^{ab}\bar{\nabla}_a\bar{\nabla}_b h_{pq} - h^{bc}\bar{R}_{abc(p}h_{q)}{}^a + \frac{1}{2}h_{ab}\bar{\nabla}_{(p}\bar{\nabla}_{q)}h^{ab} + \frac{1}{2}\bar{\nabla}_b h_{ap}\bar{\nabla}^b h^a{}_q + \frac{3}{4}\bar{\nabla}_p h_{ab}\bar{\nabla}_q h^{ab} - \bar{\nabla}_b h^a{}_{(p}\bar{\nabla}_{q)}h_a{}^b + \frac{1}{2}\bar{\nabla}_b h^a{}_p\bar{\nabla}_a h^b{}_q, \tag{27}$$

but we will not need this partially gauge-fixed result for this present article.[14]

Our task now is to compare t_{ab} with $G^{(2)}_{ab}$ and demonstrate that the energy-momentum self-coupling of h^{ab} (determined by S_2) is consistent with general relativity. Details of the calculation of $G^{(2)}_{ab}$ can be found in Appendix B; the conclusion is

$$G^{(2)}_{ab} = -\kappa t_{ab} + O(h^3), \tag{28}$$

and thus, to second-order, the vacuum Einstein field equations are

$$\hat{G}_{abcd}h^{cd} = \kappa t_{ab} \tag{29}$$

as advertised.

As a corollary of (29), we can confirm Padmanabhan's observation that general relativity cannot be derived from energy-momentum self-coupling the Fierz-Pauli Lagrangian. Only once the contribution from $\bar{H}_{abcd}$ is included will Einstein's gravity result from an energy-momentum self-coupled graviton. This realization casts doubt on Mannheim's recent treatment of gravitational energy-momentum [12], in which a tensor is constructed by applying (14) to a covariantized Fierz-Pauli Lagrangian, rather than S_2.

III. PERTURBATIVE GRAVITY

Here we develop the formalism to uncover the root cause of the second-order energy-momentum self-coupling (29), and reveal how the process continues to arbitrary order. The vast majority of this section applies to any metric theory of pure gravity[15] and can be generalized to include interactions with matter (see Sec. IV). Only in Sec. III E will we commit to general relativity, fix our action $S = S_{\mathrm{EH}}$, the Einstein-Hilbert action, and derive the formula (4) for S_2.

We shall concern ourselves with an expansion of the inverse metric g^{ab} about a nondynamical background $\bar{g}^{ab}$, which is itself an exact solution of the vacuum field equations:

$$g^{ab} = \bar{g}^{ab} + \lambda h^{ab}, \tag{30}$$

$$0 = \frac{\delta S[\bar{g}]}{\delta\bar{g}^{ab}}, \tag{31}$$

where λ, a dimensionless expansion parameter, is constant over spacetime.

Following (30), the action of the exact theory $S[g]$ becomes a λ-dependent functional of $\bar{g}^{ab}$ and h^{ab}, which can be Taylor expanded thusly:

$$S[g] = S[\bar{g} + \lambda h] = \sum_{n=0}^{\infty}\lambda^n S_n[\bar{g}, h], \tag{32}$$

where S_n is the "nth partial action" given by

$$S_n[\bar{g}, h] = \frac{1}{n!}(\partial^n_\lambda S[\bar{g} + \lambda h])_{\lambda=0}. \tag{33}$$

The derivative ∂_λ acts on each instance of λh^{ab} in the integrand of $S[\bar{g} + \lambda h]$ by Leibniz's law, removing the factor of λ. The "bare" h^{ab} left behind may still be covered by spacetime derivatives ∂_a, but these can be moved onto the remainder of the integrand by integration by parts. This operation generates the usual functional derivative:

[14]Gauge transformations are covered in Sec. III C; we note here only that because t_{ab} is not invariant under the infinitesimal gauge transformation $\delta h^{ab} = 2\bar{\nabla}^{(a}\varepsilon^{b)}$, only the first formula (26) can be used in all gauges. Although gauge invariance would be a highly desirable property if we intended to argue that t_{ab} was a physically meaningful tensor in full general relativity, it is an impossible request to make of the tensor we seek, which should be proportional to the gauge dependent tensor $G^{(2)}_{ab}$.

[15]We require only that the dynamics are determined by an action that is a coordinate-independent integral of the metric and its derivatives.

$$\partial_\lambda S[\bar{g} + \lambda h] = \int d^4x h^{ab}(x) \frac{\delta}{\delta \bar{g}^{ab}(x)} S[\bar{g} + \lambda h]. \quad (34)$$

In truth, the left-hand side of this equation differs from the right by the surface term $\int d^4x \partial_a J^a$ created when integrating by parts. As this is only a functional of the fields on the boundary (or as $x^\mu \to \infty$ if the integral of S runs over the entire manifold) it will not contribute to equations of motion or energy-momentum tensors, the calculation of which are dependent only on variations of the field that vanish on the boundary (or have compact support). Hence these surface terms may be neglected for our present purposes.

It follows from the repeated application of (34) that

$$\partial^n_\lambda S[\bar{g} + \lambda h] = \left[\int d^4x h^{ab} \frac{\delta}{\delta \bar{g}^{ab}} \right]^n S[\bar{g} + \lambda h], \quad (35)$$

and thus the partial actions (33) are given by

$$S_n[\bar{g}, h] = \frac{1}{n!} \left[\int d^4x h^{ab} \frac{\delta}{\delta \bar{g}^{ab}} \right]^n S[\bar{g}]. \quad (36)$$

An important consequence of this relation is that, using S_2 as our starting point, we can generate the entire set of partial actions $\{S_n: n \geq 3\}$ by calculating

$$S_n[\bar{g}, h] = \frac{2}{n!} \left[\int d^4x h^{ab} \frac{\delta}{\delta \bar{g}^{ab}} \right]^{n-2} S_2[\bar{g}, h], \quad (37)$$

which is possible provided S_2 is known in a *neighborhood* of whichever particular background (a solution of (31)) we are interested in. Note that the first two partial actions do not contribute to the dynamics of h^{ab}: $S_0 = S[\bar{g}]$ is manifestly independent of h^{ab}, and S_1 vanishes once the background equation (31) has been enforced. We conclude, therefore, that S_2 contains all the information necessary to reconstruct the "dynamical" part of the action

$$S_{\rm dyn}[\bar{g}, h] \equiv \sum_{n=2}^{\infty} \lambda^n S_n[\bar{g}, h], \quad (38)$$

which itself contains all the dynamical information of the full action S. This is absolutely key to the calculations of Sec. II, in which we saw the first consequence of this reconstruction process, the recovery of the second-order equation of motion from an action that one would expect to encode only first-order dynamics.

A. Field equations

In general, we could let λ be a free parameter and, on demanding $\delta S[g]/\delta g^{ab} = 0$ for fixed $\bar{g}^{ab}$, derive a λ-dependent equation of motion $E_\lambda[\bar{g}, h] = 0$ for our dynamical field h^{ab}. Any h^{ab} that solved this equation would correspond to a metric $g^{ab} = \bar{g}^{ab} + \lambda h^{ab}$ that solved the field equations *exactly*.[16] However, if we are interested in approximating small variations of the metric (i.e. the limit $\lambda h^{ab} \to 0$) we can choose some order N to which we want the equation of motion to hold:

$$\frac{\delta S[g]}{\delta g^{ab}} = O(\lambda^{N+1}). \quad (39)$$

This is equivalent to

$$\frac{1}{\lambda} \frac{\delta S^{N+1}_{\rm dyn}[\bar{g}, h]}{\delta h^{ab}} = O(\lambda^{N+1}), \quad (40)$$

where $S^{N+1}_{\rm dyn}$ is defined by discarding from $S_{\rm dyn}$ those terms that can be neglected in (39):

$$S^{N+1}_{\rm dyn}[\bar{g}, h] \equiv \sum_{n=2}^{N+1} \lambda^n S_n[\bar{g}, h]. \quad (41)$$

We shall adopt this "Nth-order approximation" picture for the development of our formalism, as we can always write $N = \infty$ if we wish to discuss the exact theory.

For the sake of continuity with the previous section, we introduce the notation

$$\left. \frac{\delta S_2[\bar{g}, h]}{\delta h^{ab}} \right|_{\delta S[\bar{g}]/\delta \bar{g}^{ab} = 0} \equiv \kappa^{-1} \sqrt{-\bar{g}} \hat{G}_{abcd} h^{cd}, \quad (42)$$

where, because S_2 is second-order in h^{ab}, $\hat{G}_{abcd}$ will be a linear differential operator dependent only on $\bar{g}^{ab}$.[17] The equation of motion (40) now takes the form

$$\lambda \hat{G}_{abcd} h^{cd} = -\frac{\kappa}{\lambda \sqrt{-\bar{g}}} \frac{\delta}{\delta h^{ab}} \sum_{n=3}^{N+1} \lambda^n S_n[\bar{g}, h], \quad (43)$$

where it should be taken as given that terms $O(\lambda^{N+1})$ have been neglected. This is the Nth-order approximation to the equation of motion for h^{ab} that is consistent with the dynamics of g^{ab} prescribed by the action S. The first-order contribution has been separated from the sum so as to evoke the picture of a wave equation $\lambda \hat{G}_{abcd} h^{cd} = 0$ with a source. In the next section we will see that the source term on the right of (43) is indeed the energy-momentum tensor of the field h^{ab}, neglecting terms $O(\lambda^{N+1})$.

