International Journal of Modern Physics D
Vol. 17, Nos. 3 & 4 (2008) 367–398

FROM GRAVITONS TO GRAVITY: MYTHS AND REALITY

T. PADMANABHAN
IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India
nabhan@iucaa.ernet.in

Received 29 November 2007
Communicated by D. V. Ahluwalia-Khalilova

There is a general belief, reinforced by statements in standard textbooks, that: (i) one can obtain the full nonlinear Einstein theory of gravity by coupling a massless, spin 2 field h_{ab} self-consistently to the total energy–momentum tensor, including its own; (ii) this procedure is unique and leads to Einstein–Hilbert (EH) action; and (iii) it uses only standard concepts in Lorentz-invariant field theory and does not involve any geometrical assumptions. After providing several reasons why such beliefs are suspect — and critically re-examining several previous attempts — we provide a detailed analysis aimed at clarifying the situation. First, we prove that it is *impossible* to obtain the EH action, starting from the standard action for gravitons in linear theory and iterating repeatedly. This result follows from the fact that EH action has a part (viz. the surface term arising from second derivatives of the metric tensor) which is nonanalytic in the coupling constant, when expanded in terms of the graviton field. Thus, at best, one can only hope to obtain the remaining, quadratic, part of the EH Lagrangian (viz. the Γ^2 Lagrangian) if no additional assumptions are made. Second, we use the Taylor series expansion of the action for Einstein's theory, to identify the tensor $\mathcal{S}^{ab}$, to which the graviton field h_{ab} couples to the lowest order (through a term of the form $\mathcal{S}^{ab}h_{ab}$ in the Lagrangian). We show that the second rank tensor $\mathcal{S}^{ab}$ is *not* the conventional energy–momentum tensor T^{ab} of the graviton and provide an explanation for this feature. Third, we construct the full nonlinear Einstein theory with the source being a spin 0 field, a spin 1 field or relativistic particles by explicitly coupling the spin 2 field to this second rank tensor $\mathcal{S}^{ab}$ order by order and summing up the infinite series. Finally, we construct the theory obtained by self-consistently coupling h_{ab} to the conventional energy–momentum tensor T^{ab} order by order and show that this does *not* lead to Einstein's theory. The implications are discussed.

Keywords: Graviton; general relativity; Einstein–Hilbert action; spin-2 field.

1. Introduction and Motivation

1.1. *Conventional wisdom...*

The two classical fields — electromagnetism and gravity — are described by a vector field and a second rank symmetric tensor field, respectively. Considerations based on the Lorentz group suggest interpreting them (when suitable restrictions are imposed) as corresponding to massless spin 1 and spin 2 fields. The vector field

A_i couples to a conserved current J_i but does not contribute to this current. (That is, the photon does not carry charge.) In contrast, the tensor field is believed to be coupled to the energy–momentum tensor; since the field itself carries energy, it has to couple to itself in a nonlinear fashion. (The situation is similar to Yang–Mills fields which carry isotopic charge and hence are nonlinear.) It may, therefore, be possible to obtain a correct theory for gravity by starting with a massless spin 2 field h_{ab} coupled to the energy–momentum tensor T_{ab} of other matter sources to the lowest order, introducing self-coupling of h_{ab} to its own energy–momentum tensor at the next order and iterating the process. This will lead to a completely field-theoretic description of gravity in a Minkowski background and is conceptually quite attractive.

This attempt has a long history. The field equation for a free massless spin 2 field was originally obtained by Fierz and Pauli.[1] The first attempt to study the consequences of coupling this field to its own energy–momentum tensor seems to have been made by Kraichnan in unpublished work done in 1946–47. The first published attempt to derive the nonlinear coupling was by Gupta,[2] and Kraichnan published some of his results soon after.[3,4] Feynman provided a derivation[5] in his Caltech lectures on gravitation during 1962–63. The problem was readdressed through a clever technique by Deser.[6] (This problem and related ideas have been explored from several other points of view in the literature; see e.g. Refs. 7–12. We shall not discuss these approaches.) Virtually all these approaches claim to obtain not only Einstein's field equations but also the Einstein–Hilbert action.

1.2. *. . . And why it is suspect*

This result is widely quoted in the literature (see e.g. p. 424 of Ref. 13) and, at first sight, seems eminently reasonable. However, deeper examination raises several disturbing questions, if the result is really valid.

- In the conventional derivations, the final metric arises as $g_{ab} = \eta_{ab} + \lambda h_{ab}$, where $\lambda \propto \sqrt{G}$ has the dimension of length and h_{ab} has the correct dimension of $(\text{length})^{-1}$ in natural units, with $\hbar = c = 1$. The iteration is in powers of λ, starting from the zeroth order Lagrangian $L_0 \simeq (\partial h)^2$ for a spin 2 field, which has the dimension of $(\text{length})^{-4}$. (We have dropped the tensor indices to simplify the notation.) The final result in all the published works is the Einstein–Hilbert Lagrangian $L_{\text{EH}} = (1/4\lambda^2)R$. Since the scalar curvature has the structure $R \simeq (\partial g)^2 + \partial^2 g$, substitution of $g_{ab} = \eta_{ab} + \lambda h_{ab}$ gives to the lowest order
$$L_{\text{EH}} \propto \frac{1}{\lambda^2} R \simeq (\partial h)^2 + \frac{1}{\lambda}\partial^2 h. \tag{1}$$
Thus the full Einstein–Hilbert Lagrangian is nonanalytic in λ! *It will be quite surprising if, starting from $(\partial h)^2$ and doing a honest iteration on λ, one can obtain a piece which is nonanalytic in λ.* At best, one can hope to get the quadratic part of L_{EH} which gives rise to the Γ^2 action but not the four-divergence term involving $\partial^2 g$.

- To carry out this program, one needs to identify the energy–momentum tensor T^G_{ab} for the graviton field h_{ab} order by order in the coupling constant. At this stage, we are working in flat space–time, Cartesian coordinates [with metric $\eta_{ab} = \mathrm{dia}(-1, 1, 1, 1)$] with the Lorentz group as the invariance group. If we are honest (and do not use anything we learnt in our general relativity course!), we must provide a prescription to find T^G_{ab} within this context. There is indeed a natural conserved second rank tensor which arises from Lorentz symmetry, usually called the canonical energy–momentum tensor. This tensor, unfortunately, is not symmetric. It can be made symmetric but the procedure is not unique. For every choice of T^G_{ab} one can obtain a nonlinear theory clearly showing that further choices are to be made somewhere along the line.
- A sharper way of stating the above difficulty is the following: The same textbooks which assert that Einstein's theory can be obtained by coupling h_{ab} to itself self-consistently will also state in some other section (in Ref. 13, this happens on p. 467) that the gravitational field does not have a well-defined energy–momentum tensor! It will be rather strange if a unique energy–momentum tensor exists for the gravitational field order by order in the perturbation series but somehow "disappears" when all the terms are summed up. (The nonuniqueness of the energy–momentum tensor for Einstein's theory is well known and is extensively discussed in the literature; see e.g. Ref. 14–28.)
- In implementing this program, one needs to be clear whether general covariance is an *assumption* or a *result*. The starting point — Lorentz-invariant field theory in flat space–time with metric η_{ab} — has no notion of general covariance. If the source of the final equations is an energy–momentum tensor which is *assumed* to be generally covariant, it is equivalent to assuming that the left-hand side of the equations is generally covariant. It is then no big deal to obtain Einstein's theory, if we are prepared to *assume* general covariance. [It is sometimes claimed that the gauge invariance of a spin 2 field under $h_{ab}(x) \to h_{ab}(x) + \partial_a \xi_b(x) + \partial_b \xi_a(x)$ "becomes" the general covariance of the full theory. This is simply wrong; see the discussion around Eq. (32) in Sec. 3 below.]
- A term in the Lagrangian proportional to $\lambda h_{ab} T^{ab}$, where T^{ab} is due to *external* matter fields (assumed to be independent of h_{ab} to this order), will lead to the equation of motion of the type $\partial^2 h = \lambda T$. Hence a coupling of the type $\lambda h_{ab} T^{ab}$ is equivalent to requiring the source to be T_{ab}. Consider now the coupling of gravity to itself through a term of the type $\lambda h_{ab} \mathcal{S}^{ab}(h)$, where $\mathcal{S}^{ab}$ explicitly depends on the graviton field h_{ab}. When this term is varied with respect to h_{ab} to get the equations of motion, we will obtain *two* terms: $\mathcal{S}^{ab} + (\partial \mathcal{S}^{ij}/\partial h_{ab}) h_{ij}$, both of which will act as a source to gravity at the next order. If we want the *source* to be the energy–momentum tensor of the graviton field, T_{ab}, then the *coupling* cannot be of the form $h_{ab} T^{ab}(h)$ since this will lead to the wrong source. Thus we need to find out the form of the tensor $\mathcal{S}^{ab}$ — a question which does not seem to have attracted any attention in the literature. (We will see that $\mathcal{S}^{ab}$ is an interesting subject in its own right.)

None of the previous derivations addresses these issues, and most of them downplay the role of *assuming* general covariance. All these attempts make different tacit assumptions and it is difficult to judge which of these derivations can be thought of as being "from first principles" in the sense that it is completely independent of our knowledge of the end result. This difficulty becomes apparent when one follows the details of many of these derivations. The technology used is very strongly influenced by the known final result. For example, Kraichnan's pioneering work explicitly uses a term like $\eta_{ab}R^{ab}(\eta)$ (η_{ab} is the Minkowski metric and R^{ab} is the Ricci tensor) cleverly to obtain the result [see Eqs. (13)–(17) of Refs. 3 and 4] in spite of the fact that $R^{ab}(\eta)$ vanishes for the flat metric η_{ab}! It is impossible (at least for the author) to imagine that someone could have "guessed" this form for the action without knowing the result. Feynman's derivation also suffers from several shortcomings. To begin with, it is considerably less general than the one by Kraichnan, since Feynman assumes a particular form for the matter action and a coupling. But more relevant to our discussion is the manner in which he constructs the solutions to a consistency condition (see the discussion in Secs. 6.3 and 6.4 of Ref. 5). Since this approach *assumes* general covariance (in the form of $\nabla_a T^{ab} = 0$) and relies heavily on constructing generally covariant scalars, it is predestined to give Einstein–Hilbert action. The by-far-cleverest mathematical procedure is the one employed by Deser, in which he exploits the fact that, with a suitable choice of variables, the gravitational action becomes a cubic polynomial allowing the iteration to stop at a finite order. To achieve this mathematical economy, he has to start with the Palatini variational form [see his Eq. (2)] based on the Lagrangian $f^{ab}R_{ab}$, where $f^{ab} = \sqrt{-g}g^{ab}$ is the preferred variable. It is no surprise that he obtains $\sqrt{-g}R$ as the final result. (This is the *only* previous work that actually attempts an iteration; we shall comment on Deser's derivation in more detail in Sec. 6.3.)

In particular, *the fact that the action for the final theory contains the second derivatives of the field is always put in by hand.* Kraichnan's work has this explicitly; Feynman's derivation assumes a condition equivalent to general covariance to obtain solutions to a functional constraint and he explicitly chooses the "simplest" generally covariant scalar, thereby getting $\sqrt{-g}R$; Deser starts with an expression which is numerically the same as $\sqrt{-g}R$, but since he uses the first order form of the variation, the question of *second* derivatives is not directly applicable until, of course, the final result is obtained. Thus, we really do not know whether the original program (of coupling the field to its own energy–momentum tensor order by order and iterating the result) can be made to yield any other form of action principle even when the field equations are the same.

As an aside, we will comment on a related issue of obtaining a classical field theory from a quantum formulation. Of the two classical fields in nature, electromagnetic and gravitational, the former is simpler in structure because the vector field A_k couples to the current J_k but the photon does not carry a charge. Therefore the theory is linear with no self-coupling at the tree level and can be easily obtained as the lowest order limit of quantum electrodynamics. In fact one can do better in this case; a summation of the relevant Feynaman diagrams in an increasing number of loops, for example, will lead to the Euler–Heisenberg effective Lagrangian[29–33] containing quantum corrections to Maxwell equations. The situation in Yang–Mills theory is closer to gravity because the field has nonzero isospin current leading to self-coupling. But, unfortunately, the structure of the action functionals in YM theory and gravity is quite different, therefore preventing us from exploiting our knowledge in the case of the YM field. The conventional Lagrangian for the YM field is quadratic in curvature while the action in gravity is linear in curvature; which, of course, is possible only because of the existence of a metric in the case of gravity. For the YM field we have only connection and curvature.

One may be tempted to argue that most of the issues and objections raised above are irrelevant in the strictly classical context. In classical general relativity, one could argue that what matters is the equations of motion and not the action functional. This, however, is a rather restricted point of view and one needs to realize that the true world is quantum-mechanical and if one can gain insight into the nature of quantum theory from the structure of classical action functional, it is worth exploring. Of course, quantum theory has taught us that action functionals are as important (if not more) as the field equations. In the case of gravity, there are two action functionals which are of primary relevance. The first is the Einstein–Hilbert action which uses the Lagrangian $R\sqrt{-g}$, and the second is the Γ^2 action involving only the squares of the first derivatives. It can be shown that these two actions can be thought of as providing the momentum representation and the coordinate representation (respectively) of the theory and differ by a surface term which is directly related to the entropy of horizons in the semiclassical theory and has been the basis of a series of investigations.[34–44] Since the existence of horizons is probably the most remarkable feature of classical gravity that could serve as a link with the quantum description of space–time, it is important to try and understand whether this surface term can arise from the spin 2 field approach.