B. Energy-momentum tensor

First we shall demonstrate that the dynamical part of the action (38) can be generated from S_2 by a simple energy-momentum self-coupling procedure. Observe that, as a

[16] It is advisable to set $\lambda = 1$ before attempting to solve $E_\lambda[\bar{g}, h] = 0$, as this constant can always be absorbed into the magnitude of h^{ab}. Although this refinement was convenient for Sec. II, here we shall keep λ as it provides a simple method for tracking the powers of h^{ab} in expressions and is useful as a variable for differentiation.

[17] The operator $\hat{G}_{abcd}$ defined here coincides with the definition in (5) once $S = S_{\rm EH}$ has been fixed. This is shown in Sec. III E by deriving S_2.

consequence of (36), we have

$$S_n[\bar{g}, h] = \frac{1}{n}\int d^4x h^{ab}\frac{\delta S_{n-1}[\bar{g}, h]}{\delta \bar{g}^{ab}}. \tag{44}$$

Defining the nth partial energy-momentum tensor t^n_{ab} by applying Hilbert's prescription to the nth partial action,

$$t^n_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S_n[\bar{g}, h]}{\delta \bar{g}^{ab}}, \tag{45}$$

we conclude that

$$S_n[\bar{g}, h] = \frac{-1}{n}\int d^4x\sqrt{-\bar{g}}h^{ab}t^{n-1}_{ab}. \tag{46}$$

This makes manifest the energy-momentum self-coupling procedure that allows us to generate the dynamical part of the action (38) to arbitrary order, given only S_2. The nth partial action is nothing more than the integral of the contraction of h^{ab} with the energy-momentum tensor of the previous partial action (divided by $-n$). The dynamical part of the action is therefore given by

$$S^{N+1}_{\rm dyn}[\bar{g}, h] = \lambda^2 S_2[\bar{g}, h] - \int d^4x\sqrt{-\bar{g}}h^{ab}\sum_{n=2}^{N}\frac{\lambda^{n+1}t^n_{ab}}{n+1}. \tag{47}$$

Note that, for the particular case of general relativity ($S = S_{\rm EH}$), the background Eq. (7) also sets $S_0 = 0$, thus $S_{\rm dyn} = S_{\rm EH}$ (modulo surface terms) and the energy-momentum self-coupling procedure recovers the *entire* action of the full theory, not just the dynamical part.

Because of factors of $n+1$ dividing each t^n_{ab} in (47), it is not the case that in the action h^{ab} couples directly to its (Nth-order) total energy-momentum tensor, given by

$$T^N_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S^N_{\rm dyn}}{\delta \bar{g}^{ab}} = \sum_{n=2}^{N}\lambda^n t^n_{ab}. \tag{48}$$

Instead, the numerical denominators account for the $n+1$ factors of h^{ab} in $h^{ab}t^n_{ab}$, and ensure that the equations of motion do indeed have T^N_{ab} as the source. To prove this, note that for any symmetric field l^{ab} (vanishing on the boundary, or with compact support) we have

$$\begin{aligned}\int d^4x l^{ab}\frac{\delta S_n[\bar{g}, h]}{\delta h^{ab}} &= \int d^4x\frac{l^{ab}}{n!}\frac{\delta}{\delta h^{ab}}(\partial^n_\lambda S[\bar{g}+\lambda h])_{\lambda=0}\\ &= \frac{1}{n!}(\partial_\mu(\partial^n_\lambda S[\bar{g}+\lambda(h+\mu l)])_{\lambda=0})_{\mu=0}\\ &= \frac{1}{n!}(\partial^n_\lambda\partial_\mu S[\bar{g}+\lambda(h+\mu l)])_{\lambda=\mu=0}\\ &= \frac{1}{n!}(\partial^n_\lambda(\lambda\partial_\alpha S[\bar{g}+\lambda h+\alpha l]))_{\lambda=\alpha=0},\end{aligned}$$

where $\alpha \equiv \lambda\mu \Rightarrow \partial_\mu = \lambda\partial_\alpha$. Thus,

$$\begin{aligned}\int d^4x l^{ab}\frac{\delta S_n[\bar{g}, h]}{\delta h^{ab}} &= \frac{1}{n!}(\lambda\partial^n_\lambda\partial_\alpha S[\bar{g}+\lambda h+\alpha l]\\ &\quad + n\partial^{n-1}_\lambda\partial_\alpha S[\bar{g}+\lambda h+\alpha l])_{\lambda=\alpha=0}\\ &= \frac{1}{(n-1)!}(\partial_\alpha\partial^{n-1}_\lambda S[\bar{g}+\lambda h+\alpha l])_{\lambda=\alpha=0}\\ &= (\partial_\alpha S_{n-1}[\bar{g}+\alpha l, h])_{\alpha=0}\\ &= \int d^4x l^{ab}\frac{\delta S_{n-1}[\bar{g}, h]}{\delta \bar{g}^{ab}}.\end{aligned} \tag{49}$$

Hence we have the following important result:

$$\frac{\delta S_n[\bar{g}, h]}{\delta h^{ab}} = \frac{\delta S_{n-1}[\bar{g}, h]}{\delta \bar{g}^{ab}}. \tag{50}$$

Or, using definition (45),

$$\frac{\delta S_n[\bar{g}, h]}{\delta h^{ab}} = -\sqrt{-\bar{g}}t^{n-1}_{ab}. \tag{51}$$

Therefore the equation of motion (43) takes on the form

$$\lambda\hat{G}_{abcd}h^{cd} = \kappa\lambda^{-1}\sum_{n=3}^{N+1}\lambda^n t^{n-1}_{ab}, \tag{52}$$

or, recalling (48),

$$\lambda\hat{G}_{abcd}h^{cd} = \kappa T^N_{ab}. \tag{53}$$

We have derived the relation we sought, demonstrating that any metric theory of pure gravity can be formulated as a first-order wave equation with its own energy-momentum tensor as a source. For every $N \geq 1$, we can derive the equation of motion (53) by applying the variational principle to the action $S^{N+1}_{\rm dyn}$; the left-hand side is the wave equation for the linearized theory, and the right-hand side is the energy-momentum tensor prescribed by the action $S^N_{\rm dyn}$. This energy-momentum tensor is, to some extent, incomplete: it does not include the $O(\lambda^{N+1})$ contribution from the highest-order partial action S_{N+1}. This contribution could be calculated, if so desired, and added by hand to the field equations (53) so that the right-hand side read κT^{N+1}_{ab}, but this equation would no longer be a stationary configuration of the action $S^{N+1}_{\rm dyn}$. To remedy this, we could introduce a correction to the action $\lambda^{N+2}S_{N+2}$ that would generate the extra term in the equation of motion; the appropriate functional is given by (46) and couples h^{ab} to the highest-order partial energy-momentum tensor t^{N+1}_{ab}. But now once again the energy-momentum tensor T^{N+1}_{ab} is incomplete, and we can apply this same line of reasoning anew. So long as there is no N for which t^N_{ab} vanishes identically, this process can continue indefinitely, and as $N \to \infty$ the exact field equations are recovered, along with the action $S_{\rm dyn} = S - S_0 - \lambda S_1$.

All that remains is to connect our formalism to the specific results of the previous section. For the sake of completeness, however, we shall first discuss the gauge

symmetries of the theory, and deduce the conservation law for T^{N+1}_{ab}.

C. Gauge transformations

Because the action $S[g]$ is a coordinate-system independent integral, any diffeomorphism ϕ: $\mathcal{M} \to \mathcal{M}$ gives rise to a gauge transformation of the theory through the action of ϕ^*, the map comprising the pullback of ϕ on covector indices and the pushforward of ϕ^{-1} on vector indices:

$$S[\phi^* g] = S[g]. \tag{54}$$

Taylor expanding both sides about $\bar{g}^{ab}$ and applying the background equation reveals the gauge invariance of the dynamical part of the action:

$$S^{N+1}_{\rm dyn}[\bar{g}, h'] = S^{N+1}_{\rm dyn}[\bar{g}, h], \tag{55}$$

where

$$\lambda h'^{ab} \equiv \phi^* g^{ab} - \bar{g}^{ab}. \tag{56}$$

In the context of an Nth-order approximation, we must insist that $\phi^* = 1 + O(\lambda)$, otherwise these transformations will map the small metric fluctuations λh^{ab} onto fluctuations comparable in magnitude to $\bar{g}^{ab}$. We can write a general diffeomorphism of this form as $\phi^* = e^{\lambda \mathcal{L}_\xi}$, where $\mathcal{L}_\xi$ is the Lie derivative along a vector field $\xi^a = O(1)$. The gauge transformations of the theory are hence given by

$$h^{ab} \to h'^{ab} = h^{ab} + \delta h^{ab},$$
$$\delta h^{ab} \equiv \lambda^{-1} \sum_{n=1}^{N} \frac{(\lambda \mathcal{L}_\xi)^n}{n!} \bar{g}^{ab} + \sum_{n=1}^{N-1} \frac{(\lambda \mathcal{L}_\xi)^n}{n!} h^{ab}, \tag{57}$$

where we have discarded all terms $O(\lambda^N)$, as these will only contribute terms $O(\lambda^{N+1})$ to the equation of motion, and terms $O(\lambda^{N+2})$ to $S^{N+1}_{\rm dyn}$. If we wish we can let $\xi^a = \varepsilon^a$, an infinitesimal vector field, and derive the infinitesimal gauge transformation

$$\delta h^{ab} = \begin{cases} \mathcal{L}_\varepsilon(\bar{g}^{ab} + \lambda h^{ab}) & N \geq 2, \\ -2\bar{\nabla}^{(a}\varepsilon^{b)} & N = 1. \end{cases} \tag{58}$$

Because these gauge transformations (infinitesimal or otherwise) are symmetries of $S^{N+1}_{\rm dyn}$, they map solutions of the equation of motion (53) to other solutions. We can therefore use the equation of motion to deduce the transformation law for T^N_{ab}:

$$\delta T^N_{ab} \equiv T^N_{ab}[\bar{g}, h'] - T^N_{ab}[\bar{g}, h] = \frac{\lambda}{\kappa} \hat{G}_{abcd} \delta h^{cd}. \tag{59}$$

This verifies the earlier remark that the energy-momentum tensor is gauge dependent, except in the trivial case $N = 1$, for which $T^N_{ab} = 0$ by definition. It may come as a surprise that the energy-momentum tensor does not inherit the gauge invariance of the action from which it was derived. It should be stressed, however, that $S^{N+1}_{\rm dyn}$ is not *identically* gauge invariant: the relation (55) is only true when the background equation is obeyed. For general $\bar{g}^{ab}$, the diffeomorphism invariance of $S[g]$ only furnishes the gauge transformation law $\delta S^{N+1}_{\rm dyn} = -\lambda \delta S_1$, the right-hand side of which has a nonvanishing energy-momentum tensor responsible for the variation in T^N_{ab}. Equivalently, the gauge dependence of T^N_{ab} can be seen to result from the noncommutativity of gauge transformations and the functional derivative $\delta / \delta \bar{g}^{ab}$ used to define T^N_{ab} [13]; these operations would only commute if the gauge invariance of $S^{N+1}_{\rm dyn}$ extended to a neighborhood of the solutions of the background equation, rather than being confined to the solutions themselves.