1.3. *Plan of the paper*

We will try to address these issues in as straightforward ("dumb") a manner as possible. Section 2 reviews the background material related to a spin 2 field, and a few important results, needed later, are obtained. In particular, we introduce a new second rank tensor $\mathcal{S}^{ab}$ associated with any matter Lagrangian which can be obtained by a well-defined procedure. This tensor, in general, is different from the standard energy–momentum tensor T^{ab} but coincides with the energy–momentum tensor for a relativistic particle, spin 0 field or spin 1 field.

In Sec. 3, we start with an action for Einstein gravity for the metric $g_{ab} = \eta_{ab} + \lambda h_{ab}$, expand it in a functional Taylor series in h_{ab} and determine the form of the self-coupling at the lowest order. We show that, to the lowest nontrivial order, the coupling is of the form $h^{ab}\mathcal{S}_{ab}$, where $\mathcal{S}_{ab}$ is the quantity introduced in Sec. 2.

We also exhibit the nonanalytic nature of the Einstein–Hilbert Lagrangian in λ and prove that the Lagrangian can never be obtained by an iteration in λ.

In Sec. 4 we provide a general procedure for coupling the field h_{ab} self-consistently to *any* second rank tensor which can be expressed as a functional derivative of matter action. This leads to a well-defined "rule" for coupling the field h_{ab} to matter fields. We first use it with the tensor $\mathcal{S}^{ab}$ we have defined in Sec. 2 and show that it leads to a generally covariant Lagrangian for a relativistic particle, spin 0 field or spin 1 field but not in general. We then use the same prescription to couple h_{ab} to itself and show that *the resulting theory is Einstein's theory* (Sec. 4.1). Thus, at least in the limited case of a spin 2 field interacting with a relativistic particle, spin 0 a field or spin 1 field [the only cases in which we have any observational evidence for gravitational theory!], we have an iterative procedure for obtaining the full theory when the self-coupling of h_{ab} is *not* to the energy–momentum tensor. In Sec. 4.2 we repeat the analysis by coupling h_{ab} using the standard energy–momentum tensor T^{ab}. In the case of *all* matter fields, this leads to the standard generally covariant action. But when we use this prescription for coupling h_{ab} to itself, we do *not* get Einstein's theory but a more complicated one which explicitly depends on the background Lorentzian metric or the field h_{ab}. We shall also show (in Sec. 4.3) that previous results, when properly analyzed, agree with our claims.

The analysis in this paper goes contrary to the conventional wisdom, and Sec. 5 discusses the issues which arise from this work.

2. Action and Energy–Momentum Tensor for the Spin 2 Field

In this section we will collect together the results which are required later. (A more pedagogical description is provided in App. A; this may be useful since the results are somewhat scattered in the literature.[45–50] The action for the noninteracting, massless, spin 2 field h_{ab} is built out of scalars which are quadratic in the derivatives $\partial_a h_{bc}$. The most general expression will be the sum of different scalars obtained by contracting pairs of indices in $\partial_a h_{bc}\partial_i h_{jk}$ in a different manner. If we assume that the field equations should be invariant under the gauge transformation

$$h_{ab}(x) \to h_{ab}(x) + \partial_a \xi_b(x) + \partial_b \xi_a(x), \tag{2}$$

then the resulting expression for the quadratic part of the action can be written in different, equivalent forms:

$$\begin{aligned} A_h &= \frac{1}{4}\int d^4x \partial_a h_{bc}\partial_i h_{jk}\left[\eta^{ai}\eta^{bc}\eta^{jk} - \eta^{ai}\eta^{bj}\eta^{ck} + 2\eta^{ak}\eta^{bj}\eta^{ci} - 2\eta^{ak}\eta^{bc}\eta^{ij}\right] \\ &= \frac{1}{4}\int d^4x[\partial_i h^a_a \partial^i h^j_j - \partial_a h_{bc}\partial^a h^{bc} + 2\partial_a h_{bc}\partial^c h^{ba} - 2\partial_a h^b_b \partial_i h^{ia}] \\ &= \frac{1}{4}\int d^4x\left[\frac{1}{2}\partial_i \bar{h}^a_a \partial^i \bar{h}^j_j - \partial_a \bar{h}_{bc}\partial^a \bar{h}^{bc} + 2\partial_a \bar{h}_{bc}\partial^c \bar{h}^{ba}\right]; \quad \bar{h}_{ab} \equiv h_{ab} - \frac{1}{2}\eta_{ab}h^i_i. \end{aligned} \tag{3}$$

[If we assume that the action is quadratic in the first derivatives and gauge-invariant, its form is uniquely given by the above equation, except for one *very specific* four-divergence term which can be added. This is discussed in App. A; see Eq. (A.4).] We shall use the more compact notation

$$A_h = \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk} M^{abcijk}(\eta^{mn}), \tag{4}$$

where the tensor $M^{abcijk}(\eta^{mn})$ is symmetric in bc, jk and under the triple exchange $(a,b,c) \leftrightarrow (i,j,k)$ and is given by

$$M^{abcijk}(\eta^{mn}) = \left[\eta^{ai}\eta^{bc}\eta^{jk} - \eta^{ai}\eta^{bj}\eta^{ck} + 2\eta^{ak}\eta^{bj}\eta^{ci} - 2\eta^{ak}\eta^{bc}\eta^{ij}\right]_{\text{symm}}, \tag{5}$$

where the subscript "symm" indicates that the expression inside the square brackets should be suitably symmetrized in bc, jk and under the exchange $(a,b,c) \leftrightarrow (i,j,k)$. In the expression for the action, since M^{abcijk} is multiplied by $\partial_a h_{bc}\partial_i h_{jk}$, we need not worry about symmetrization and use the expression given inside the square brackets as it is. The gauge invariance of the action leads to the *identity*

$$M^{abcijk}\partial_b\partial_a\partial_i h_{jk} = 0. \tag{6}$$

To the lowest order, we can couple h_{ab} to other fields by adding an interaction Lagrangian of the form $(\lambda/2)T^{ab}h_{ab}$, where T^{ab} is some tensor built out of the matter variables and λ is a coupling constant. The total action will be

$$A_{\text{tot}} = \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk} M^{abcijk}(\eta^{mn}) + \frac{\lambda}{2}\int d^4x h^{ab}T_{ab} + A_{\text{matter}}. \tag{7}$$

Obviously, only the symmetric part of T^{ab} is relevant to this coupling. The variation of h_{ab} will now lead to the field equation $M^{abcijk}\partial_a\partial_i h_{jk} = \lambda T^{bc}$. The condition in Eq. (6) now implies that $\partial_a T^{ab} = 0$. Thus the field described by the action in Eq. (3) can only be sourced by a conserved, symmetric part of a second rank tensor. The above fact — which is, of course, fairly standard — shows the intimate connection between the conservation of the source and the gauge invariance of the field.

It is this conservation law, $\partial_a T^{ab} = 0$, which leads to an inconsistency if we assume that T^{ab} is the standard expression for the energy–momentum tensor for matter fields. When the *matter variables* are varied, the equation of motion will now be affected by h_{ab} because of the $(\lambda/2)T^{ab}h_{ab}$ coupling. But the condition $\partial_a T^{ab} = 0$ is equivalent to the equations of motion for the matter field, when it is *unaffected* by h_{ab}. Hence, in general, it will not be possible to satisfy both these conditions and find a consistent set of solutions. The conventional wisdom is to attempt to find a consistent theory in which the field equations for h_{ab} should imply not the condition $\partial_a T^{ab} = 0$ but a modified one of the form $\partial_a(T^{ab} + t^{ab}) = 0$, where t_{ab} is the energy–momentum tensor for the spin 2 field. This will require coupling the field to its own energy–momentum tensor recursively and the hope is to show that — when the recursion is carried out to infinite orders — the resulting theory will be Einstein's gravity. This brings us to the question of defining the energy–momentum tensor for the spin 2 field.

For any system described by a Lorentz–invariant Lagrangian $L(\phi_A, \partial_a\phi_A)$, where ϕ_A denotes a generic matter field with A representing possible tensor indices, one can show that

$$\partial_b\left[\partial_a\phi_A\left(\frac{\partial L}{\partial(\partial_b\phi_A)}\right) - \delta^b_a L\right] = 0 \tag{8}$$

when the equation of motion is satisfied.[51] This allows us to define an infinite number of conserved second rank tensors of the form

$$T^{ba} \equiv \left[\partial^a\phi_A\left(\frac{\partial L}{\partial(\partial_b\phi_A)}\right) - \eta^{ab}L\right] + \partial_c\psi^{cba}, \tag{9}$$

where $\psi^{cba} = -\psi^{bca}$ is an arbitrary third rank tensor antisymmetric in the first two indices, so that $\partial_c\partial_b\psi^{cba} = 0$ identically. It is possible to choose ψ^{cba} in an infinite number of ways and still ensure that T^{ba} is symmetric. Thus Lorentz-invariant field theories possess an infinite number of conserved symmetric second rank tensors, any of which can be legitimately thought of as an energy–momentum tensor. For the spin 2 field this prescription gives

$$T^{pq} = \frac{1}{2}M^{pbcijk}\partial_i h_{jk}\partial^q h_{bc} - \eta^{pq}L + \partial_c\psi^{cpq}. \tag{10}$$

This nonuniqueness shows that it is not possible to proceed further without making extra assumptions regarding the form of T^{pq}.

One needs to be clear about the *different* kinds of ambiguity in the definition of T^{pq}. The first ambiguity is in the choice of ψ^{cpq}. The second ambiguity has to do with the fact that we can add to our Lagrangian a total divergence with a undetermined coefficient [as shown in Eq. (A.4) of App. A]. This changes the form of T^{pq}. Third, the T^{pq} defined in Eq. (10) is *not* gauge-invariant. In fact, one can prove a general theorem[52] that the energy–momentum tensor for the spin 2 field cannot be made gauge-invariant for any choice of ψ^{cpq}. This raises serious questions about whether the resulting theory after infinite iteration will possess any trace of the original gauge symmetry. If one is to be honest, in the sense that *no structures other than those sanctioned by Lorentz-invariant field theory are to be used*, then it is not possible to proceed any further and obtain a *unique* nonlinear theory.

Let us, however, reduce the standards of honesty and introduce another definition of a second rank symmetric tensor (which is based on what we learnt in our general relativity course, but we will not mention it!) along the following lines: We note that the Lagrangian for any field also depends on the Lorentz metric η^{ab}, i.e. $L = L(\phi_A, \partial_a\phi_A, \eta_{ab})$. The functional derivative of the action A_{matter} with respect to η_{ab} will define a symmetric, second rank tensor which we can attempt to use. But since $\eta_{ab} = \text{dia}(-1, 1, 1, 1)$ is a constant, $(\delta A/\delta\eta_{ab})$ is mathematically ill-defined and the functional derivative actually depends on the rule for its definition. We shall see below that several rules are possible, but let us first consider the conventional wisdom again.

We begin by noting that, even though we are in flat space–time, we can use any set of coordinates to describe the physics. Let us assume that, in a curvilinear coordinate system we choose, the space–time metric is $\gamma_{ab}(x)$. We will further assume that the action in the curvilinear coordinates is obtained with replacing η_{ab} with γ_{ab}, ordinary derivatives into covariant derivatives and changing the volume element from d^4x to $d^4x\sqrt{-\gamma}$. Thus the action has a kinematic dependence on γ_{ab}, which we shall explicitly exhibit by writing it as

$$A(\phi_A, \partial\phi_A, \eta_{ab}) \to A_\nabla(\phi_A, \nabla\phi_A, \gamma_{ab}) = \int d^4x\sqrt{-\gamma}\, L_\nabla(\phi_A, \nabla\phi_A, \gamma_{ab}). \tag{11}$$

(The subscript ∇ in A_∇ is to remind ourselves that, in obtaining this action, ordinary derivatives have been changed to covariant derivatives; this will turn out to be important later on.) It is now possible to obtain a second rank symmetric tensor T^{ab} by taking the functional derivative of the action with respect to γ_{ab} and then setting $\gamma_{ab} = \eta_{ab}$:

$$\delta A_\nabla = \frac{1}{2}\int d^4x\,\sqrt{-\gamma}\,T^{ab}\delta\gamma_{ab}, \quad T^{ab}(x) \equiv \left[\frac{2}{\sqrt{-\gamma}}\frac{\delta A_\nabla}{\delta\gamma_{ab}(x)}\right]_{\gamma=\eta}. \tag{12}$$

More explicitly, this leads to the energy–momentum tensor

$$T^{ab}(x) = \left[\frac{2}{\sqrt{-\gamma}}\left\{\frac{\partial L\sqrt{-\gamma}}{\partial\gamma_{ab}} - \partial_c\left(\frac{\partial L\sqrt{-\gamma}}{\partial(\partial_c\gamma_{ab})}\right)\right\}\right]_{\gamma=\eta}. \tag{13}$$

The procedure described above provides one possible prescription for obtaining T^{ab} in flat space–time. Note that γ_{ab} for us is purely a bookkeeping device and, at the end of the calculations, we shall set $\gamma_{ab} = \eta_{ab}$.