D. Conservation law

It should be expected that $S^{N+1}_{\rm dyn}[\bar{g}, h]$ inherits the diffeomorphism invariance of $S[g]$, and that this symmetry endows the energy-momentum tensor with a covariant conservation law with respect to the background metric. The derivation proceeds in close analogy to the proof of $\nabla^a T^{\rm matter}_{ab} = 0$ from general relativity.

We again appeal to the diffeomorphism invariance of the action (54) but this time expand $S[g]$ about $\bar{g}^{ab}$ (a solution of the background equation) and $S[\phi^* g]$ about $\phi^* \bar{g}^{ab}$ (which will also be a solution). The result,

$$S^{N+1}_{\rm dyn}[\phi^* \bar{g}, \phi^* h] = S^{N+1}_{\rm dyn}[\bar{g}, h], \tag{60}$$

affirms that $S^{N+1}_{\rm dyn}$ is diffeomorphism invariant.[18] Now let ϕ be an infinitesimal diffeomorphism: $\phi^* = 1 + \mathcal{L}_\varepsilon$ for an arbitrary infinitesimal vector field ε^a with compact support. Then (60) becomes

$$0 = \int d^4x \left[\frac{\delta S^{N+1}_{\rm dyn}}{\delta \bar{g}^{ab}} \mathcal{L}_\varepsilon \bar{g}^{ab} + \frac{\delta S^{N+1}_{\rm dyn}}{\delta h^{ab}} \mathcal{L}_\varepsilon h^{ab} \right]. \tag{61}$$

Clearly the second term vanishes [to $O(\lambda^{N+1})$] if h^{ab} solves the equation of motion (53), and thus

$$0 = \int d^4x \frac{\delta S^{N+1}_{\rm dyn}}{\delta \bar{g}^{ab}} \bar{\nabla}^a \varepsilon^b + O(\lambda^{N+2})$$
$$= \int d^4x \sqrt{-\bar{g}} \varepsilon^b \bar{\nabla}^a T^{N+1}_{ab} + O(\lambda^{N+2}). \tag{62}$$

As this equation holds for any ε^a it follows that

$$\bar{\nabla}^a T^{N+1}_{ab} = 0 \tag{63}$$

is valid up to and including $O(\lambda^{N+1})$. Because this relation holds whenever h^{ab} solves its equation of motion, and because gauge transformations map solutions to solutions, the conservation law is gauge invariant.

[18] Note that diffeomorphism invariance is equivalent to being independent of coordinate system, and is a distinct property from gauge invariance as defined in Sec. III C.

It is important to recognize that (63) applies to the $(N+1)$th-order energy-momentum tensor: this is the highest-order approximation to the energy-momentum tensor that can be constructed from our truncated action $S_{\rm dyn}^{N+1}$, and is a better approximation than the tensor T_{ab}^{N} which features in the equations of motion appropriate to this order. Of course, the conservation law for T_{ab}^{N} follows from (63) by discarding the highest-order term, and ensures the consistency of the equation of motion (53) with the identity $\bar{\nabla}^a \hat{G}_{abcd} h^{cd} = 0$, which holds for all h^{ab} once the background equation has been enforced.

E. Constructing the graviton action

It is now time to close the circle of our discussion and connect the abstract formalism to our earlier calculation. We shall derive here the graviton action S_2, the ansatz of Sec. II, by applying the perturbative formalism to the particular case

$$S[g] = \frac{1}{\kappa}\int d^4x\sqrt{-g}R \equiv S_{\rm EH}[g], \tag{64}$$

the Einstein-Hilbert action. To proceed, we will use Eq. (36) to derive S_1, and then S_2, by successive functional derivatives $\delta/\delta\bar{g}^{ab}$ acting on $S_{\rm EH}[\bar{g}]$. The first derivative generates

$$S_1[\bar{g}, h] = \frac{1}{\kappa}\int d^4x\sqrt{-\bar{g}}\bar{G}_{ab}h^{ab}, \tag{65}$$

which of course vanishes for all h^{ab} when $\bar{g}^{ab}$ solves the background equation $\bar{G}_{ab} = 0$. A second variation in $\bar{g}^{ab}$ gives rise to

$$\delta S_1 = \frac{1}{\kappa}\int d^4x\sqrt{-\bar{g}}\Big[\delta\bar{R}_{ab}\Big(h^{ab} - \frac{1}{2}h\bar{g}^{ab}\Big) + \delta\bar{g}^{cd}\frac{1}{2}(h_{cd}\bar{R} - h\bar{R}_{cd} - \bar{g}_{cd}\bar{G}_{ab}h^{ab})\Big].$$

Replacing $\delta\bar{R}_{ab} \rightarrow \delta\bar{g}^{cd}\hat{R}_{cdab}$ in accordance with (24), we determine $\delta S_1/\delta\bar{g}^{ab}$ and assemble

$$\begin{aligned} S_2 &= \frac{1}{2}\int d^4x h^{cd}\frac{\delta S_1}{\delta\bar{g}^{cd}} \\ &= \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}\Big[h^{cd}\hat{R}_{cdab}\Big(h^{ab} - \frac{1}{2}h\bar{g}^{ab}\Big) \\ &\quad + \frac{1}{2}h^{cd}(h_{cd}\bar{R} - h\bar{R}_{cd} - \bar{g}_{cd}\bar{G}_{ab}h^{ab})\Big] \\ &= \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}(\hat{G}_{abcd} + \bar{H}_{abcd})h^{cd}. \end{aligned} \tag{66}$$

In the last line we referred to the definitions (5) and (6), and made use of the identity

$$\hat{R}_{abef}\Big(\delta^e_c\delta^f_d - \frac{1}{2}\bar{g}^{ef}\bar{g}_{cd}\Big) \equiv \hat{G}_{abcd}. \tag{67}$$

This completes the derivation of the graviton action (4) and confirms that it can be used as the starting point of an energy-momentum self-coupling procedure (46) that generates the Einstein field equations and the Einstein-Hilbert action (modulo surface terms) to arbitrary order.

The preceding calculation helps to reveal the advantage of using h^{ab}, a perturbation in the *inverse* metric, as our fundamental degree of freedom. Had we instead taken the usual approach, expanding $g_{ab} = \bar{g}_{ab} + \lambda\mathfrak{h}_{ab}$ and taking $\mathfrak{h}_{ab}$ as fundamental, the perturbative formalism would have unfolded identically but for the placement of indices. However, the calculation of S_2 from $S_{\rm EH}$ would have differed dramatically. The Lagrangian of S_1 would instead be proportional to $\bar{G}^{ab}\mathfrak{h}_{ab}$, and because the Ricci tensor is naturally covariant, the variation of $\bar{G}^{ab} = \bar{R}_{cd}\bar{g}^{ca}\bar{g}^{db} - \frac{1}{2}\bar{R}_{cd}\bar{g}^{cd}\bar{g}^{ab}$ under $\delta\bar{g}^{ab}$ would have been complicated by the extra two factors of $\bar{g}^{ab}$ on the first term, compared to the relevant tensor in our approach: $\bar{G}_{ab} = \bar{R}_{ab} - \frac{1}{2}\bar{R}_{cd}\bar{g}^{cd}\bar{g}_{ab}$. This trend continues at every order; the $\mathfrak{h}_{ab}$ convention leads to a greater proliferation of terms in each partial energy-momentum tensor because the Lagrangian of S_n has the form $(\bar{\nabla}_a)^2(\mathfrak{h}_{ab})^n$ so must be contracted with further $n+1$ factors of $\bar{g}^{ab}$ to render it a scalar.[19] Each of these metric factors generates a term in the partial energy-momentum tensor, and thus acts as compound interest for the process of energy-momentum self-coupling. In comparison, our convention leads to Lagrangians of the form $(\bar{\nabla}_a)^2(h^{ab})^n$, which only need only $n-1$ additional factors of $\bar{g}_{ab}$.[20] Clearly the inefficiency of the $\mathfrak{h}_{ab}$ approach stems from the natural covariance of derivative operators (∂_a or $\bar{\nabla}_a$) and curvature tensors; the advantages of the contravariant expansion $g^{ab} = \bar{g}^{ab} + h^{ab}$ are therefore not peculiar to the Einstein-Hilbert action, and are expected to be even more distinguished in higher derivative theories of gravity.