It is rather surprising that this definition of T^{ab} is routinely used in field-theoretic approaches to gravity (for example, in Ref. 5) as though it had nothing to do with curved space–time. *This attitude is incorrect.* The variation of γ_{ab} to $\gamma_{ab} + \delta\gamma_{ab}$ for arbitrary choices of $\delta\gamma_{ab}$ takes one from flat space–time in curvilinear coordinates to genuine curved space–times. (Just varying the coordinates in flat space–time will only have four function degrees of freedom, while we need 10.) The evaluation of the functional derivative in Eq. (13) requires the *strong* assumption that the action in Eq. (11) is valid in arbitrary curved space–time with metric γ_{ab}. We have come a long way from the basic concepts of Lorentz-invariant quantum field theory in flat space–time.

To use this definition for the spin 2 field, we first write the action in Eq. (3) in arbitrary curvilinear coordinates with metric γ_{ab}, using our rule in Eq. (11), as

$$\begin{aligned} A_\nabla &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\nabla_a h_{bc}\nabla_i h_{jk}\,M^{abcijk}(\gamma^{mn}) \\ &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\nabla_a h_{bc}\nabla_i h_{jk} \\ &\quad\times\left[\gamma^{ai}\gamma^{bc}\gamma^{jk} - \gamma^{ai}\gamma^{bj}\gamma^{ck} + 2\gamma^{ak}\gamma^{bj}\gamma^{ci} - 2\gamma^{ak}\gamma^{bc}\gamma^{ij}\right]. \end{aligned} \tag{14}$$

In this expression, the covariant derivative operator ∇ is defined with respect to the metric γ_{ab} and involves the first derivatives $\partial_a \gamma_{bc}$. Varying this action with respect to γ_{ab} will lead to a symmetric energy–momentum tensor for the spin 2 field when we use the prescription in Eq. (13). The actual expression for this tensor is fairly complicated but — fortunately — we do not need it. We, however, stress the following fact: since $[\partial L/\partial(\partial_c \gamma_{ab})]$ involves first derivatives of h_{ab}, the tensor T^{ab} will involve *second derivatives* of h_{ab}. It can again be shown by detailed algebra that this T^{ab} is indeed of the form in Eq. (10) for a specific choice of ψ^{cpq}; thus, our rule uses one of many choices in Eq. (10). We can now attempt to obtain the nonlinear theory by coupling h_{ab} iteratively to the energy–momentum tensor defined by Eq. (13). We shall show in Sec. 4.2 that *contrary to popular belief, the resulting theory is not Einstein's theory.*

There is, however, a more important issue which needs to be raised as regards the procedure used to obtain Eq. (13). What we have done is essentially to introduce the curvilinear metric γ_{ab} into the matter action (which was originally defined in flat space–time Cartesian coordinates) *by a particular rule* and then evaluate the functional derivative in Eq. (13). At the end of the calculation, we set $\gamma_{ab} \to \eta_{ab}$. The rule in Eq. (11) is strongly motivated by general covariance and, of course, leads to a generally covariant matter action in the curvilinear coordinates. But since we do *not* have the right to assume general covariance (and only Lorentz invariance), the rule we have specified is only one of many possible ways of introducing γ_{ab} into the matter action.

To bring this point sharply into focus and to derive some important consequences in the coming sections, we shall introduce another rule which leads to the definition of *another* symmetric second rank tensor. To do this, we will construct a modified action in the curvilinear coordinates by replacing η_{ab} with γ_{ab} and changing the volume element from d^4x to $d^4x\sqrt{-\gamma}$, *but without* changing ordinary derivatives into covariant derivatives. The action again acquires a kinematic dependence on γ_{ab}, which we shall explicitly exhibit by writing it as

$$A(\phi_A, \partial\phi_A, \eta_{ab}) \to A_\partial(\phi_A, \partial\phi_A, \gamma_{ab}) = \int d^4x\sqrt{-\gamma}\, L_\partial(\phi_A, \partial\phi_A, \gamma_{ab}). \tag{15}$$

(The subscript ∂ in A_∂ is to remind ourselves that, in obtaining this action, ordinary derivatives are retained as they were.) We can again obtain a second rank symmetric tensor $\mathcal{S}_{ab}$ by taking the functional derivative of the action with respect to γ_{ab} and then setting $\gamma_{ab} = \eta_{ab}$:

$$\delta A_\partial = \frac{1}{2}\int d^4x\,\sqrt{-\gamma}\,\mathcal{S}^{ab}\delta\gamma_{ab}, \quad \mathcal{S}^{ab}(x) \equiv \left[\frac{2}{\sqrt{-\gamma}}\frac{\delta A_\partial}{\delta\gamma_{ab}(x)}\right]_{\gamma=\eta}. \tag{16}$$

More explicitly,

$$\mathcal{S}^{ab}(x) \equiv \left[\frac{2}{\sqrt{-\gamma}}\frac{\partial L_\partial\sqrt{-\gamma}}{\partial\gamma_{ab}}\right]_{\gamma=\eta} = 2\left[\frac{\partial L\sqrt{-\gamma}}{\partial\gamma_{ab}}\right]_{\gamma=\eta}. \tag{17}$$

This procedure provides another possible prescription for obtaining $\mathcal{S}^{ab}$ in flat space–time. Once again, γ_{ab} is purely a bookkeeping device and, at the end of the calculations, we shall set $\gamma_{ab} = \eta_{ab}$.

Both $\mathcal{S}^{ab}$ and T^{ab} are Lorentz-invariant tensors. In general, there are two crucial differences between these two tensors: (i) T^{ab} is obtained from a generally covariant Lagrangian and hence is generally covariant; $\mathcal{S}^{ab}$ *need not* be generally covariant; (ii) T^{ab} satisfies the identity $\nabla_a T^{ab} = 0$, since it arises from an action which is a generally covariant scalar; $\mathcal{S}^{ab}$ *need not* satisfy this identity. Having said these, one must note that these two tensors are identical whenever the action does not depend on the derivatives of the metric. For a spin 0 field, a spin 1 field and for a relativistic particle, the generally covariant action in Eq. (11) is independent of the derivatives $\partial_a \gamma_{bc}$ of the metric tensor. Hence, in all these three — physically important cases — the two definitions lead to identical energy–momentum tensors: $T^{ab} = \mathcal{S}^{ab}$. Even our apparently noncovariant definition will lead to a generally covariant energy–momentum tensor which satisfies the condition $\nabla_a \mathcal{S}^{ab} = 0$.

For the spin 2 field these two definitions differ. To find $\mathcal{S}^{ab}$, we need to start with the action in the form

$$\begin{aligned} A_\partial &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\partial_a h_{bc}\partial_i h_{jk}\, M^{abcijk}(\gamma^{mn}) \\ &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\partial_a h_{bc}\partial_i h_{jk} \\ &\quad\times\left[\gamma^{ai}\gamma^{bc}\gamma^{jk} - \gamma^{ai}\gamma^{bj}\gamma^{ck} + 2\gamma^{ak}\gamma^{bj}\gamma^{ci} - 2\gamma^{ak}\gamma^{bc}\gamma^{ij}\right], \end{aligned} \tag{18}$$

which differs from that in Eq. (14) by the fact that we have *not* changed ∂_a's to ∇_a's. The tensor $\mathcal{S}^{ab}$ can now be calculated using Eq. (17):

$$\mathcal{S}^{pq}(x) \equiv \left[\frac{2}{\sqrt{-\gamma}}\frac{\partial L\sqrt{-\gamma}}{\partial \gamma_{ab}}\right]_{\gamma=\eta} = \frac{1}{2}\left[\frac{\partial\sqrt{-\gamma}M^{abcijk}(\gamma^{mn})}{\partial\gamma_{pq}}\right]_{\gamma=\eta}\partial_a h_{bc}\partial_i h_{jk}. \tag{19}$$

Again, it is possible to write down the explicit expression for this but, fortunately, we will not need it. However, it is obvious from this expression that this tensor is quadratic in $\partial_a h_{bc}$ and does *not* involve second derivatives of h_{ab}. We shall see in the next section that it is this object which governs, through a term $\mathcal{S}^{ab}h_{ab}$ in the Lagrangian, the coupling of gravity to itself at the lowest nontrivial order.

3. Sneak Preview: Reverse Engineering of Einstein's Theory

We want to obtain Einstein's theory by starting from the action for a spin 2 field in flat space–time and coupling it to some kind of energy–momentum tensor for h_{ab} and iterating the process. Given the ambiguities in the definition of the energy–momentum tensor described in the last section, it makes sense to do it the other way round and identify the correct form of the tensor to which h_{ab} couples in Einstein's theory. This exercise is straightforward: (i) start with an action functional $A_g[g_{ab}]$ which leads to Einstein's field equations for the metric tensor g_{ab}; (ii) define the spin 2 field through $g_{ab} = \eta_{ab} + \lambda h_{ab}$, where h_{ab} has the dimension (length)$^{-1}$ and λ has the dimensions of length; (iii) substitute $g_{ab} = \eta_{ab} + \lambda h_{ab}$ in $A_g[g_{ab}]$ and expand in a Laurent–Taylor series in λ (or, which is the same thing, do a functional Taylor series in h_{ab}). Now, if our ideas are correct, two things must happen: (a) the lowest order term should give the action functional for the spin 2 field in flat space–time with a suitable choice for λ; (b) the next order term will have a Lagrangian of the form $\lambda h_{ab} K^{ab}$ and we should be able to read off K^{ab}. We will carry out this exercise and then comment on various issues.

The conventional action principle for general relativity is the Einstein–Hilbert action given by (with $\lambda^2 = 4\pi G$)

$$A_{\rm EH} \equiv \frac{1}{4\lambda^2}\int R\sqrt{-g}d^4x \equiv \frac{1}{4\lambda^2}\int d^4x[\sqrt{-g}L_{\rm quad} - \partial_j P^j] \equiv A_{\rm quad} + A_{\rm sur}, \quad (20)$$

where

$$L_{\rm quad} = g^{ab}\left(\Gamma^i_{ja}\Gamma^j_{ib} - \Gamma^i_{ab}\Gamma^j_{ij}\right) \quad (21)$$

and

$$P^c = \sqrt{-g}\left(g^{ck}\Gamma^m_{km} - g^{ik}\Gamma^c_{ik}\right) = \sqrt{-g}\,(g^{ac}g^{ji} - g^{ia}g^{cj})\partial_i g_{ac}. \quad (22)$$

The quantity $L_{\rm quad}$ is what is usually called the Γ^2 Lagrangian and is quadratic in the first derivatives of the metric. The term $\partial_i P^i$ integrates to a surface term and is usually ignored while deriving the field equations by assuming that "suitable" boundary conditions can be imposed. A more formal route is to add a suitable boundary term to cancel this, thereby essentially reducing the action to one based on $L_{\rm quad}$.

Classically, there is no way of deciding whether $A_{\rm EH}$ or $A_{\rm quad}$ is the "correct" action, since they lead to the same field equations. Let us first consider $A_{\rm quad}$ and determine to what second rank tensor it self-couples at the lowest order: since $\Gamma \simeq \partial g$, one can express $A_{\rm quad}$ as a quadratic expression in $\partial_a g_{bc}$. Straightforward algebra shows this to be of the form

$$\begin{aligned} A_{\rm quad} &= \frac{1}{4\lambda^2}\int d^4x\,\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk} \\ &\quad\times\left[g^{ai}g^{bc}g^{jk} - g^{ai}g^{bj}g^{ck} + 2g^{ak}g^{bj}g^{ci} - 2g^{ak}g^{bc}g^{ij}\right] \\ &= \frac{1}{4\lambda^2}\int d^4x\,\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk}\,M^{abcijk}(g_{mn}), \end{aligned} \quad (23)$$

with the same functional form as the M^{abcijk} defined in Eq. (4)! *This is a miracle* and all the results of this paper are essentially an exploitation of this miracle. This result means that if, after obtaining the flat space–time action for spin 2 in Eq. (4), we have (i) "just replaced" η_{ab} by g_{ab} in M^{abcijk} and d^4x by $d^4x\sqrt{-g}$ and (ii) used $g_{ab} = \eta_{ab} + \lambda h_{ab}$ in the derivatives $\partial_a h_{bc}$, we will have got the correct Einstein theory. All the (infinite) iterations are required only to understand why this is legitimate.