IV. MATTER

To avoid over-complicating our discussion, we have so far focused exclusively on *pure gravity*. Here we will go some way to remedy this simplification, and generalize the formalism of the previous section to include the perturbations of matter fields, and the effects of nonvacuum backgrounds.

In the most general case, let the action S be a functional of g^{ab} and a generic matter field Ψ^A, where A will serve as a placeholder for any number of internal or spacetime

[19] There are of course the instances of $\bar{g}^{ab}\partial_c\bar{g}_{de}$ in each $\bar{\nabla}_a$, but these occur equally in either convention.

[20] This does not mean that *all* terms in such a Lagrangian will contain only $n-1$ additional factors of $\bar{g}_{ab}$; there will often be cases in which $\bar{g}^{ab}$ is contracted with $(\bar{\nabla}_a)^2$ and thus $n+1$ factors of the metric (and its inverse) will be present. These cases only represent a small proportion of all possible terms, particularly as n becomes large, and are no worse than the terms afforded by the $\mathfrak{h}_{ab}$ convention.

indices. We then expand S about a background $(\bar{g}^{ab}, \bar{\Psi}^A)$ as follows:

$$g^{ab} = \bar{g}^{ab} + \lambda h^{ab}, \tag{68}$$

$$\Psi^A = \bar{\Psi}^A + \lambda \psi^A, \tag{69}$$

$$\Rightarrow S[g, \Psi] = \sum_{n=0}^{\infty} \lambda^n S_n[\bar{g}, h, \bar{\Psi}, \psi], \tag{70}$$

where $\bar{g}^{ab}$ and $\bar{\Psi}^A$ satisfy the background equations

$$\frac{\delta S[\bar{g}, \bar{\Psi}]}{\delta \bar{g}^{ab}} = 0, \qquad \frac{\delta S[\bar{g}, \bar{\Psi}]}{\delta \bar{\Psi}^A} = 0. \tag{71}$$

As before, each partial action can be calculated from the partial action at the previous order; with matter included, the appropriate recurrence relation is

$$S_n = \frac{-1}{n} \int d^4x \sqrt{-\bar{g}} (h^{ab} t^{n-1}_{ab} + \psi^A j^{n-1}_A), \tag{72}$$

where

$$t^n_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}} \frac{\delta S_n}{\delta \bar{g}^{ab}}, \qquad j^n_A \equiv \frac{-1}{\sqrt{-\bar{g}}} \frac{\delta S_n}{\delta \bar{\Psi}^A}. \tag{73}$$

There are two aspects of this coupling scheme that differ from pure gravity. The first is immediately apparent: the $h^{ab} t_{ab}$ term has been joined by an analogous coupling between matter fluctuations ψ^A and its "source current" j_A. The second difference is hidden within the definitions of t_{ab} and j_A; because the $\{S_n\}$ now represent the partial actions for gravity and matter together, $h^{ab} t_{ab}$ and $\psi^A j_A$ are no longer just self-couplings, and will in general contain terms coupling h^{ab} to ψ^A. In particular, t^n_{ab} should now be interpreted as the (nth-order) energy-momentum tensor due to *all* the fields: h^{ab}, ψ^A, and the background matter $\bar{\Psi}^A$.

Proceeding as before, we can now demand that the dynamical fields h^{ab} and ψ^A solve the field equations of the action $S^{N+1}_{\rm dyn} = \sum_{n=2}^{N+1} \lambda^n S_n$, and generate approximate solutions of the exact field equations (prescribed by S) accurate to $O(\lambda^N)$. Instead of using the definition (42) for $\hat{G}_{abcd}$, we write the general form of S_2, modulo surface terms, as

$$S_2 = \frac{1}{2} \int d^4x \sqrt{-\bar{g}} (h^{ab} \hat{G}_{abcd} h^{cd}/\kappa - 2 h^{ab} \hat{I}_{abA} \psi^A + \psi^A \hat{W}_{AB} \psi^B), \tag{74}$$

once the background equations (71) have been enforced. In the above equation, $\hat{G}_{abcd}$, $\hat{I}_{abA}$, and $\hat{W}_{AB}$ are linear operators that depend only on background fields, $\hat{G}_{abcd}$ and $\hat{W}_{AB}$ are self-conjugate, in the sense given by (11), and $\hat{I}_{abA}$ is conjugate to $\hat{I}^\dagger_{Aab}$:

$$\int d^4x \sqrt{-\bar{g}} A^{ab} \hat{I}_{abA} B^A = \int d^4x \sqrt{-\bar{g}} B^A \hat{I}^\dagger_{Aab} A^{ab}, \tag{75}$$

for all A^{ab} or B^{ab}, provided one has compact support. These definitions lead to equations of motion, accurate to $O(\lambda^N)$, as follows:

$$\lambda \hat{G}_{abcd} h^{cd} = \kappa T^N_{ab} + \lambda \kappa \hat{I}_{abA} \psi^A, \tag{76}$$

$$\lambda \hat{W}_{AB} \psi^B = J^N_A + \lambda \hat{I}^\dagger_{Aab} h^{ab}, \tag{77}$$

where

$$T^N_{ab} \equiv \sum_{n=2}^{N} \lambda^n t^n_{ab}, \qquad J^N_A \equiv \sum_{n=2}^{N} \lambda^n j^n_A. \tag{78}$$

Although this formalism is quite general, it is probably too general to be usefully employed. Indeed, the complications involved in describing matter as a background field *and* a dynamical perturbation generally serve to obscure the physical interpretation of the mathematics. An interesting example of this occurs when one tries to rederive $\bar{\nabla}^a T^{N+1}_{ab} = 0$ by applying the argument of Sec. III D. The result that now follows is

$$\bar{\nabla}^a T^{N+1}_{ab} = \frac{1}{2\sqrt{-\bar{g}}} \frac{\delta}{\delta \varepsilon^b} \int d^4x \sqrt{-\bar{g}} J^{N+1}_A \mathcal{L}_\varepsilon \bar{\Psi}^A, \tag{79}$$

the physical interpretation of which is far from clear. Rather than continue with this formulation in its full generality, it will therefore be more instructive to examine two special cases. First, we set $\bar{\Psi}^A = 0$ and consider small matter fields $\lambda \psi^A$ interacting with λh^{ab}. Second, by setting $\psi^A = 0$ we can study the effect of a background matter field $\bar{\Psi}^A$ on the propagation of the graviton. In principal, one could reach these special cases starting from the formalism we have just described, but it will be simpler and more illuminating to build them up from scratch.

A. Matter perturbations

In a region where the matter fields are small enough that their effects on spacetime curvature can be described by small perturbations λh^{ab} in the inverse metric, we can model the dynamics by taking $\bar{\Psi}^A = 0$, and describe the matter field using $\lambda \psi^A$ alone. As it is often the case for gravitational theories, let us suppose that the action S is the sum of a gravitational action S_g and a matter action S_Ψ:

$$S[g, \Psi] = S_g[g] + S_\Psi[g, \Psi]. \tag{80}$$

Moreover, for the sake of simplicity, we take Ψ^A to be a *free* field:

$$S_\Psi[g, \lambda\Psi] = \lambda^2 S_\Psi[g, \Psi] \quad \forall\ g^{ab}, \Psi^A. \tag{81}$$

This assumption will mean that the perturbative expansion of S can be described by an energy-momentum coupling procedure only. To see this explicitly, we expand the action about a background $(\bar{g}^{ab}, 0)$:

$$S[\bar{g}+\lambda h, \lambda\psi] = \sum_{n=0}^{\infty} \lambda^n (S_{gn}[\bar{g}, h] + S_{\Psi n}[\bar{g}, h, \psi]), \quad (82)$$

where each gravitational partial action

$$\begin{aligned} S_{gn}[\bar{g}, h] &= \frac{1}{n!}(\partial_\lambda^n S_g[\bar{g}+\lambda h])_{\lambda=0} \\ &= \frac{1}{n!}\left[\int d^4x h^{ab}\frac{\delta}{\delta \bar{g}^{ab}}\right]^n S_g[\bar{g}], \end{aligned} \quad (83)$$

much as before, and the matter partial actions

$$\begin{aligned} S_{\Psi n}[\bar{g}, h, \psi] &= \frac{1}{n!}(\partial_\lambda^n S_\Psi[\bar{g}+\lambda h, \lambda\psi])_{\lambda=0} \\ &= \frac{1}{n!}(\partial_\lambda^n(\lambda^2 S_\Psi[\bar{g}+\lambda h, \psi]))_{\lambda=0} \\ &= \frac{1}{(n-2)!}(\partial_\lambda^{n-2} S_\Psi[\bar{g}+\lambda h, \psi])_{\lambda=0} \\ &= \frac{1}{(n-2)!}\left[\int d^4x h^{ab}\frac{\delta}{\delta \bar{g}^{ab}}\right]^{n-2} S_\Psi[\bar{g}, \psi]. \end{aligned} \quad (84)$$

Defining the partial energy-momentum tensors for h^{ab} and ψ^A as

$$t^{gn}_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S_{gn}}{\delta \bar{g}^{ab}}, \qquad t^{\Psi n}_{ab} \equiv \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S_{\Psi n}}{\delta \bar{g}^{ab}}, \quad (85)$$

respectively, we see that the partial actions are coupled as

$$S_n[\bar{g}, h] = -\int d^4x\sqrt{-\bar{g}}h^{ab}\left(\frac{t^{gn-1}_{ab}}{n} + \frac{t^{\Psi n-1}_{ab}}{n-2}\right). \quad (86)$$

These partial actions lead to the Nth-order equations of motion

$$\lambda\hat{G}_{abcd}h^{cd} = \kappa T^N_{ab} = \sum_{n=2}^{N}\lambda^n(t^{gn}_{ab} + t^{\Psi n}_{ab}) \quad (87)$$

$$\lambda\hat{W}_{AB}\psi^B = \sum_{n=2}^{N}\left[\frac{-\lambda^n}{(n-1)\sqrt{-\bar{g}}}\frac{\delta}{\delta\psi^A}\int d^4x\sqrt{-\bar{g}}h^{ab}t^{\Psi n}_{ab}\right]. \quad (88)$$

The first equation confirms that the energy-momentum tensors of ψ^A and h^{ab} combine as the source for the graviton. The second equation describes how the coupling between h^{ab} and t^{Ψ}_{ab} acts as a source for ψ^A. Note that even when the matter field is not free, because S_Ψ never contains terms linear in the matter fields, $\hat{I}_{abA}$ must be at least linear in $\bar{\Psi}^A$, so we will always have $\hat{I}_{abA} = 0$ when $\bar{\Psi}^A = 0$.