Let us now proceed with the original program reverse-engineering this action to find out what it couples to at the lowest order. This is quite straightforward. Substituting $g_{ab} = \eta_{ab} + \lambda h_{ab}$ in Eq. (23), we get $\partial_a g_{bc} \partial_i g_{jk} = \lambda^2 \partial_a h_{bc} \partial_i h_{jk}$ exactly. The Taylor series expansion of $\sqrt{-g} M^{abcijk}$ gives

$$\sqrt{-g} M^{abcijk} = M^{abcijk}(\eta^{mn}) + \lambda \left[\frac{\partial \sqrt{-g} M^{abcijk}(g^{mn})}{\partial g_{pq}} \right]_{g=\eta} h_{pq} + \mathcal{O}(\lambda^2). \quad (24)$$

Putting them together, we get the expansion

$$\begin{aligned} A_{\text{quad}} &= \frac{1}{4} \int d^4x \, \partial_a h_{bc} \partial_i h_{jk} \, M^{abcijk}(\eta^{mn}) \\ &\quad + \frac{\lambda}{4} \int d^4x \partial_a h_{bc} \partial_i h_{jk} \left[\frac{\partial \sqrt{-g} M^{abcijk}(g^{mn})}{\partial g_{pq}} \right]_{g=\eta} h_{pq} + \mathcal{O}(\lambda^2). \end{aligned} \quad (25)$$

But the integrand of the second term contains precisely the quantity we defined in Eq. (19). Using it, we get the final answer,

$$A_{\text{quad}} = \frac{1}{4} \int d^4x \, \partial_a h_{bc} \partial_i h_{jk} \, M^{abcijk}(\eta^{mn}) + \frac{\lambda}{2} \int d^4x \, \mathcal{S}^{pq} h_{pq} + \mathcal{O}(\lambda^2), \quad (26)$$

with $\mathcal{S}^{pq}$ given by

$$\mathcal{S}^{pq} = \frac{1}{2} \left[\frac{\partial \sqrt{-\gamma} M^{abcijk}(\gamma^{mn})}{\partial \gamma_{pq}} \right]_{\gamma=\eta} \partial_a h_{bc} \partial_i h_{jk}. \quad (27)$$

We, therefore, have proved that the coupling of gravity to itself, at least to the lowest order, is to a strange beast, defined in a noncovariant way. To $\mathcal{O}(\lambda)$, the field h_{ab} couples to a quantity which is quadratic in the first derivatives, $\partial_a h_{bc}$, of the field and does not couple to an object which has second derivatives of the field. The standard energy–momentum tensor T^{pq} defined in Eq. (13) involves second derivatives of the field, and hence a naive coupling of the form $T^{pq} h_{pq}$ will not match what we have found by explicit computation above. The proof in Eq. (26), of course, gives the result only to the lowest order and one needs to know whether gravity consistently couples to $\mathcal{S}^{ab}$ [as defined by Eq. (19)] to all orders in the coupling constant. It should be intuitively obvious that it will, but we shall provide an explicit proof in Sec. 4.1.

It is, anyway, easy to understand that the self-coupling term in the gravitational Lagrangian *cannot* be of the form $\lambda h_{ab} T_G^{ab}$, where T_G^{ab} is the energy–momentum tensor of the graviton. The reason is the following: a term in the Lagrangian proportional to $\lambda h_{ab} T^{ab}$, where T^{ab} is due to *external* matter fields (assumed to be independent of h_{ab} to this order), will lead to an equation of motion of the type $\partial^2 h = \lambda T$. Hence a coupling of the type $\lambda h_{ab} T^{ab}$ is equivalent to requiring the source to be T_{ab}. Consider now the coupling of gravity to itself through a term of the type $\lambda h_{ab} C^{ab}(h)$, where C_{ab} is some tensor which explicitly depends on the graviton field h_{ab}. When this term is varied with respect to h_{ab} to get the equations of motion, we will obtain *two* terms, $C^{ab} + (\partial C_{ij}/\partial h_{ab}) h^{ij}$, both of which will act as a source to gravity at the next order. If we want the source to be the energy momentum tensor of the graviton field, T^G_{ab}, then the coupling *cannot* be of the form $h_{ab}T^{ab}(h)_G$ since this will lead to the wrong source. What we find is that the coupling in the Lagrangian should be to $\mathcal{S}^{ab}$ if the source of the gravity is to be the energy–momentum tensor to the lowest order. (In this paper, we shall use the terminology "A is coupled to be B" if the Lagrangian in the relevant context has the term AB. This does not necessarily mean that B acts as a source term in Euler Lagrange equations when A is varied, since — in general — B could be a functional of A.) For the same reason, the tensors $\mathcal{S}_{ab}$ and T_{ab} are not the same as any of the energy–momentum pseudotensors[14–28] suggested for the gravitational field. The energy–momentum pseudotensors t_{ab} are introduced at the level of equations of motion [usually requiring the Einstein equations to be recast in the form $\partial_a[t^a_b + T^a_b] = 0$]; that is, they act as the source in the field equations and will have a structure $t^{ab} = C^{ab} + (\partial C_{ij}/\partial h_{ab})h^{ij}$, if the coupling term in the action is $C_{ab}h^{ab}$. In principle, one can find a C^{ab}, for a given t^{ab}, by inverting this functional equation with a suitable choice of ψ^{ijk}; but, in practice, the algebra is tedious and unnecessary in our approach.

So far we have obtained the lowest order self-coupling from the Γ^2 action, and we will now turn our attention to the surface term. We will prove a strong result: *It is impossible to obtain the Einstein–Hilbert action, especially the $A_{\rm sur}$ term, by starting from an action for the spin 2 field which is quadratic in the first derivatives and iterating in powers of λ.* To see this qualitatively, note the structure of the Taylor series expansion for $A_{\rm sur}$; symbolically,

$$\begin{aligned} A_{\rm sur} &\sim \frac{1}{\lambda^2}\int d^4x\partial[\lambda\partial h(\eta + \lambda h + \lambda^2h^2 + \mathcal{O}(\lambda^3))] \\ &\sim \frac{1}{\lambda}\int d^4x\partial[\eta\,\partial h] + \int d^4x\partial[h\partial h] + \lambda\int d^4x\partial[h^2\partial h] + \mathcal{O}(\lambda^2). \end{aligned} \tag{28}$$

(The Lagrangian in $A_{\rm sur}$ is linear in the first derivative of g_{ab}. The substitution $g_{ab} = \eta_{ab} + \lambda h_{ab}$ will lead to a $\lambda\partial h$ type term which — on multiplication by the prefactor λ^{-2} — will give rise to the lowest order term which scales as λ^{-1}.) The third term of the above expansion, which is $\mathcal{O}(\lambda)$ and has the structure $\partial(h^2\partial h)$, can be combined with the coupling term $(\lambda/2)\mathcal{S}^{pq}h_{pq}$ in Eq. (26), which is also $\mathcal{O}(\lambda)$. If we write

$$\partial(h^2\partial h) \sim h[h\partial^2h + (\partial h)^2], \tag{29}$$

it might seem that the $\mathcal{S}^{pq}$ in the coupling term $(\lambda/2)\mathcal{S}^{pq}h_{pq}$ in Eq. (26) changes with a contribution from the second derivatives of the field. One might wonder whether this would help us to get $A_{\rm sur}$ by using the coupling $(\lambda/2)T^{pq}h_{pq}$ and exploiting the second derivatives ∂^2h in T^{pq}. *This idea, however, will not work.* The term we need [$\mathcal{O}(\lambda)$ term] is the *third* term in the Taylor series expansion in Eq. (28), and to get $A_{\rm EH}$ as the final answer, we *must* obtain the first two terms as well. The two *leading terms* in Eq. (28) are $\mathcal{O}(1/\lambda)$ and $\mathcal{O}(1)$ and hence cannot be obtained through the iterative process by coupling in ascending powers of λ.

More explicitly, the two leading terms in $A_{\rm sur}$ in the Einstein–Hilbert Lagrangian are

$$A_{\rm sur} = -\frac{1}{4\lambda}\int d^4x\,\partial_a\partial_b[h^{ab} - \eta^{ab}h^i_i] + \frac{1}{4}\int d^4x\partial_c\left[\frac{1}{2}h\partial^c h - \frac{1}{2}h\partial_k h^{kc} - h^{kc}\partial_k h \right.$$
$$\left. - h^{ab}\partial^c h_{ab} + h^{bc}\partial_k h^k_b + h^{ak}\partial_k h^c_a\right] + \mathcal{O}(\lambda). \tag{30}$$

Thus the leading term in the Einstein–Hilbert Lagrangian (the term in the first line of the above equation),

$$A_{\rm sur} \approx -\frac{1}{4\lambda}\int d^4x\partial_a\partial_b[h^{ab} - \eta^{ab}h^i_i], \tag{31}$$

is nonanalytic in λ when expanded in $g_{ab} = \eta_{ab} + \lambda h_{ab}$. Note that one cannot get out of this nonanalyticity by cheap tricks like rescaling h_{ab} to λh_{ab}. The dimension of the genuine graviton field and the form of the zeroth order Lagrangian uniquely fix the scaling of h_{ab} and there is no freedom for dimensionful scaling left. (In fact, it turns out that this nonanalyticity is vital for the interpretation of the surface term as horizon entropy in semiclassical gravity; so it is not a trivial issue.) For us, the importance of the nonanalyticity lies in the following fact: if one starts with the quadratic spin 2 graviton action and iterate with self-coupling, it is impossible to obtain $R\sqrt{-g}$, since it requires obtaining a piece nonanalytic in λ. One may wonder how previous "derivations" obtained this piece. This was added by hand; since all the terms in $A_{\rm sur}$ are four-divergences, any part of it can be added by hand without affecting the equations of motion, and this is what was done. (We will discuss this in detail, in the context of the derivation by Deser,[6] in Sec. 4.3.)

The Lagrangian in the leading order surface term in Eq. (31), viz. $L_{\rm sur} \equiv \partial_a\partial_b[h^{ab} - \eta^{ab}h^i_i]$, has some interesting properties. First, it is gauge-invariant under the transformation (2), as can be easily checked. (This, of course, means that one cannot set this term to zero by a gauge choice.) Second, it provides a simple counterexample to the belief that if a functional is invariant under infinitesimal gauge transformations [that is, under Eq. (2), with ξ treated as first order infinitesimal], then the expression will be generally covariant. This belief originates from the fact that a metric tensor transforms like in Eq. (2) under infinitesimal coordinate transformations $[x^i \to x^i + \xi^i(x)]$ and one thinks (erroneously) of the finite coordinate transformations as arising from "exponentiating" the infinitesimal ones. Explicitly, the functional

$$F = \partial_i\left[\sqrt{-g}\,\partial_l g_{jk}(g^{jk}g^{li} - g^{lk}g^{ji})\right] \tag{32}$$

is clearly not invariant under arbitrary coordinate transformations. But if we take $g_{ab} = \eta_{ab} + \lambda h_{ab}$, then the expression for F becomes, to linear order in h,

$$F_{\rm lin} \approx \lambda\partial_i[\partial_l h_{jk}(\eta^{jk}\eta^{li} - \eta^{lk}\eta^{ji})] = -\lambda\partial_i\partial_l[h^{il} - \eta^{il}h^j_j], \tag{33}$$

which is the same as the integrand in Eq. (31). This expression, however, is gauge-invariant under the transformations in Eq. (2). This shows that it is possible to have scalars which are invariant under infinitesimal gauge transformations but are not generally covariant.

4. From Gravitons to Gravity: General Procedure

After the sneak preview of the results to come, we shall now return to the original task of coupling the spin 2 field to matter, as well as to itself, self-consistently to all orders. We shall start with the issue of coupling the spin 2 field to other matter fields self-consistently to all orders, to see how an externally specified $h_{ab}(x)$ affects the dynamics of $\phi_A(x)$.

Consider a field $\phi_A(x^a)$ described by a Lagrangian density $L(\phi_A, \partial\phi_A, \eta_{ab})$ in *flat* space–time, in the Cartesian coordinates in which the metric is $\eta_{ab} = \text{dia}(-1, 1, 1, 1)$. The index A formally denotes all the indices the field carries depending on its spin; we will assume that the field is bosonic, for simplicity. To couple h_{ab} to this matter field, we need to first find a suitable, second rank tensor field K^{ab}, defined in terms of the matter variables. In Sec. 2, we introduced two such tensors (among an infinite number of possibilities) in Eqs. (13) and (17). Both can be generically expressed as functional derivatives in the form

$$\delta A_0 = \frac{1}{2}\int d^4x\,\sqrt{-\gamma}\,K^{ab}\delta\gamma_{ab}, \quad K^{ab}(x) \equiv \left[\frac{2}{\sqrt{-\gamma}}\frac{\delta A_0}{\delta\gamma_{ab}(x)}\right]_{\gamma=\eta}. \tag{34}$$

If we use $A_0 = A_\nabla(\phi_A, \nabla\phi_A, \gamma_{ab})$ (with covariant derivatives), we get the conventional energy–momentum tensor $K^{ab} = T^{ab}$. If, instead, we use $A_0 = A_\partial(\phi_A, \partial\phi_A, \gamma_{ab})$, we get $K^{ab} = \mathcal{S}^{ab}$, which is a hybrid object, except for the relativistic particle, spin 0 field or spin 1 field for which the definitions coincide and $T^{ab} = \mathcal{S}^{ab}$. For most of our algebra below, we need not specify whether we are using A_∇ or A_∂ and we will use the generic symbols A and K^{ab} to represent either definition.

We now want to couple the second rank symmetric tensor field h_{ab} to K^{ab}. To the lowest order, this is done by changing the action from A_0 to $A_{\leq 1} \equiv A_0 + A_1$, where A_1 is chosen such that

$$\delta A_1 = \frac{\lambda}{2}\int d^4x_1\sqrt{-\gamma}\,K^{ab}(x_1)\delta h_{ab}(x_1) = \lambda\int d^4x_1\left[\frac{\delta A_0}{\delta\gamma_{ab}(x_1)}\right]_{\gamma=\eta}\delta h_{ab}(x_1), \tag{35}$$

where λ is a coupling constant. To the lowest order, K^{ab} is independent of h_{ab} and we can integrate this to obtain the action

$$\begin{aligned} A_{\leq 1} = A_0 + A_1 &= A_0 + \lambda\int d^4x_1\left[\frac{\delta A_0}{\delta\gamma_{ab}(x_1)}\right]_{\gamma=\eta} h_{ab}(x_1) \\ &= A_0 + \frac{\lambda}{2}\int d^4x_1\sqrt{-\gamma}\,K^{ab}(x_1)h_{ab}(x_1). \end{aligned} \tag{36}$$

The addition of this coupling will, however, change the definition of K^{ab}, since the second term A_1 contributes to K^{ab} via Eq. (34). To take this into account, we need to add a term A_2 in a manner similar to what we did in Eq. (35); that is, we need to impose

$$\delta A_2 = \lambda \int d^4x_2 \left[\frac{\delta A_1}{\delta\gamma_{cd}(x_2)}\right]_{\gamma=\eta} \delta h_{cd}(x_2). \tag{37}$$

Using the form for A_1 in Eq. (36) we get

$$\delta A_2 = \lambda^2 \int d^4x_1 d^4x_2 \left[\frac{\delta^2 A_0}{\delta\gamma_{cd}(x_2)\delta\gamma_{ab}(x_1)}\right]_{\gamma=\eta} h^{ab}(x_1)\delta h^{cd}(x_2). \tag{38}$$

Integrating, we get the second order correction as

$$A_2 = \frac{\lambda^2}{2} \int d^4x_1 d^4x_2 \left[\frac{\delta^2 A_0}{\delta\gamma_{cd}(x_2)\delta\gamma_{ab}(x_1)}\right]_{\gamma=\eta} h^{ab}(x_1)h^{cd}(x_2). \tag{39}$$

It is obvious that this term will bring about another correction, etc. The sum of the infinite series of terms in the action will be

$$A_\infty = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \int d^4x_1 \cdots d^4x_n \left[\frac{\delta^n A_0}{\delta\gamma_{ab}(x_1)\cdots\delta\gamma_{ij}(x_n)}\right]_{\gamma=\eta} h_{ab}(x_1)\cdots h_{ij}(x_n), \tag{40}$$

which is just a functional Taylor series expansion leading to

$$A_\infty = A_0(\gamma_{ab} + \lambda h_{ab})\,\Big|_{\gamma=\eta}. \tag{41}$$

An alternative way of obtaining the same result is to note that, at every order, we have the recurrence relation, similar to Eqs. (35) and (38),

$$\delta A_{n+1} = \lambda \int d^4x_1 \left[\frac{\delta A_n}{\delta\gamma_{ab}(x_1)}\right]_{\gamma=\eta} \delta h_{ab}(x_1), \tag{42}$$

which is the same as

$$\frac{\delta A_{n+1}}{\lambda\delta h_{ab}(x)} = \left[\frac{\delta A_n}{\delta\gamma_{ab}(x)}\right]_{\gamma=\eta}. \tag{43}$$

Summing both sides to infinite orders, we find that A_∞ satisfies the relation

$$\frac{\delta A_\infty}{\lambda\delta h_{ab}(x)} = \left[\frac{\delta A_\infty}{\delta\gamma_{ab}(x)}\right]_{\gamma=\eta}, \tag{44}$$

which has the general solution given by Eq. (41). This is our key result.

Thus we can consistently couple a field h_{ab} to K^{ab} by the rule given in Eq. (41): this allows us to compute the effect of an external h_{ab} on the system if we insist that the external field consistently couple to a tensor K^{ab} which can be expressed as in Eq. (34). Since the curvilinear metric γ_{ab} was introduced only as a bookkeeping device to allow for variation of the action, the final action is given by replacing γ_{ab} with η_{ab} at the end of the calculation. Note that there is a subtle difference between Eq. (41) and the expression obtained by replacing η_{ab} with $(\eta_{ab} + \lambda h_{ab})$ in the original action. The latter will miss, for example, the $\sqrt{\det|\gamma|} \to \sqrt{\det|\eta + h|}$ kind of factors. We need to introduce γ_{ab} in order to provide a placeholder in the final expression.

Here comes the parting of ways. If we had chosen $A = A_\nabla(\phi_A, \nabla\phi_A, \gamma_{ab})$ (with covariant derivatives), then $K^{ab} = T^{ab}$ would be the standard energy–momentum tensor and our result would give the final matter action as

$$A_\infty = A_0(\phi_A, \nabla_{(g)}\phi_A, g_{ab}), \tag{45}$$

where we have used the abbreviation $g_{ab} \equiv \eta_{ab} + \lambda h_{ab}$ and ∇ is defined with respect to this metric. This is a generally covariant matter action in the space–time with metric g_{ab} and agrees with all the textbook results. It should, however, be stressed that this cannot be considered a *derivation* of general covariance of matter action when self-consistently coupled to the spin 2 field. This is because we made a rule for finding T^{ab} which has general covariance with respect to curved space–time built in as an *assumption*. All that has been shown is that this extends to an interpretation of g_{ab} as a metric tensor and only the combination $g_{ab} \equiv \eta_{ab} + \lambda h_{ab}$ is relevant for the matter sector.

Since the final matter Lagrangian we obtained is generally covariant, its energy–momentum tensor has zero covariant divergence, leading to

$$\nabla_a \left[\frac{\delta A_\infty}{\delta g_{ab}}\right] = 0, \tag{46}$$

where ∇ is the covariant derivative operator corresponding to $g_{ab} = \eta_{ab} + \lambda h_{ab}$.

On the other hand, if we had chosen $A = A_\partial(\phi_A, \partial\phi_A, \gamma_{ab})$ (without covariant derivatives), then $K^{ab} = \mathcal{S}^{ab}$ would be the best we have introduced in Sec. 2.4 and our result would give the final matter action as

$$A_\infty = A_0(\phi_A, \partial\phi_A, g_{ab}). \tag{47}$$

In general, this is not a generally covariant matter Lagrangian since we have not replaced partial derivatives by covariant derivatives. The metric appears only through the $\sqrt{-g}$ factor and by the replacement of η_{ab} with g_{ab}. This, of course, does not matter for the Lagrangians of a relativistic particle, spin 0 field or spin 1 field. In all these cases this prescription does lead to a generally covariant Lagrangian, though this is not by design. Equation (46) also holds for the spin 0 or spin 1 field or for the relativistic particles, *but not in general.* (We shall comment on this in Sec. 5.)

4.1. *How to obtain Einstein gravity from the spin 2 field*

To see where all this is leading to, let us consider next the real issue: that of coupling the graviton field to itself. In our approach this is ridiculously simple. We merely use the fact that the analysis leading to Eq. (41) was completely independent of the form of A_0 as well as the nature of the fields ϕ_A. Hence we can use the same prescription when ϕ_A is the second rank symmetric tensor field h_{ab} itself.

The key ambiguity, of course, is whether we want to use A_∇ and couple to T^{ab} or use A_∂ and couple to $\mathcal{S}^{ab}$. Let us use the second procedure first: the A_∂ for the graviton field is obtained from the first line of Eq. (3) by replacing η^{ab} with γ^{ab} in $M^{abcijk}(\eta)$ and multiplying by $\sqrt{-\gamma}$. This is given by Eq. (18):

$$\begin{aligned} A_\partial &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\partial_a h_{bc}\partial_i h_{jk}\,M^{abcijk}(\gamma^{mn}) \\ &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\partial_a h_{bc}\partial_i h_{jk} \\ &\quad\times\left[\gamma^{ai}\gamma^{bc}\gamma^{jk} - \gamma^{ai}\gamma^{bj}\gamma^{ck} + 2\gamma^{ak}\gamma^{bj}\gamma^{ci} - 2\gamma^{ak}\gamma^{bc}\gamma^{ij}\right]. \end{aligned} \tag{48}$$

Our prescription now requires that A_∞ for the field h_{ab} be obtained by replacing γ_{ab} with $g_{ab} \equiv \eta_{ab} + \lambda h_{ab}$. This leads to the action

$$\begin{aligned} A_\infty &= \frac{1}{4\lambda^2}\int d^4x\,\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk} M^{abcijk}(g^{mn}) \\ &= \frac{1}{4\lambda^2}\int d^4x\,\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk} \\ &\quad\times\left[g^{ai}g^{bc}g^{jk} - g^{ai}g^{bj}g^{ck} + 2g^{ak}g^{bj}g^{ci} - 2g^{ak}g^{bc}g^{ij}\right], \end{aligned} \tag{49}$$

which is precisely the Γ^2 action in general relativity. The variation of this action will lead to Einstein's field equations *in vacuum*, and we have achieved our goal.

As an aside, we note that the above approach will work with any generic action functional for the spin 2 field. A different choice of the spin 2 action functional will lead to a different final theory. This aspect is briefly discussed in App. B, since it is irrelevant to the main issues of this paper.

4.2. *The failure of the conventional procedure to lead to Einstein's theory*

Let us see what happens if we follow the conventional procedure. For this, we need to start with the A_∇ of graviton action and couple to the standard T^{ab}. This action is given by Eq. (14):

$$\begin{aligned} A_\nabla &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\nabla_a h_{bc}\nabla_i h_{jk}\,M^{abcijk}(\gamma^{mn}) \\ &= \frac{1}{4}\int d^4x\,\sqrt{-\gamma}\,\nabla_a h_{bc}\nabla_i h_{jk} \\ &\quad\times\left[\gamma^{ai}\gamma^{bc}\gamma^{jk} - \gamma^{ai}\gamma^{bj}\gamma^{ck} + 2\gamma^{ak}\gamma^{bj}\gamma^{ci} - 2\gamma^{ak}\gamma^{bc}\gamma^{ij}\right]. \end{aligned} \tag{50}$$

Our prescription now requires that A_∞ be obtained by replacing γ_{ab} with $g_{ab} \equiv \eta_{ab} + \lambda h_{ab}$. This leads to the action

$$\begin{aligned} A_\infty &= \frac{1}{4}\int d^4x\,\sqrt{-g}\,\nabla_a h_{bc}\nabla_i h_{jk}\,M^{abcijk}(g^{mn}) \\ &= \frac{1}{4}\int d^4x\,\sqrt{-g}\,\nabla_a h_{bc}\nabla_i h_{jk} \\ &\quad\times\left[g^{ai}g^{bc}g^{jk} - g^{ai}g^{bj}g^{ck} + 2g^{ak}g^{bj}g^{ci} - 2g^{ak}g^{bc}g^{ij}\right], \end{aligned} \tag{51}$$

where the ∇ operator is with respect to the metric g_{ab}. Now, since $\nabla_i g_{jk} = 0$, we get

$$\begin{aligned} \nabla_a h_{bc} &= \frac{1}{\lambda}\nabla_a[g_{bc} - \eta_{bc}] = -\frac{1}{\lambda}\nabla_a\eta_{bc} = -\frac{1}{\lambda}[\Gamma^i_{ba}\eta_{ic} + \Gamma^i_{ca}\eta_{ib}] \\ &= -\frac{1}{\lambda}[g^{pi}\Gamma_{pba}(g_{ic} - \lambda h_{ic}) + (b \leftrightarrow c)] \\ &= -\frac{1}{\lambda}[\Gamma_{pba}(\delta^p_c - \lambda g^{pi}h_{ic}) + (b \leftrightarrow c)] \end{aligned} \tag{52}$$

and a similar expression for $\nabla_i h_{jk}$. Since these are multiplied by $M^{abcijk}(g^{mn})$, which is symmetric in $(b,c),(i,j)$, we can ignore the $(b \leftrightarrow c)$ term, etc. We thus obtain

$$\begin{aligned} \nabla_a h_{bc}\nabla_i h_{jk}\,M^{abcijk}(g^{mn}) &= \frac{4}{\lambda^2}[\Gamma_{pba}\Gamma_{qki}(\delta^p_c - \lambda g^{pi}h_{ic})(\delta^q_j - \lambda g^{lq}h_{lj})] \\ &\quad\times M^{abcijk}(g^{mn}). \end{aligned} \tag{53}$$

Of the four terms which arise on expanding out the product, $(\delta^p_c - \lambda g^{pi}h_{ic})(\delta^q_j - \lambda g^{lq}h_{lj})$, the first term can be transformed, again using the symmetry of $M^{abcijk}(g^{mn})$ in $(b,c),(i,j)$, to give

$$\begin{aligned} \frac{4}{\lambda^2}\Gamma_{cba}\Gamma_{jki}M^{abcijk}(g^{mn}) &= \frac{1}{\lambda^2}M^{abcijk}(g^{mn})[\Gamma_{cba} + \Gamma_{bca}][\Gamma_{jki} + \Gamma_{kji}] \\ &= \frac{1}{\lambda^2}M^{abcijk}(g^{mn})\partial_a g_{bc}\,\partial_i g_{jk}. \end{aligned} \tag{54}$$

This is precisely the Lagrangian term in Einstein's theory, in the form of Γ^2 action [of Eq. (49)]. Unfortunately, there are three more terms in Eq. (53) of the $\Gamma\Gamma hh$ and $\Gamma\Gamma h$ forms. They do not vanish, they are not total divergences and they depend explicitly on h_{ab}. Thus the action functional obtained by coupling the spin 2 field to the standard energy–momentum tensor is not that of Einstein's theory.

The manner in which we led the reader to the result should not come as a surprise. We have, in fact, shown explicitly in Sec. 3 that the standard action for Einstein's theory does couple to $\mathcal{S}^{ab}$ in the lowest order. So, clearly, we will not get Einstein's theory if we force the conventional energy–momentum tensor T^{ab} on to the spin 2 field.