B. Nonvacuum background

For a nonvacuum spacetime, we expect to be able to approximate (at least to first-order) the behavior of a gravitational perturbation by ignoring the perturbations in the matter field that it might induce. Alternatively, we may have in mind a particular nonvacuum solution of the field equations $(\bar{g}^{ab}, \bar{\Psi}^A)$ and wish to find nearby solutions (approximate or exact) with precisely the same matter content. For these two scenarios, we can set $\psi^A = 0$ and investigate the effect that the background $\bar{\Psi}^A$ has on the dynamics of h^{ab}.

Considerations of this nature highlight an interesting feature of our prior discussion of the graviton action. In Sec. II we saw the importance of a contribution to the action $h^{ab}H_{abcd}h^{ab}$ that vanished in the vacuum; the obvious question to ask is whether a similar term exists in the nonvacuum case, and whether or not it will vanish on the *nonvacuum* background equations. To answer these questions we will derive the graviton action for a nonvacuum background, which will also include the cosmological constant as a special case.

Let us restrict our attention to general relativity in the presence of a matter field:

$$S[g, \Psi] = S_{\mathrm{EH}}[g] + S_\Psi[g, \Psi], \quad (89)$$

$$S_\Psi[g, \Psi] \equiv 2\int d^4x\sqrt{-g}\mathcal{L}_\Psi(g^{ab}, \Psi^A, \partial_a\Psi^A). \quad (90)$$

The factor of 2 in the definition of the matter Lagrangian $\mathcal{L}_\Psi$ compensates for our slightly unusual normalization of S_{EH}.[21] It should be noted that we have assumed that $\mathcal{L}_\Psi$ does not depend on derivatives of the metric. This is the case for the Lagrangians of all the fields of the standard model except the spin-$\frac{1}{2}$ fermion, which in any case should be coupled to gravity using the vierbein formalism, e.g. [14]; such an approach is beyond the scope of this article. The results of this section can be generalized to allow $\mathcal{L}_m$ to depend on $\partial_c g^{ab}$ without any great difficulty, but this is an added algebraic complication that seems to add little insight to our investigation.

We proceed by expanding the action about a background $(\bar{g}^{ab}, \bar{\Psi}^A)$ just as in (68) and (69), but now, as $\psi^A = 0$, the coupling scheme (72) reverts to the familiar energy-momentum coupling of Sec. III. Following precisely the same method as Sec. III E, we can compute S_2 by two successive functional derivatives (with respect to $\bar{g}^{ab}$) applied to $S[\bar{g}, \bar{\Psi}]$. The first derivative yields

$$S_1 = \frac{1}{\kappa}\int d^4x\sqrt{-\bar{g}}(\bar{G}_{ab} - \kappa\bar{T}^{\Psi}_{ab})h^{ab}, \quad (91)$$

where

$$\bar{T}^{\Psi}_{ab} = \frac{-1}{\sqrt{-\bar{g}}}\frac{\delta S_\Psi[\bar{g}, \bar{\Psi}]}{\delta\bar{g}^{ab}} = -2\frac{\partial\bar{\mathcal{L}}_\Psi}{\partial\bar{g}^{ab}} + \bar{g}_{ab}\bar{\mathcal{L}}_\Psi \quad (92)$$

is the energy-momentum tensor of the background matter.

[21] All our actions are twice as large as the usual definition. This normalization has no effect on the classical equations of motion, but has allowed us to define the energy-momentum tensor without a factor of 2, simplifying the algebra of Secs. II and III.

The second derivative yields the graviton action:

$$S_2 = \frac{1}{2}\int d^4x h^{ab}\frac{\delta S_1}{\delta \bar{g}^{ab}}$$
$$= \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}\Big[h^{ab}\hat{G}_{abcd}h^{cd} - (\bar{G}_{ab} - \kappa\bar{T}^{\Psi}_{ab})h^{ab}h + 2\kappa h^{ab}h^{cd}\frac{\partial^2\bar{\mathcal{L}}_\Psi}{\partial\bar{g}^{ab}\partial\bar{g}^{cd}} + (\bar{R} + 2\kappa\bar{\mathcal{L}}_\Psi)\Big(\frac{1}{2}h_{ab}h^{ab} - \frac{1}{4}h^2\Big)\Big]. \quad (93)$$

This is the action we sought: the generalization of Eq. (4) to a nonvacuum background.

If we are only interested in the linear theory, and have no wish to calculate the energy-momentum tensor, then we are free to enforce the background equation

$$\bar{G}_{ab} = \kappa\bar{T}^{\Psi}_{ab}, \quad (94)$$

in the graviton action. In sharp contrast to the vacuum case, however, the background equation does not reduce S_2 to $\frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}h^{ab}\hat{G}_{abcd}h^{cd}$, or indeed any other covariantization of the massless spin-2 Fierz-Pauli action. Instead, it appears as though the background matter has endowed the graviton with mass:

$$S_2 = \frac{1}{2\kappa}\int d^4x\sqrt{-\bar{g}}(h^{ab}\hat{G}_{abcd}h^{cd} + \alpha), \quad (95)$$

where the "mass-term" α is given by

$$\alpha \equiv -\frac{1}{2}M\Big(h^{ab}h_{ab} - \frac{1}{2}h^2\Big) + N_{abcd}h^{ab}h^{cd}, \quad (96)$$

with

$$M \equiv 2\kappa\Big(\bar{\mathcal{L}}_\Psi - \bar{g}^{ab}\frac{\partial\bar{\mathcal{L}}_\Psi}{\partial\bar{g}^{ab}}\Big), \qquad N_{abcd} \equiv 2\kappa\frac{\partial^2\bar{\mathcal{L}}_\Psi}{\partial\bar{g}^{ab}\partial\bar{g}^{cd}}. \quad (97)$$

We refer to α as a mass-term because it is quadratic in h^{ab}, free from derivatives, and has been added to the kinetic term $h^{ab}\hat{G}_{abcd}h^{cd}$ in the Lagrangian. However, as we will see for the specific case of the cosmological constant, α does not by itself determine whether the graviton is actually *massive*, i.e. whether it propagates *subluminally*; the curvature of the background will play an equally important role in the field equations. In particular, while it is tempting to identify a mass m for the graviton according to $m^2 = M$ (at least when $N_{abcd} = 0$) we will soon see that the background matter often sets $M < 0$, so this idea is essentially untenable.

To explore these issues, it will be instructive to calculate α for a few simple examples. First, consider a scalar field background $\bar{\Phi}$ with Lagrangian

$$\bar{\mathcal{L}}_\Phi = -\frac{1}{2}\bar{g}^{ab}\partial_a\bar{\Phi}\partial_b\bar{\Phi} - V(\bar{\Phi}); \quad (98)$$

the mass-term is

$$\alpha_\Phi = \kappa V(\bar{\Phi})\Big(h_{ab}h^{ab} - \frac{1}{2}h^2\Big). \quad (99)$$

To ensure that the scalar field has positive energy density, we must insist that $V(\bar{\Phi}) \geq 0$; hence $M \leq 0$ as previously warned. Equation (99) can also be used to find the corresponding mass-term for a cosmological constant. In this case the Lagrangian is $\mathcal{L}_\Lambda = -\Lambda/\kappa$, which we can reach from $\mathcal{L}_\Phi$ by setting $\partial_a\bar{\Phi} = 0$ and $V = \Lambda/\kappa$. Clearly this gives

$$\alpha_\Lambda = \Lambda\Big(h_{ab}h^{ab} - \frac{1}{2}h^2\Big), \quad (100)$$

which similarly suffers from $M < 0$ if the cosmological constant is positive.