4.3. *Comments on the previous work*

Finally, we shall discuss how the previous "derivations" escaped this problem. As described in Sec. 1, none of the previous derivations other than that of Deser[6] actually performs any iteration. All the rest of them tacitly or explicitly bring in general covariance for the gravity sector, after which it is trivial to obtain $R\sqrt{-g}$ as the Lagrangian; hence, we need not discuss them any further. Deser does perform the iteration using the first order form of gravity and using $\sqrt{-g}g^{ab}$ as the chosen variable. Let us first summarize this approach briefly in a slightly different language, to bring the essential ingredients into focus.

We first note that the standard Einstein–Hilbert action can be expressed in the first order Palatini form using the variables Γ^a_{bc} and $f^{ab} \equiv \sqrt{-g}g^{ab}$:

$$A_{\rm EH} = \frac{1}{4\lambda^2}\int d^4x f^{ab}R_{ab}(\Gamma), \quad f^{ab} \equiv \sqrt{-g}g^{ab}. \tag{55}$$

Varying f^{ab} and Γ^a_{bc} independently in this action will lead to standard Einstein equations. If we substitute $f^{ab} = \eta^{ab} + \lambda q^{ab}$ into this action (without any approximations), then the Lagrangian becomes

$$L_{\rm EH} = \frac{1}{4\lambda^2}[(\eta^{ab} + \lambda q^{ab})(R^L_{ab}(\Gamma) + R^Q_{ab}(\Gamma))], \tag{56}$$

where $R^L_{ab}(\Gamma)$ and $R^Q_{ab}(\Gamma)$ are the linear and quadratic parts of the Ricci tensor:

$$R^L_{ab}(\Gamma) = \partial_c\Gamma^c_{ab} - \partial_b\Gamma_a, \quad R^Q_{ab}(\Gamma) = \Gamma_c\Gamma^c_{ab} - \Gamma^c_{ad}\Gamma^d_{bc}, \quad \Gamma_c \equiv \Gamma^i_{ci}. \tag{57}$$

Expanding out the product in Eq. (56), we get four terms, which we will group as

$$L_{\rm EH} = L_0 + L_1 + L_2 = \frac{1}{4\lambda^2}\eta^{ab}R^L_{ab}(\Gamma) + \frac{1}{4\lambda^2}[\eta^{ab}R^Q_{ab}(\Gamma) + \lambda q^{ab}R^L_{ab}(\Gamma)] + \frac{1}{4\lambda^2}\lambda q^{ab}R^Q_{ab}(\Gamma). \tag{58}$$

We now notice something remarkable:

- The first term, $\eta^{ab}R^L_{ab} = \partial_i(\Gamma^i - \Gamma^{ij}{}_j)$, is a total divergence. Let us assume that we are allowed to drop this term.
- The second term has the piece $[\eta^{ab}R^Q_{ab}(\Gamma) + \lambda q^{ab}R^L_{ab}(\Gamma)]$, which is essentially equivalent to the zeroth order action for the spin 2 graviton (in the second order formalism) *plus a very specific* four-divergence term. [This will be apparent if we write the term $qR^L \sim q\partial\Gamma \sim \partial(q\Gamma) - \Gamma\partial q$; we find that it has a quadratic term plus a *very specific* total divergence term $\partial(q\Gamma)$.] Let us assume that we are allowed to start with this very specific term as the lowest order graviton Lagrangian.
- Granted these two wishes, we can obtain from this L_1 term the tensor $t_{ab} = (\delta L_1/\delta\eta^{ab}) = (1/4\lambda^2)R^Q_{ab}(\Gamma)$, purely in a formal way. If we think of this as the energy–momentum tensor for the graviton, then the next order coupling should be $\lambda q^{ab}t_{ab} = (1/4\lambda^2)[\lambda q^{ab}R^Q_{ab}(\Gamma)]$, which is precisely the last piece, L_2, in the

Einstein–Hilbert Lagrangian! What is more, this last term is independent of η^{ab}, so $(\delta L_2/\delta\eta^{ab}) = 0$ and *no further iterations are required.* [Since we are working with $\sqrt{-g}g^{ab}$ and not g^{ab}, the variation actually gives $T_{ab} - (1/2)g_{ab}T$ rather than the energy–momentum tensor itself; but this is irrelevant to our discussion.]

When one attempts to do this properly, one faces an important issue: as we have said before, one cannot really define things like $(\delta L/\delta\eta^{ab})$, where η^{ab} is the Minkowski metric. So we first need to write L_1 in curved space–time with a metric γ^{ab} and compute variations with respect to this metric. What do we do to the ∂_a in the definition of $R^L_{ab}(\Gamma) = \partial_c\Gamma^c_{ab} - \partial_b\Gamma_a$? Let us suppose that we change them to ∇_a for the metric γ^{ab}. Then the t_{ab} computed from the functional derivative will pick up additional terms. The action L_1 in curved space–time is

$$L_1(\gamma) = \frac{1}{4\lambda^2}[\sqrt{-\gamma}\gamma^{ab}R^Q_{ab}(\Gamma) + \lambda q^{ab}R^L_{ab}(\Gamma)], \quad \bar{\gamma}^{ab} \equiv \sqrt{-\gamma}\gamma^{ab}, \tag{59}$$

where the R_{ab} is now evaluated with partial derivatives replaced by covariant derivatives with respect to γ_{ab}, etc. The variation of this Lagrangian with respect to $\bar{\gamma}^{ab}$ gives [see Eq. (8) of Ref. 6]

$$t_{ab} = \frac{1}{4\lambda^2}\left[R^Q_{ab}(\Gamma) + \lambda\sigma_{ab}\right], \tag{60}$$

where the second term σ_{ab} arises from the variation of the $\partial_a\gamma_{bc}$ terms, because we have changed ∂_a to ∇_a. Its explicit form,

$$\begin{aligned}\sigma_{ab} = \partial^c \bigg[&-\eta_{ab}\left(q^j_i\Gamma^i_{cj} - \frac{1}{2}q\Gamma_c\right) - 2q^i_c\Gamma_{(ab)i} - 2q^i_{(a}\Gamma_{cib)} \\ &+ 2q^i_{(a}\Gamma_{b)ci} - q_{ab}\Gamma_c + 2q_{c(a}\Gamma_{b)}\bigg],\end{aligned} \tag{61}$$

is irrelevant to us. We are now in trouble, since the next order coupling $t_{ab}q^{ab}$ will have an unwanted term, $q^{ab}\sigma_{ab}$, in addition to the term we want (proportional to $q^{ab}R^Q_{ab}$). Deser simply drops this term, saying (see his comment after Eq. (9) of Ref. 6) "Note that we have not added the full $h^{\mu\nu}\tau_{\mu\nu}$, but rather used the simple part of $\tau_{\mu\nu}$ only" — without any additional justification! Then, of course, one gets the L_2 as the next term and iteration stops there.

It should now be obvious that Deser's derivation requires the following implicit assumptions:

- One should drop the L_0 term in Eq. (58) unceremoniously, saying that it is a total divergence [Deser does this in going from Eq. (2) to Eq. (4) in Ref. 6]. It is precisely this term which, in second order formalism, has the non-analytic behavior $1/\lambda$ and is displayed as the first term in Eq. (30). Sure, it is a four-divergence, but one can *never* get it from the graviton's quadratic action and one needs to add and subtract this term, at will, to get Einstein–Hilbert action.

- One should start with L_1, which is not the graviton action that is quadratic in the derivatives of the field (which a particle physicist would have written down from first principles) but the one with a very specific total divergence added to it. This is precisely the $\mathcal{O}(1)$ in Eq. (30). There is no way anyone could have guessed this specific total divergence term without knowing the final answer!
- One should drop the terms in t_{ab} which arise from varying $\partial_a \gamma_{bc}$. But this is precisely the same as using our quantity $\mathcal{S}^{ab}$! Or, rather, not changing ∂ to ∇ when one takes the graviton action from flat space–time to curved space–time. So, in real terms, the two derivations match mathematically and our conclusion stands: *Gravity self-couples to* $\mathcal{S}^{ab}$, *not to* T^{ab}. It just was not realized before.

If we are not attempting to *derive* Einstein's theory from the spin 2 theory but only want to *reinterpret* it in the field-theory language, then we may be willing to live with the first two issues mentioned above. We can use the hindsight gained from general relativity and add/subtract four divergences at will to/from the action, to obtain the necessary final form. (We must then admit that the Venusian physicists whom Feynman keeps referring to in Ref. 4 would never have got there.) But the third issue is not a matter of opinion or a point of view; what quantity gravity couples to in becoming nonlinear is a well-defined mathematical question. Our analysis — and proper interpretation of previous work — gives an unconventional answer.

5. Conclusions

We have shown that it is not possible to obtain the Einstein–Hilbert *action* starting from the standard graviton action and iterating in the coupling constant. This is because of the existence of the total divergence term in the Einstein–Hilbert action which is nonanalytic in the coupling constant, when expanded in terms of the graviton field. This result is crucial because a series of previous investigations[34–44] have shown that the surface term is vital in the thermodynamics of horizons and in semiclassical gravity. In fact, I started this investigation to understand how the surface term — and, hence, possibly the entropy of horizons — can be interpreted in terms of the graviton field in a Minkowski background. The result shows that one simply cannot understand the surface term in a standard field-theoretical language, using the graviton field. There is more to gravity than gravitons. (There is sufficient evidence to assume that gravity is not a fundamental field but an emergent phenomenon like elasticity. In that case, the peculiar structure of the gravitational Lagrangian needs to be understood as an effective low energy phenomenon.[53–58])

In a strictly classical theory, what matters is the equation of motion and not the form of the action principle. Hence, the fact that we cannot get the surface term in the Einstein–Hilbert action is not of concern if we are only interested in Einstein's equations. Our analysis shows that it is indeed possible to obtain the quadratic Γ^2 action (and thus Einstein's equations) by starting from the the graviton action and iterating in the coupling constant. But to do this, we need to couple h_{ab} to a second rank tensor $\mathcal{S}^{ab}$ which is different from the standard energy–momentum tensor T_G^{ab} of the graviton. Indeed, as we explained in Sec. 3 [see the discussion after Eq. (27)], if the *source* of gravity at each order of iteration has to be the energy–momentum tensor of the graviton evaluated at the previous order, then the *coupling* in the Lagrangian *cannot* be of the form $h_{ab}T_G^{ab}$ since the h_{ab} dependence of the T_G^{ab} will lead to an extra term on variation. A term in the Lagrangian of the form $h_{ab}\mathcal{S}^{ab}$ does lead to the energy–momentum tensor as the source of gravity. Identifying the nature of $\mathcal{S}^{ab}$ and bringing it into focus has been one of the results of this paper.

If we were only interested in pure gravity, this would have been the whole story. But, in that case, it would have been an unnecessary exercise. The *linear* spin 2 field, uncoupled to anything, is a perfectly consistent theory and we need not try to couple it to itself. So the whole exercise has meaning only when we have both matter and spin 2 field and we try to couple them consistently. Then we need to assume that the spin 2 field couples to itself through $\mathcal{S}^{ab}$, while it couples to matter through T^{ab}. This assumption will lead consistently to Einstein's theory and seems to be the most viable option, if we want to obtain standard gravity coupled to matter, starting from the graviton action. (Of course, in a world made of a spin 2 field coupled to matter made of only relativistic particles, spin 0 fields and spin 1 fields, one can assume that all the coupling is through $\mathcal{S}^{ab}$; this is because for matter made of these constituents, $\mathcal{S}^{ab} = T^{ab}$.)

Two facts need to be borne in mind as regards this option. First, we do not know anything about the coupling of the spin 2 field to itself except through standard gravity; and the analysis in Sec. 3 shows that gravity does couple to itself through a term, $h_{ab}\mathcal{S}^{ab}$. Second, there is no conflict with the principle of equivalence even though the self-coupling term is $h_{ab}\mathcal{S}^{ab}$ while the coupling to the external source is through $h_{ab}T^{ab}$. What matters for the principle of equivalence is the fact that the source for gravity is always the energy–momentum tensor This is indeed assured in our approach and — as has been stressed several times — this requires a self-coupling term of the form $h_{ab}\mathcal{S}^{ab}$ in the Lagrangian.

Acknowledgments

I thank S. Deser for his time and patience during the many rounds of lively e-mail discussion, stretching over six months, regarding earlier versions of this draft, which resulted in considerable improvement of the contents and presentation. We achieved a fair amount of convergence, but not in all issues. I thank Apoorva Patel, R. Nityananda and K. Subramanian for detailed comments on an earlier draft, and D. Lynden-Bell, A. Lasenby and J. Bjorken for discussions on several related issues.