At this point, the reader may be suspicious that the formulas for α_Φ and α_Λ (with $M < 0$ and $N_{abcd} = 0$) signify that h^{ab} is a *tachyon* in the presence of a scalar field background or a cosmological constant. Indeed, if the background were flat and M constant over spacetime, we could derive the field equations from (95), observe that their divergence enforces the de Donder gauge condition

$$\partial^\alpha h_{\alpha\beta} - \frac{1}{2}\partial_\beta h = 0,$$

and, substituting this back into the equations of motion, conclude that the dynamics of the graviton were described by

$$(\partial^2 - M)h^{\alpha\beta} = 0.$$

This argument appears to justify the relation $m^2 = M$ for the graviton's mass, and motivate the conclusion that $M < 0$ betrays tachyonic behavior. It is important to realize, however, that the field equation above is of little relevance to the actual physical system we were discussing. In reality, M will not be constant, and the presence of background matter will inevitably preclude background flatness. To understand how this last consideration alters the dynamics of the graviton, we shall briefly examine the field equation for h^{ab} in the presence of a cosmological constant. First, we substitute (100) into (95) and derive the field equation

$$\hat{G}_{abcd}h^{cd} + \Lambda\Big(h_{ab} - \frac{1}{2}\bar{g}_{ab}h\Big) = 0. \quad (101)$$

In contrast to the naive approach, the covariant divergence of this equation vanishes identically, and so cannot be used to relate $\bar{\nabla}_b h$ and $\bar{\nabla}^a h_{ab}$. In place of this, the gauge invariance of the vacuum theory remains intact,[22] and the

[22]If we wish to extend our discussion of gauge invariance (Sec. III C) to include background matter *in general*, we would need to account for the gauge-fixing implicit in our starting assumption $\psi^A = 0$, which is obviously not preserved by a (first-order) infinitesimal diffeomorphism $\delta\psi^A = \mathcal{L}_\varepsilon\bar{\Psi}^A$. However, because Λ is constant over spacetime, no such difficulty arises here, and the transformations $\delta h^{ab} = -2\bar{\nabla}^{(a}\varepsilon^{b)}$ remain a symmetry of the equations of motion.

field equation may be simplified by setting $h = 0$, $\bar{\nabla}_a h^{ab} = 0$:

$$\bar{\nabla}^2 h_{ab} - 2\bar{R}_{dabc} h^{dc} = 0. \tag{102}$$

Surprisingly, the contribution from α_Λ has been canceled by a term proportional to the background Ricci tensor, resulting in a field equation that is identical in form to the first-order *vacuum* field equation (13) in this gauge. Of course, this does not indicate that the cosmological constant has no effect on the propagation of h^{ab}, only that these effects are limited to the constraints imposed on the background geometry by the background equation $\bar{R}_{ab} = \Lambda \bar{g}_{ab}$. For this reason, it does not seem particularly natural to interpret $2\bar{R}_{dabc} h^{dc}$ as endowing the graviton with a mass; Eq. (102) can instead be understood as a (partially gauge-fixed) massless spin-2 field equation that has been generalized to cosmological backgrounds. Quite aside from this, there is also the technical issue of interpreting the four-index tensor $\bar{R}_{abcd}$ as a mass: only if this tensor can be defined in terms of a single scalar variable (and the background metric) could the argument be made that this single variable described the graviton's mass. For a nonzero cosmological constant, the only background with this property is de Sitter space: $\bar{R}_{dabc} = \frac{\Lambda}{3}(\bar{g}_{db}\bar{g}_{ac} - \bar{g}_{dc}\bar{g}_{ab})$, thus the gauge-fixed field equation (102) becomes

$$\left(\bar{\nabla}^2 - \frac{2\Lambda}{3}\right) h^{ab} = 0. \tag{103}$$

If we were so inclined, we might interpret this as a field equation for a graviton with $m^2 = 2\Lambda/3$, and note that this relation has the *correct* sign for positive Λ, unlike the formula $m^2 = -2\Lambda$ suggested by our preliminary inspection of α_Λ. In truth, however, further investigation is needed before we can either adopt or discard this interpretation. This is not only because (102) [of which (103) is a special case] can be understood as a generalization of a massless field equation to cosmological backgrounds, but also because of the subtleties involved in interpreting the wave operator $\bar{\nabla}^2$ in curved space, and issues of whether or not to use a conformal coupling. Clearly, more work must be done to ascertain the physical ramifications of α_Λ, and the mass-term α in general, before we can understand the degree to which its effects can be thought of as giving mass to the graviton.

Although massive gravitons and the cosmological constant were historically viewed as entirely separate concepts, recent work has brought to light a number of interesting connections between the two. Deser and Waldron [15] have demonstrated that, in (anti-)de Sitter background spacetimes, a massive spin-2 field is stable if and only if $m^2 \geq 2\Lambda/3$, or $m = 0$. While it is intriguing that our de Sitter background field equation (103) suggests precisely the same special value of $m^2 = 2\Lambda/3$, Deser and Waldron's analysis differs significantly from our own, so this superficial observation may be misleading. In particular, whereas our mass-term arises as a direct result of the perturbative expansion, Deser and Waldron add their mass-term to the action *by hand*. Thus it is far from clear that the massive gravitons of their paper correspond to the physical system considered above. In contrast, Novello and Neves [16] claim to prove that $m^2 = -2\Lambda/3$, with the implication that $\Lambda \leq 0$. This approach considers an unusual generalization of the spin-2 field equation to curved backgrounds, making a nonstandard choice for the covariantization ambiguous term discussed in Sec. II A. Thus, while their calculations arguably describe a spin-2 field, this does not appear to be a natural way to describe the spin-2 field that results from perturbations of the metric (or its inverse) in Einstein's theory. It is our intention to disentangle the connections between these two approaches, and our own, in a later publication.

For the sake of completeness, we conclude this section with an example of a mass-term that can have $M > 0$, and $N_{abcd} \neq 0$. Unlike α_Λ, however, we shall not attempt to derive any of the implications for the equations of motion. Consider an electromagnetic 1-form background $\bar{A}_a$, with Lagrangian

$$\bar{\mathcal{L}}_A = -\frac{1}{4}\bar{F}^2 = -\frac{1}{4}\bar{g}^{ab}\bar{g}^{cd}\bar{F}_{ac}\bar{F}_{bd}, \tag{104}$$

and note that $\bar{F}_{ab} \equiv 2\partial_{[a}\bar{A}_{b]}$ is independent of the metric. The calculation yields

$$\alpha_A = -\frac{1}{4}\kappa\bar{F}^2\left(h_{ab}h^{ab} - \frac{1}{2}h^2\right) - \kappa h^{ab}h^{cd}\bar{F}_{ac}\bar{F}_{bd}, \tag{105}$$

which has the aforementioned properties.

V. CONCLUSION

Contrary to the prevailing maxim, coupling the classical Fierz-Pauli graviton to its own energy and momentum *does not* recreate general relativity order by order. However, there is an alternative action for the graviton (4) for which energy-momentum self-coupling *is* consistent with Einstein's theory. Using this action, the energy-momentum tensor of the graviton (26), added as a source to the graviton's first-order equation of motion (13), builds a field equation consistent with the Einstein equations to *second-order*. Furthermore, the perturbative formalism developed in Sec. III reveals that our action provides sufficient information to reconstruct general relativity to *arbitrary* accuracy: a simple recurrence relation (46) identifies the energy-momentum tensor at one order as the appropriate contribution to the action at the next. To any order N, this scheme assembles an action that dictates field equations (53) in which the graviton's Nth-order energy-momentum tensor is the source.

The formal machinery used to understand vacuum perturbations is easily extended to include matter, although the physical interpretation of the most general approach, in

which matter comprises both a background field and a small perturbation, is less than transparent. Focusing on matter perturbations separately from nonvacuum backgrounds serves to clarify the formalism significantly. In a vacuum background, the interactions between the graviton and perturbations of a free matter field lead to a field equation (87) in which the source for the graviton is the sum of gravitational and matter energy-momentum. This interaction inevitably induces a source in the field equations for matter (88). Alternatively, one may neglect matter perturbations and examine the consequences of a nonvacuum background. In this case, the dynamics and energy-momentum of the graviton are prescribed by the action (93), generalizing our previous ansatz. Surprisingly, the background matter appears to induce a mass-term in the graviton action, although it is currently unclear to what extent its interpretation as a mass is valid at the level of the field equations. The mass-terms induced by a scalar field (99), a cosmological constant (100) and electromagnetism (105) have been calculated.

ACKNOWLEDGMENTS

L. M. B. is supported by STFC and St. John's College, Cambridge. The MATHEMATICA package *Ricci* was used for the bookkeeping part of the calculations that produced Eqs. (26) and (B15). We thank Stanley Deser and Tomás Ortín for their helpful comments, and Thanu Padmanabhan for an enlightening discussion.

APPENDIX A: PADMANABHAN'S ANALYSIS

The recent article by Padmanabhan [8] unearths many significant shortcomings of the well-known arguments [1–4] that supposedly derive Einstein's equations by coupling the Fierz-Pauli graviton to its own energy-momentum tensor. Here we attempt to summarize his observations, and explain their relation to this present work.

In broad terms, Padmanabhan's criticisms fall into three areas:

(1) The Einstein-Hilbert action consists of a bulk term (the Γ^2 action) and a surface term. The latter includes a piece *linear* in $h_{\alpha\beta}$, so there can be no way to construct it from a self-coupling procedure that starts with an action that is already *quadratic* in $h_{\alpha\beta}$.[23]

(2) The starting point, the Fierz-Pauli Lagrangian (8), describes a *Lorentz invariant* field theory, and yet the end result, general relativity, is *generally covariant*. It is claimed that this metamorphosis only occurs because general covariance has been *assumed* in the various derivations, in which case it is "no big deal to obtain Einstein's theory." More generally, the classic bootstrapping arguments wield ideas developed in general relativity (such as Hilbert's definition of the energy-momentum tensor) or use knowledge of the end result to achieve their goal. Hence they cannot be regarded as a derivation of general relativity *from first principles*.

(3) The first-order field equation can only take a *symmetric* tensor as its source; the canonical energy-momentum tensor (15) is not necessarily symmetric, and although it can be made to be so, this process is not unique. Therefore the energy-momentum self-coupling procedure is ill-defined. The Hilbert definition *is* uniquely determined by the action, but to use it would violate criticism 2. *Crucially*, even if we allow ourselves to use Hilbert's definition, we still fail to recover the correct source-term for the second-order field equation.

It is to this very last crucial point that we have devoted the bulk of this paper. We now wish to explain our position with regards to the first two criticisms, and also Padmanabhan's proposed solution to the third.

Response to criticism 1.—Our approach expressly avoids discussing surface terms. This has greatly streamlined our formalism, and because such terms are completely irrelevant for determining field equations or energy-momentum tensors, the only price to pay for this simplicity is that we can only claim to reconstruct the Einstein-Hilbert action *modulo surface terms*.[24] In this sense, Padmanabhan's first criticism still stands, although it is unclear whether it has any great importance. If the action is an integral over the whole manifold, and asymptotic conditions apply to h^{ab} such that the surface term at infinity vanishes, then of course there is no distinction between the Einstein-Hilbert action and the action we have constructed. Even if the action is an integral over a manifold with a boundary, so long as we consider the action to be a functional over all fields with a particular boundary configuration (just as we might think of the action of a particle as a functional over all paths with particular end-points) the two actions differ only by an irrelevant constant. Besides, in situations where contributions from the boundary really are important, one does not typically use the Einstein-Hilbert action anyway: the Gibbons-Hawking-York boundary term [17,18] must be included to remove the dependence on second derivatives of the metric. This allows the field equations to be derived using a variational principle that only demands that the variation in the fields (and not also their derivatives) vanishes on the boundary.

[23]The argument given by Padmanabhan is phrased in terms of nonanalyticity in a dimensionful coupling constant. This form of the argument depends on his particular choice of normalization for $h_{\alpha\beta}$ and S_{EH}, but is essentially equivalent to the statement given here.

[24]Note that this does not necessarily mean that we have constructed the Γ^2 action, only that the integrand of the action differs from $\sqrt{-g}R$ by some total divergence.

Padmanabhan's major concern is that the surface term of the Einstein-Hilbert action has some quantum mechanical significance. As the nature of quantum gravity has yet to be understood, it remains unclear whether or not this is the case. We stress once again that the analysis in this paper is purely classical, and that we make no claims as to a quantum mechanical interpretation. Furthermore, it is not even known whether the graviton is a useful theoretical object for describing quantum gravity. We note again that the Gibbons-Hawking-York boundary term is usually included in quantum gravity investigations for which the boundary is not negligible.

Response to criticism 2.—It is our view that Padmanabhan's concerns about general covariance are unjustified: we take the position of Weinberg [19], that ''general covariance by itself is empty of physical content.'' Any theory (Lorentz invariant or not) can be expressed in arbitrary curvilinear coordinates, so the requirement of general covariance cannot, in and of itself, constrain the sort of theory one might construct. Rather, the kinematical content of general relativity is encapsulated by *the equivalence principle*, that the effect of gravity vanishes locally in an inertial coordinate system; thus expressing physical equations in coordinate invariant notation is an invaluable tool for describing how their dynamics are modified by gravity. It is possible that when Padmanabhan refers to ''general covariance'' he is referring to the equivalence principle also. As the latter is tantamount to identifying the gravitational field with a dynamical metric, he would certainly be correct to criticize any ''derivation'' that contained such a step; needless to say, we do not appeal to the equivalence principle in our approach.

General covariance aside, though, Padmanabhan's objection to the use of curved-space ideas is a valid one, indicating that none of the classic arguments constitute a derivation from first principles. Our approach certainly makes use of curved-space concepts; however our goals are perhaps not quite so bold as the other derivations that Padmanabhan has scrutinized: we do not pretend to derive general relativity purely from the ideas of Lorentz-invariant field theory. It should be stressed, however, that even if some of the *kinematical* content of general relativity is in some way assumed (curved spacetime, functional derivatives with respect to the metric, etc.) it is still a ''big deal'' to derive the *dynamical* content of the theory, Einstein's equations.

Response to criticism 3.—We have already explained our position with regards to the definition of the energy-momentum tensor in Sec. II C; the only reason that Hilbert's definition is unpalatable to Padmanabhan is that his aim is to start with as little curved-space mathematics as he can. However, the failure of the Hilbert energy-momentum tensor to give the correct second-order term for the Einstein field equations is a more significant stumbling block. We have explained our remedy, the use of a different starting action, in the body of this paper. Padmanabhan, on the other hand, eschews energy-momentum self-coupling and introduces a new object $S^{\alpha\beta}$ that he defines with the following algorithm. Start with a Lorentz invariant Lagrangian $\mathcal{L}(\eta_{\alpha\beta}, h_{\alpha\beta}, \partial_\gamma h_{\alpha\beta})$ expressed in Lorentzian coordinates $\{x^\alpha\}$. Replace every instance of $\eta_{\alpha\beta}$ with the metric $\bar{g}_{\alpha\beta}$ to produce a new Lagrangian $\tilde{\mathcal{L}}(\bar{g}_{\alpha\beta}, h_{\alpha\beta}, \partial_\gamma h_{\alpha\beta})$; note that this is *not* the same as expressing $\mathcal{L}$ in an arbitrary coordinate system because the partial derivatives ∂_α have not been upgraded to covariant derivatives $\bar{\nabla}_\alpha$. We can now define

$$S^{\alpha\beta} \equiv 2\frac{\partial\sqrt{-\bar{g}}\,\tilde{\mathcal{L}}}{\partial\bar{g}_{\alpha\beta}}\bigg|_{\bar{g}=\eta}. \tag{A1}$$

The subscript reminds us that we must set $\bar{g}_{\alpha\beta} = \eta_{\alpha\beta}$ after taking the metric derivative, as we are supposedly working in Lorentzian coordinates. Padmanabhan claims to be able to reconstruct the Γ^2 action by coupling $h_{\alpha\beta}$ to this new object $S^{\alpha\beta}$. Unfortunately $S^{\alpha\beta}$ has a number of highly undesirable properties, suggesting that it is a rather unnatural object, ill-defined in its current form.[25]

Firstly, as it has been constructed from a Lagrangian rather than an action, $S^{\alpha\beta}$ depends directly on surface terms. This introduces a very large ambiguity, as $S^{\alpha\beta}$ will depend on whether we write the integrand of the action in the form $(\partial h)^2$, as Padmanabhan does, in the form $h\partial^2 h$, or as some arbitrary combination of both. Each possibility defines a different $S^{\alpha\beta}$ and (presumably) leads to a different self-coupled limit for the graviton. It seems that the only remedy for this ambiguity is to artificially stipulate that $\mathcal{L}$ contain no second derivatives, although we note in passing that even this leaves us free to add surface terms of the form $\partial^\alpha(\phi A_\alpha)$ in theories for fields other than the graviton.

The second troubling aspect to $S^{\alpha\beta}$ is the ''half-covariantizing'' algorithm used to construct $\tilde{\mathcal{L}}$. It should be clear that this procedure has only been defined in Lorentzian coordinates, thus the matrix $S^{\alpha\beta}$ does not really constitute the components of a tensor, as we have not explained how their values change when expressed in another coordinate system.[26] There are essentially two ways to extend the definition (A1) to include curvilinear coordinates. The trivial solution is to construct the tensor $S^{ab} \equiv S^{\alpha\beta}(\partial_\alpha)^a(\partial_\beta)^b$ using the vectors $\{(\partial_\alpha)^a\}$, partial

[25]In private communication, Padmanabhan has indicated that he shares our concerns about $S^{\alpha\beta}$ and does not believe it to be of any fundamental importance; hence we present the case against $S^{\alpha\beta}$ for the sake of completeness rather than rebuttal.

[26]The insistence that we be able to calculate the components of this object in arbitrary coordinates has nothing to do with curved spacetime or general relativity. Rather, this reflects the perfectly reasonable expectation that we should be able to express Padmanabhan's self-coupling procedure in *flat-space* spherical polar coordinates, for example, or any other coordinate system we choose.

derivatives with respect to the Lorentzian coordinates used to calculate $S^{\alpha\beta}$ in the first place. This obviously defines a genuine tensor, so the components $S^{\alpha'\beta'}$ of S^{ab} in some curvilinear coordinate system $\{x^{\alpha'}\}$ can be calculated, and they will be related to $S^{\alpha\beta}$ by the usual transformation rules. It should be clear, however, that this solution is rather unnatural: suppose we have a Lagrangian expressed in a curvilinear coordinate system, then the only way to calculate the components $S^{\alpha'\beta'}$ in that system is to first transform to Lorentzian coordinates, calculate $S^{\alpha\beta}$ according to (A1), and then transform back to our original coordinate system. Also, because this process picks out a special set of coordinates, there is also no reason to expect that S^{ab} can be written as a tensorial function of h_{ab}, $\bar{g}_{ab}$, and $\bar{\nabla}_a$. The *natural* way to proceed would be to generalize the definition (A1) in such a way that we could calculate $S^{\alpha'\beta'}$ working in any coordinate system. It might seem that a viable solution would be to define the tensor

$$S^{ab} \equiv \frac{2}{\sqrt{-\bar{g}}} \frac{\partial \sqrt{-\bar{g}}\mathcal{L}}{\partial \bar{g}_{ab}} \bigg|_{\bar{\Gamma}}, \tag{A2}$$

where $\mathcal{L} = \mathcal{L}(\bar{g}_{ab}, h_{ab}, \bar{\nabla}_c h_{ab})$ is the *fully* covariant Lagrangian, and the subscript indicates that the Christoffel symbols $\bar{\Gamma}^a{}_{bc}$ are to be treated as independent of the metric and held constant in the derivative. This expression generalizes (A1) to define a tensor S^{ab} in a coordinate invariant fashion; because the Christoffel symbols are held constant, no term arises from a variation of the covariant derivatives, and S^{ab} will reduce to $S^{\alpha\beta}$ in Lorentzian coordinates. This expression gives us some insight into the geometrical meaning of Padmanabhan's half-covariantized algorithm; in particular, it reveals that the derivative $\partial/\partial\bar{g}_{\alpha\beta}$ used to define $S^{\alpha\beta}$ is in fact exploring geometries (infinitesimally close to Minkowski spacetime) with connections that are not metric compatible.[27] It is perhaps unsurprising that this $\bar{\Gamma}$-constant derivative introduces a new layer of ambiguity to the procedure, as we can now alter S^{ab} by adding terms proportional to $0 = \bar{\nabla}_c \bar{g}_{ab}$ to the Lagrangian. Although this might seem a rather contrived objection, it is in fact a very common consideration. For example, suppose the Lagrangian includes a term of the form $\bar{\nabla}_a h^a{}_b$; should we calculate S^{ab} by acting with $\partial/\partial\bar{g}|_{\bar{\Gamma}}$ on $\bar{\nabla}_a(\bar{g}^{ac}h_{cb})$, or should we first commute the metric past the covariant derivative, and act on $\bar{g}^{ac}\bar{\nabla}_a h_{cb}$ instead? Note that this issue would have been invisible in Lorentzian coordinates because

$$\frac{\partial \bar{\nabla}_c \bar{g}_{ef}}{\partial \bar{g}_{ab}} \bigg|_{\bar{\Gamma}} = -2\bar{\Gamma}^{(a}{}_{c(e}\delta^{b)}_{f)}, \tag{A3}$$

which we would have automatically set to zero. It seems the only way to avoid this uncertainty in S^{ab} is to introduce another artificial constraint on the Lagrangian: we insist that it be written in such a way that no derivatives act on the metric. This should be achieved by commuting covariant derivatives through the metric, rather than integrating by parts, due to the aforementioned issues with surface terms.