Appendix A. Physics of the Spin 2 Field: A Brief Review

In this Appendix we will briefly review the theory of the spin 2 field and collect together different results which are required later. (This is done especially since

I could not find a convenient source for pedagogical discussion of the spin 2 field; the results are somewhat scattered in the literature.[45–50])

A.1. Action functional for the spin 2 field

The action for the non-interacting, massless, spin 2 field h_{ab} is built out of scalars which are quadratic in the derivatives $\partial_a h_{bc}$. The most general expression will be the sum of different scalars obtained by contracting pairs of indices in $\partial_a h_{bc}\partial_i h_{jk}$ in a different manner. Since this product is symmetric in (b, c) and (j, k) and also under the interchange $(a, b, c) \to (i, j, k)$, it is easy to figure out that, *a priori*, seven different contractions are possible. For example, if a is contracted with i, then there are two possibilities for contracting b (with either c or j; contracting b with k is the same as contracting b with j). These contractions will lead to the terms $c_1\partial_a h_{bc}\partial_i h_{jk}\eta^{ai}\eta^{bc}\eta^{jk} + c_2\partial_a h_{bc}\partial_i h_{jk}\eta^{ai}\eta^{bj}\eta^{ck} = c_1\partial_a h^b_b\partial^a h^j_j + c_2\partial_a h_{bc}\partial^a h^{bc}$ in the Lagrangian with as-yet-undetermined constants (c_1, c_2). For brevity, we will denote these two terms symbolically as $(ai, bc, jk), (ai, bj, ck)$. Next, if a is contracted with b, there are again two inequivalent possibilities for contracting c, leading to $(ab, ci, jk), (ab, ck, ij)$. Finally, if a is contracted with k, there are three possible ways of contracting b, giving $(ak, bj, ci), (ak, bc, ij), (ak, bi, cj)$.

Of these, the contraction (ak, bc, ij) is the same as (ab, ci, jk) since $(ak, bc, ij) = (ic, jk, ab)$ under $(a, b, c) \leftrightarrow (i, j, k)$ and, of course, $(ic, jk, ab) = (ab, ci, jk)$. Similarly, $(ak, bj, ci) = (ic, jb, ka) = (ib, jc, ka)$; the first equality comes from $(a, b, c) \leftrightarrow (i, j, k)$ symmetry while the second arises from $b \leftrightarrow c$ symmetry. Since $(ib, jc, ka) = (ak, bi, cj)$ trivially, we need to retain only the first two of the three possibilities in the last set. Thus, dropping the two contractions (ab, ci, jk), (ak, bi, cj) out of the seven possibilities, we are left with five different contractions: $(ai, bc, jk), (ai, bj, ck), (ab, ck, ij), (ak, bj, ci), (ak, bc, ij)$. This will correspond to an action for the spin 2 field of the form

$$\begin{aligned} A = \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk}\, &\big[c_1\eta^{ai}\eta^{bc}\eta^{jk} + c_2\eta^{ai}\eta^{bj}\eta^{ck} \\ &+ c_3\eta^{ab}\eta^{ck}\eta^{ij} + c_4\eta^{ak}\eta^{bj}\eta^{ci} + c_5\eta^{ak}\eta^{bc}\eta^{ij}\big] \\ = \frac{1}{4}\int d^4x\, &[c_1\partial_a h^b_b\partial^a h^j_j + c_2\partial_a h_{bc}\partial^a h^{bc} \\ &+ c_3\partial_a h^{ab}\partial_i h^i_b + c_4\partial_a h_{bc}\partial^c h^{ba} + c_5\partial_a h^b_b\partial_i h^{ia}]. \end{aligned} \tag{A.1}$$

Each term in the action in Eq. (A.1) is of the kind $\partial_a h_{bc}\partial_i h_{jk}J^{abcijk}(\eta)$, where J^{abcijk} is a cubic in η^{lm} and hence is constant, i.e. all components are 0 or ± 1. This allows one to "swap" the derivatives ∂_i and ∂_a by adding a total divergence, using the identity

$$[\partial_a h_{bc}\partial_i h_{jk} - \partial_i h_{bc}\partial_a h_{jk}]J^{abcijk} = \partial_a[h_{bc}\partial_i h_{jk}(J^{abcijk} - J^{ibcajk})]. \tag{A.2}$$

Using this result, one can convert the c_3 term to the c_4 term and rewrite the action as

$$\begin{aligned} A &= \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk}\big[c_1\eta^{ai}\eta^{bc}\eta^{jk} + c_2\eta^{ai}\eta^{bj}\eta^{ck} \\ &\quad + (c_3+c_4)\eta^{ak}\eta^{bj}\eta^{ci} + c_5\eta^{ak}\eta^{bc}\eta^{ij}\big] + A_{\rm div} \\ &= \frac{1}{4}\int d^4x\big[c_1\partial_a h^b_b\partial^a h^j_j + c_2\partial_a h_{bc}\partial^a h^{bc} \\ &\quad + (c_3+c_4)\partial_a h_{bc}\partial^c h^{ba} + c_5\partial_a h^b_b\partial_i h^{ia}\big] + A_{\rm div} \\ &\equiv A_h + A_{\rm div}, \end{aligned} \tag{A.3}$$

where

$$\begin{aligned} A_{\rm div} &= \frac{c_3}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk}\big[\eta^{ab}\eta^{ck}\eta^{ij} - \eta^{ak}\eta^{bj}\eta^{ci}\big] \\ &= \frac{c_3}{4}\int d^4x\big[\partial_a h^{ab}\partial_i h^i_b - \partial_a h_{bc}\partial^c h^{ba}\big] \\ &= \frac{c_3}{4}\int d^4x\, \partial_a\big[h^{ab}\partial_i h^i_b - h^{ib}\partial_i h^a_b\big], \end{aligned} \tag{A.4}$$

which, being a total divergence, does not contribute to the equations of motion if suitable boundary conditions are imposed. Notice that there are *no* further ambiguities related to "swapping" of derivatives in the action in Eq. (A.2); this is clearly not possible in the c_1, c_2 or c_5 terms, since the swapping leads to identical terms. Hence the only ambiguity is in the choice between the c_3 term and the c_4 term.

Interestingly enough, the constants $c_1, c_2, c_5, (c_3 + c_4)$ in A_h of Eq. (A.3) can be determined except for an overall scaling by the requirement that the field equations should be invariant under the gauge transformation:

$$h_{ab}(x) \to h_{ab}(x) + \partial_a\xi_b(x) + \partial_b\xi_a(x). \tag{A.5}$$

This fixes the constants to be $c_1 = -c_2 = 1; c_3 + c_4 = -c_5 = 2$ except for an overall scaling which is left as $1/4$ for future convenience. The resulting expression for the quadratic part of the action can be written in different forms:

$$\begin{aligned} A_h &= \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk}\big[\eta^{ai}\eta^{bc}\eta^{jk} - \eta^{ai}\eta^{bj}\eta^{ck} + 2\eta^{ak}\eta^{bj}\eta^{ci} - 2\eta^{ak}\eta^{bc}\eta^{ij}\big] \\ &= \frac{1}{4}\int d^4x[\partial_i h^a_a\partial^i h^j_j - \partial_a h_{bc}\partial^a h^{bc} + 2\partial_a h_{bc}\partial^c h^{ba} - 2\partial_a h^b_b\partial_i h^{ia}] \\ &= \frac{1}{4}\int d^4x\left[\frac{1}{2}\partial_i\bar{h}^a_a\partial^i\bar{h}^j_j - \partial_a\bar{h}_{bc}\partial^a\bar{h}^{bc} + 2\partial_a\bar{h}_{bc}\partial^c\bar{h}^{ba}\right], \\ \bar{h}_{ab} &\equiv h_{ab} - \frac{1}{2}\eta_{ab}h^i_i. \end{aligned} \tag{A.6}$$

We shall use the shorter notation

$$A_h = \frac{1}{4}\int d^4x\, \partial_a h_{bc}\partial_i h_{jk} M^{abcijk}(\eta^{mn}), \tag{A.7}$$

where the tensor $M^{abcijk}(\eta^{mn})$ is symmetric in bc, jk and under the triple exchange $(a,b,c) \leftrightarrow (i,j,k)$, and is given by

$$M^{abcijk}(\eta^{mn}) = \left[\eta^{ai}\eta^{bc}\eta^{jk} - \eta^{ai}\eta^{bj}\eta^{ck} + 2\eta^{ak}\eta^{bj}\eta^{ci} - 2\eta^{ak}\eta^{bc}\eta^{ij}\right]_{\text{symm}}, \tag{A.8}$$

where the subscript "symm" indicates that the expression inside the square brackets should be suitably symmetrized in bc, jk and under the exchange $(a,b,c) \leftrightarrow (i,j,k)$. In the expression for the action, since M^{abcijk} is multiplied by $\partial_a h_{bc}\partial_i h_{jk}$, we need not worry about symmetrization and use the expression given inside the square brackets in Eq. (A.8) as it is.

A.2. Gauge conditions and true degrees of freedom

The gauge invariance which was imposed to obtain the action in Eq. (A.6) implies that we are dealing with redundant degrees of freedom in h_{ab} and — without additional restrictions — it does not carry pure spin 2, in the sense of irreducible representations of the Lorentz group. To see this explicitly, consider a h_{ik} of the form

$$h_{ik}(x) = Q_{ik}(x) + \partial_i A_k(x) + \partial_k A_i(x) + \left(\partial_i\partial_k - \frac{1}{4}\eta_{ik}\partial^2\right)\alpha(x) + \frac{1}{4}\eta_{ik}\beta(x), \tag{A.9}$$

where

$$\partial_i Q^{ik} = 0, \quad \eta_{ik}Q^{ik} = 0, \quad \partial_i A^i = 0, \tag{A.10}$$

so that h_{ik} (10 components) is separated into a transverse traceless tensor $(10-5=5$ components), a transverse vector $(4-1=3$ components) and two scalars α and β. The action in Eq. (A.6) now becomes

$$A_h = -\frac{1}{2}\int d^4x \left[\partial_a Q_{bc}\partial^a Q^{bc} - \frac{3}{8}\partial_i\epsilon\partial^i\epsilon\right], \quad \epsilon \equiv (\beta - \partial^2\alpha). \tag{A.11}$$

[This is most easily proved in the Fourier space using the third line of Eq. (A.6).] This expression shows that: (a) the action is independent of the vector degree of freedom, as one would have guessed from the gauge invariance; (b) it does depend on the scalar $\epsilon(x)$ but ϵ and Q^{ab} are decoupled from each other; (c) the residual scalar appears with the wrong sign for the kinetic energy term. Therefore, isolating the physical degrees of freedom requires imposing the conditions $h^a_a = 0$ (to set $\epsilon = 0$) and $\partial_a h^a_b = 0$ (to ensure the transverse–traceless condition on Q^{ab}). If we impose these gauge conditions *in the action itself*, it becomes

$$\begin{aligned} A_h &= \frac{1}{4}\int d^4x \partial_a h_{bc}\partial_i h_{jk}[-\eta^{bj}\eta^{ck}\eta^{ai} + 2\eta^{bj}\eta^{ci}\eta^{ak}] \\ &= \frac{1}{4}\int d^4x[-\partial_a h_{bc}\partial^a h^{bc} + 2\partial_a h_{bc}\partial^c h^{ba}]. \end{aligned} \tag{A.12}$$

We mentioned before that in the original action in Eq. (A.6) there was an ambiguity with respect to the two terms c_3 and c_4. This nonuniqueness disappears in

Eq. (A.12). There is no possibility of "swapping" the derivatives in the first term (since both are ∂_a); in the second term, swapping the derivatives will give a vanishing term because $\partial_a h^{ba} = 0$. We shall see in App. B that this action generalizes to an interesting nonlinear theory.

There is alternative way of interpreting the gauge conditions (in both electromagnetism and gravity) which is probably more physical. In electromagnetism, one can obtain the standard wave equation with a source $\Box A^k = J^k$, if one starts with a kinetic term proportional to $-\partial_i A_j \partial^i A^j$ and an interaction term proportional to $A_i J^i$. But note that, in the time derivative term in the kinetic energy for the field, $-\partial_0 A_j \partial^0 A^j = \partial_0 A_j \partial_0 A^j$, the 0 component comes with the wrong sign (if the spatial components have the correct sign). This degree of freedom will make the Hamiltonian unbounded (for example, free wave solutions for A_0 will carry negative energy) and we need to eliminate this degree of freedom from the theory. This is indeed possible when we introduce the gauge invariance and start with a gauge-invariant Lagrangian. The situation in the case of the spin 2 field is similar. We need to first eliminate unphysical degrees of freedom (which carry negative energy in classical theory and appear as ghosts in quantum theory) by eliminating the scalar and vector modes. (From the viewpoint of the Lorentz group we require the field to correspond to the conventional irreducible representation; so spin 1 and spin 0 parts of the field need to be removed by gauge conditions.) The remaining five degrees of freedom in Q^{ab} can be further reduced to the two physical degrees of freedom of a graviton by using a residual gauge transformation of the form in Eq. (A.5) with $\Box \xi^a = 0$; see e.g. p. 946 of Ref. 13). This will lead to a propogating graviton with two physical degrees of freedom.