We shall take our analysis of $S^{\alpha\beta}$ no further at this time. It is still uncertain whether this object can be generalized, naturally and uniquely, to form a genuine tensor; without such a generalization it is difficult to ascertain what sort of mathematical object the matrix of functions $S^{\alpha\beta}$ is supposed to represent. Although we cannot claim to have exhausted all possibilities, the evidence before us suggests, at the very least, that this goal is not easily achieved.

Aside from these technical issues, we should also emphasize that, unlike the energy-momentum tensor, $S^{\alpha\beta}$ has no apparent physical interpretation beyond its supposed role in a graviton self-coupling scheme. Energy-momentum self-coupling was justified by analogy with matter-gravity coupling, and advanced by the notion that the energy-momentum of *all* fields should source gravitation. In contrast, the self-coupling scheme involving $S^{\alpha\beta}$ only serves to set gravity apart from the other fields. Furthermore, our solution displays an unusual symmetry between the coupling terms in the action and source terms generated in the field equations as a result (see Sec. III B); this symmetry is broken by Padmanabhan's self-coupling procedure.

APPENDIX B: EXPANSION OF G_{ab}

Here we determine the first two terms of the expansion of the Einstein tensor

$$G_{ab} = G^{(1)}_{ab} + G^{(2)}_{ab} + O(h^3), \tag{B1}$$

induced by a perturbation of the inverse metric about a vacuum background:

$$g^{ab} = \bar{g}^{ab} + h^{ab}, \tag{B2}$$

$$\bar{G}_{ab} = 0. \tag{B3}$$

The perturbation in the metric is of course fixed by the relationship $g^{ab}g_{bc} = \delta^a_c$,

$$\Rightarrow g_{ab} = \bar{g}_{ab} - h_{ab} + h_{ac}h^c{}_b + O(h^3). \tag{B4}$$

To begin, introduce a connection $E^a{}_{bc}$ between the derivative operators ∇_a and $\bar{\nabla}_a$:

$$E^a{}_{bc} = \frac{1}{2}g^{ab}(\bar{\nabla}_b g_{cd} + \bar{\nabla}_c g_{bd} - \bar{\nabla}_d g_{bc}). \tag{B5}$$

This allow the Ricci tensor to be expressed as

$$R_{ab} = 2(\bar{\nabla}_{[c}E^c{}_{a]b} + E^c{}_{d[c}E^d{}_{a]b}). \tag{B6}$$

From (B5) it is clear that

[27] This is the same operation as the derivative used to acquire the Einstein equations from the Palatini action [20], although here we will have no cause to perform the complementary derivative $\partial/\partial\Gamma|_{\bar{g}}$.

$$E^{a(0)}_{bc} = 0, \tag{B7}$$

$$E^{a(1)}_{bc} = -\frac{1}{2}\bar{g}^{ad}(2\bar{\nabla}_{(b}h_{c)d} - \bar{\nabla}_d h_{bc}), \tag{B8}$$

$$E^{a(2)}_{bc} = -\frac{1}{2}h^{ad}(2\bar{\nabla}_{(b}h_{c)d} - \bar{\nabla}_d h_{bc}) + \frac{1}{2}\bar{g}^{ad}(2\bar{\nabla}_{(b}(h_{c)e}h^e{}_d) - \bar{\nabla}_d(h_{be}h^e{}_c)). \tag{B9}$$

Hence the terms of the expansion $R_{ab} = R^{(1)}_{ab} + R^{(2)}_{ab} + O(h^3)$ can be computed as follows:

$$R^{(1)}_{ab} = 2\bar{\nabla}_{[c}E^{c(1)}_{a]b} \tag{B10}$$

$$R^{(2)}_{ab} = 2(\bar{\nabla}_{[c}E^{c(2)}_{a]b} + E^{c(1)}_{d[c}E^{d(1)}_{a]b}). \tag{B11}$$

Thus,

$$\begin{aligned} G^{(1)}_{ab} &= R^{(1)}_{ab} - \frac{1}{2}\bar{g}_{ab}R^{(1)}_{cd}\bar{g}^{cd} \\ &= -\bar{\nabla}_c\bar{\nabla}_{(a}h_{b)}{}^c + \frac{1}{2}\bar{\nabla}^2 h_{ab} + \frac{1}{2}\bar{\nabla}_a\bar{\nabla}_b h \\ &\quad - \frac{1}{2}\bar{g}_{ab}(-\bar{\nabla}_c\bar{\nabla}_d h^{cd} + \bar{\nabla}^2 h), \end{aligned} \tag{B12}$$

which confirms that $\hat{G}_{abcd}$, as defined in (5), represents the linearized Einstein tensor:

$$\hat{G}_{abcd}h^{cd} = G^{(1)}_{ab}. \tag{B13}$$

In particular, note that both sides of this equation agree on the order of the derivatives in $\bar{\nabla}_c\bar{\nabla}_{(a}h_{b)}{}^c$; this is the descendant of the covariantization ambiguous term discussed in Sec. II A.

To find $G^{(2)}_{ab}$, start with

$$G^{(2)}_{ab} = R^{(2)}_{ab} - \frac{1}{2}\bar{g}_{ab}(R^{(2)}_{cd}\bar{g}^{cd} + R^{(1)}_{cd}h^{cd}) + \frac{1}{2}h_{ab}R^{(1)}_{cd}\bar{g}^{cd}, \tag{B14}$$

and substitute Eqs. (B10) and (B11), followed by (B8) and (B9). The bookkeeping for this calculation is characteristically laborious, but is easily accomplished using a computer algebra package; the result is

$$G^{(2)}_{ab} = -\kappa t_{ab} + \frac{1}{2}h\hat{G}_{abcd}h^{cd}, \tag{B15}$$

where t_{ab} is given by (26). As expounded in Sec. II B, and now confirmed by direct calculation (B13), the first-order approximation to the Einstein field equation is $\hat{G}_{abcd}h^{cd} = 0$, so $\hat{G}_{abcd}h^{cd} = O(h^2)$ must hold true at second-order. Clearly it follows from this that $h\hat{G}_{abcd}h^{cd} = O(h^3)$, and hence (28) is verified.

The third-order difference between $G^{(2)}_{ab}$ and $-\kappa t_{ab}$ exists because the field equation approximated to second-order in (29) is actually $\sqrt{-g}G^{ab}/\sqrt{-\bar{g}} = 0$; this is of course entirely equivalent to the usual form of the Einstein field equation $G_{ab} = 0$.

[1] S. Deser, Gen. Relativ. Gravit. **1**, 9 (1970).
[2] R. P. Feynman, F. B. Morinigo, and W. G. Wagner, *Feynman Lectures on Gravitation* (Addison-Wesley, Reading, MA, 1995), pp. 74–88.
[3] S. N. Gupta, Phys. Rev. **96**, 1683 (1954).
[4] R. H. Kraichnan, Phys. Rev. **98**, 1118 (1955).
[5] S. Deser, Classical Quantum Gravity **4**, L99 (1987).
[6] D. G. Boulware and S. Deser, Ann. Phys. (N.Y.) **89**, 193 (1975).
[7] T. Ortin, *Gravity and Strings* (Cambridge University Press, Cambridge, England, 2004), Chap. 3.2.
[8] T. Padmanabhan, Int. J. Mod. Phys. D **17**, 367 (2008).
[9] M. Fierz and W. Pauli, Proc. R. Soc. A **173**, 211 (1939).
[10] R. M. Wald, *General Relativity* (University of Chicago, Chicago, 1984), p. 437.
[11] K. Kuchar, J. Math. Phys. (N.Y.) **17**, 801 (1976).
[12] P. D. Mannheim, Phys. Rev. D **74**, 024019 (2006).
[13] G. Magnano and L. M. Sokolowski, Classical Quantum Gravity **19**, 223 (2002).
[14] A. Lasenby, C. Doran, and S. Gull, Phil. Trans. R. Soc. A **356**, 487 (1998).
[15] S. Deser and A. Waldron, Phys. Lett. B **508**, 347 (2001).
[16] M. Novello and R. P. Neves, Classical Quantum Gravity **20**, L67 (2003).
[17] J. W. York, Phys. Rev. Lett. **28**, 1082 (1972).
[18] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2752 (1977).
[19] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), pp. 91–93.
[20] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby, *General Relativity: An Introduction for Physicists* (Cambridge University Press, Cambridge, England, 2006), Chap. 19.10.