A.3. Gauge invariance and conservation of the source

The symmetries of the theory are easier to see in the momentum space, which can be done by introducing the Fourier components $f_{ab}(p)$ of $h_{ab}(x)$, defined as usual by

$$h_{ab}(x) \equiv \int \frac{d^4p}{(2\pi)^4} f_{ab}(p) e^{ipx}. \tag{A.13}$$

The action becomes

$$A_h = \frac{1}{4} \int \frac{d^4p}{(2\pi)^4} f_{bc} f^*_{jk} \left[p_a p_i M^{abcijk}(\eta^{mn})\right] \equiv \frac{1}{4} \int \frac{d^4p}{(2\pi)^4} f_{bc} f^*_{jk} N^{bcjk}, \tag{A.14}$$

with

$$\begin{aligned} N^{bcjk} &= \left[p^2(\eta^{bc}\eta^{jk} - \eta^{bj}\eta^{ck}) + 2p^k(p^c\eta^{bj} - p^j\eta^{bc})\right]_{\text{symm}} \\ &= \left(p^2\eta^{bc}\eta^{jk} - p^k p^j \eta^{bc} - p^c p^b \eta^{jk}\right) - \frac{p^2}{2}\left(\eta^{bj}\eta^{ck} + \eta^{cj}\eta^{bk}\right) \\ &\quad + \frac{1}{2}\left(p^k p^c \eta^{bj} + p^j p^c \eta^{bk} + p^k p^b \eta^{cj} + p^j p^b \eta^{ck}\right), \end{aligned} \tag{A.15}$$

where we have exhibited the symmetrized expression in full glory for once. [N^{bcjk} is symmetric in bc, jk and under the pair exchange $(b,c) \leftrightarrow (j,k)$.] The gauge transformation of the spin 2 field, given by Eq. (A.5), is equivalent to $f_{ab} \to f_{ab} + p_a\xi_b + p_b\xi_a$ in the Fourier space. Using this in Eq. (A.14) we get

$$A_h \to \frac{1}{4}\int \frac{d^4p}{(2\pi)^4}(f_{bc} + 2p_b\xi_c)(f^*_{jk} + 2p_j\xi^*_k)N^{bcjk}. \tag{A.16}$$

Straightforward computation now shows that $p_bN^{bcjk} = 0 = p_jN^{bcjk}$, making A_h invariant under the gauge transformations. This is, of course, built in by the choice of the coefficients c_i in the original action. This condition translates, in coordinate space, to the *identity* $M^{abcijk}\partial_b\partial_a\partial_i h_{jk} = 0$.

Appendix B. Comment on the Uniqueness

We briefly comment on two issues in this Appendix, postponing their detailed discussion to a future publication.

First, it may be noted that the action for the spin 2 graviton had an ambiguity in the form a four-divergence term A_{div} in Eq. (A.4). While the action in Eq. (4) correctly leads to the full Einstein theory under the substitution $d^4x \to d^4x\sqrt{-g}, \eta_{ab} \to g_{ab} = \eta_{ab} + \lambda h_{ab}$, the four-divergence term in Eq. (A.4) does not map to a four-divergence term under these substitutions:

$$\begin{aligned} A_{\text{div}} &= \frac{c_3}{4}\int d^4x\,\partial_a h_{bc}\partial_i h_{jk}\left[\eta^{ab}\eta^{ck}\eta^{ij} - \eta^{ak}\eta^{bj}\eta^{ci}\right] \\ &\to \frac{c_3}{4}\int d^4x\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk}\left[g^{ab}g^{ck}g^{ij} - g^{ak}g^{bj}g^{ci}\right], \end{aligned} \tag{B.1}$$

which cannot be expressed as a four-divergence. So, we obtain the correct Einstein theory only if we start with the correct set of terms in the linear spin 2 case. In general, consider two different actions $A^{\text{lin}}_{\text{I}}$ and $A^{\text{lin}}_{\text{II}}$ in the linear theory, which differ by a four-divergence. The substitutions $d^4x \to d^4x\sqrt{-g}, \eta_{ab} \to g_{ab} = \eta_{ab} + \lambda h_{ab}$ can be used with either $A^{\text{lin}}_{\text{I}}$ or $A^{\text{lin}}_{\text{II}}$ and will lead to nonlinear theories with the action A^{nl}_{I} or $A^{\text{nl}}_{\text{II}}$. But A^{nl}_{I} and $A^{\text{nl}}_{\text{II}}$ cannot be related by a four-divergence, in general.

Second, let us consider the spin 2 theory obtained by imposing the gauge conditions $h^a_a = 0, \partial_a h^a_b = 0$ in the action itself. The linear theory action was given by Eq. (A.12) and, as was pointed out before, there is no ambiguity as regards four-divergences in this action. Our "rule" now leads to the nonlinear action

$$A_\infty = \frac{1}{4\lambda^2}\int d^4x\,\sqrt{-g}\,\partial_a g_{bc}\partial_i g_{jk}\left[-g^{bj}g^{ck}g^{ai} + 2g^{bj}g^{ci}g^{ak}\right]. \tag{B.2}$$

To get the field equations from this action, we only need to note that it is the same as the standard Γ^2 action in Eq. (B.2), with the extra condition $g_{ab}dg^{ab} = 0$. Hence the field equations resulting from varying g_{ab} in Eq. (B.2) will be the same as those obtained from varying g_{ab} in the action in Eq. (23), keeping $\sqrt{-g}$ = constant. It is well known that we then get the equations

$$R_{ab} - \frac{1}{4}g_{ab}R = 8\pi\left(T_{ab} - \frac{1}{4}g_{ab}T\right), \tag{B.3}$$

in which both sides are trace-free. Bianchi identity can now be used to show that $\partial_a(R + 8\pi T) = 0$, requiring $R + 8\pi T$ = constant. Thus the cosmological constant arises as an (undetermined) integration constant in such models (the model has a long history; for a sample of references see Refs. 59–66) and could be interpreted as a Lagrange multiplier that maintains the condition $\sqrt{-g}$ = constant. Clearly, the gauge conditions translate to $\sqrt{-g}$ = constant in the full theory, eliminating the scalar degree of freedom.

It should be stressed that, while this theory is mathematically the same as Einstein's gravity with a cosmological constant, it is conceptually quite different from the usual approach to the cosmological constant. What is probably more interesting is that this theory takes gravity one step closer to gauge theories in the following sense: it has been known for a long time that the Christoffel symbols in gravity Γ^c_{ab} can be thought of as the elements of the matrix $(\Gamma_a)^c_b$ in exact analogy with the matrix representation of the gauge field $(A_i)^j_k$. The Riemann–Christoffel tensor can then be given the matrix representation

$$R_{ab} = \partial_a\Gamma_b - \partial_b\Gamma_a + \Gamma_a\Gamma_b - \Gamma_a\Gamma_b \tag{B.4}$$

with two matrix indices suppressed, etc. Everyone who followed this route soon realized that — unfortunately — we need to contract on a matrix index with a space–time index to get the Einstein action, etc. In the above approach, if we take $\sqrt{-g} = 1$, then the quadratic action can be expressed in the form

$$A_{\text{quad}} = \frac{1}{4\lambda^2}\int d^4x\, g^{ab}Tr[\Gamma_a\Gamma_b], \quad \sqrt{-g} = 1. \tag{B.5}$$

We hope to discuss the gauge theory connection arising from this approach in a future publication.

References

1. M. Fierz and W. Pauli, *Proc. Roy. Soc. A* **173** (1939) 211.
2. S. N. Gupta, *Phys. Rev.* **96** (1954) 1683.
3. R. H. Kraichnan, *Phys. Rev.* **98** (1955) 1118.
4. R. H. Kraichnan, *Phys. Rev.* **101** (1955) 482.
5. R. P. Feynman, F. B. Morinigo and W. G. Wagner, *Feynman Lectures on Gravitation* (Addison-Wesley, 1995).
6. S. Deser, *Gen. Relativ. Gravit.* **1** (1970) 9.
7. S. Weinberg, *Phys. Lett.* **9** (1964) 357.
8. S. Weinberg, *Phys. Rev. B* **135** (1964) 1049.
9. D. G. Boulware and S. Deser, *Ann. Phys.* **89** (1975) 193.
10. L. P. Grishchuk, A. N. Petrov and A. D. Popova, *Commun. Math. Phys.* **94** (1984) 379.

11. R. M. Wald, *Phys. Rev. D.* **33** (1986) 3613.
12. S. V. Babak and L. P. Grishchuk, *Phys. Rev. D* **61** (1986) 024038.
13. C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).
14. L. Rosenfeld, *Acad. Roy. Mem. Class. Sci.* **18** (1940) 1.
15. A. Papapetrou, *Proc. Roy. Irish Acad.* **52** (1948) 11.
16. P. G. Bergmann and R. Thompson, *Phys. Rev.* **89** (1953) 400.
17. J. N. Goldberg, *Phys. Rev.* **111** (1958) 315.
18. C. Møller, *Ann. Phys.* **4** (1958) 347.
19. A. Komar, *Phys. Rev.* **113** (1959) 934.
20. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 2nd edn. (Addison-Wesley, Massachusetts, 1962).
21. R. Arnowitt, S. Deser and C. W. Misner, in *Gravitation: An Introduction to Current Research*, ed. L. Witten (John Wiley & Sons, 1962), pp. 227–265.
22. A. Trautman, in *Gravitation: An Introduction to Current Research*, ed. L. Witten (John Wiley & Sons, 1962), pp. 169–198.
23. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
24. B. F. Schutz and R. Sorkin, *Ann. Phys.* **107** (1977) 1.
25. R. P. Wallner, *Acta Physica Austriaca* **52** (1980) 121.
26. D. Bak, D. Cangemi and R. Jackiw, *Phys. Rev. D* **49** (1994) 173.
27. J. Katz, J. Bičák and D. Lynden-Bell, *Phys. Rev. D* **55** (1997) 5957.
28. G. Magnano and L. M. Sokolowski, *Gen. Relativ. Gravit.* **30** (1998) 1281.
29. W. Heisenberg, H. Euler, *Z. Phys.* **98** (1936) 714 [English translation available in arXiv as physics/0605038].
30. J. Schwinger, *Phys. Rev.* **82** (1951) 664.
31. T. Padmanabhan, *Pramana* **37** (1991) 179.
32. C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
33. C. Kiefer *et al.*, *Class. Quant. Grav.* **8** (1991) L185.
34. T. Padmanabhan, gr-qc/0311036.
35. T. Padmanabhan, *Gen. Relativ. Gravit.* **34** (2002) 2029 [gr-qc/0205090].
36. T. Padmanabhan, *Gen. Relativ. Gravit.* **35** (2003) 2097.
37. T. Padmanabhan, *Mod. Phys. Lett. A* **17** (2002) 923 [gr-qc/0202078].
38. T. Padmanabhan, *Mod. Phys. Lett. A* **17** (2002) 1147 [hep-th/0205278].
39. T. Padmanabhan, *Class. Quant. Grav.* **19** (2002) 5387 [gr-qc/0204019].
40. T. Padmanabhan, *Class. Quant. Grav.* **21** (2004) 4485 [gr-qc/0308070].
41. T. Padmanabhan, *Mod. Phys. Lett. A* **18** (2003) 2903 [hep-th/0302068].
42. T. Padmanabhan, *Mod. Phys. Lett. A* **19** (2004) 2637 [gr-qc/0405072].
43. T. Padmanabhan, *Int. J. Mod. Phys. D* **13** (2004) 2293 [gr-qc/0408051].
44. T. Padmanabhan and A. Patel, gr-qc/0309053.
45. V. I. Ogievetsky and I. V. Polubarinov, *Ann. Phys.* **35** (1965) 167.
46. W. Thirring, *Ann. Phys.* **16** (1961) 96.
47. G. Scharf, *Quantum Gauge Theories: A True Ghost Story* (John Wiley, 2001).
48. G. 't Hooft and M. Veltman, *Phys. Rev. D* **10** (1974) 411.
49. Y. V. Baryshev, gr-qc/9912003.
50. C. Wetterich, *Gen. Relativ. Gravit.* **30** (1998) 159.
51. F. Belinfante, *Physica* **6** (1939) 887.
52. G. Magnano and L. M. Sokolowski, *Class. Quant. Grav.* **19** (2002) 223 [gr-qc/0103072].
53. T. Padmanabhan, *Int. J. Mod. Phys. D* **15** (2006) 1659 [gr-qc/0606061].
54. T. Padmanabhan, *Gen. Relativ. Gravit.* **38** (2006) 1547.

55. T. Padmanabhan, gr-qc/0705.2533.
56. T. Padmanabhan, gr-qc/0706.1654.
57. T. Padmanabhan and A. Paranjape, *Phys. Rev. D* **75** (2007) 064004 [gr-qc/0701003].
58. A. Mukhopadhyay and T. Padmanabhan, *Phys. Rev. D* **74** (2006) 124023 [hep-th/0608120].
59. A. Einstein, *Siz. Preuss. Acad. Scis.* (1919), translated as "Do gravitational fields play an essential role in the structure of elementary particles of matter?" in *The Principle of Relativity*, eds. A. Einstein *et al.* (Dover, New York, 1952).
60. J. J. van der Bij *et al.*, *Physica A* **116** (1982) 307.
61. F. Wilczek, *Phys. Rep.* **104** (1984) 111.
62. A. Zee, in *High Energy Physics*, *Proc. 20th Annual Orbis Scientiae* (*Coral Gables, 1983*), eds. B. Kursunoglu, S. C. Mintz and A. Perlmutter (Plenum, New York, 1985).
63. W. Buchmuller and N. Dragon, *Phys. Lett. B* **207** (1988) 292.
64. W. G. Unruh, *Phys. Rev. D* **40** (1989) 1048.
65. S. Weinberg, *Rev. Mod. Phys.* **61** (1989) 1.
66. T. Padmanabhan, *Phys. Rep.* **380** (2003) 235 [hep-th/0212290, Sec. 7.2].