Article

Toward a Mechanism for the Emergence of Gravity

Carlos Barceló [1,*,†] ⓘ, Raúl Carballo-Rubio [2,†], Luis J. Garay [3,4,†] ⓘ and Gerardo García-Moreno [1,3,†] ⓘ

1 Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía, 18008 Granada, Spain; gerargar@ucm.es

2 Florida Space Institute, University of Central Florida, 12354 Research Parkway, Partnership 1, Orlando, FL 32826, USA; raul.carballorubio@ucf.edu

3 Departamento de Física Teórica and IPARCOS, Universidad Complutense de Madrid, 28040 Madrid, Spain; luisj.garay@ucm.es

4 Instituto de Estructura de la Materia (IEM-CSIC), Serrano 121, 28006 Madrid, Spain

* Correspondence: carlos@iaa.es

† All the authors contributed equally to this work.

Abstract: One of the main problems that emergent-gravity approaches face is explaining how a system that does not contain gauge symmetries ab initio might develop them effectively in some regime. We review a mechanism introduced by some of the authors for the emergence of gauge symmetries in [JHEP 10 (2016) 084] and discuss how it works for interacting Lorentz-invariant vector field theories as a warm-up exercise for the more convoluted problem of gravity. Then, we apply this mechanism to the emergence of linear diffeomorphisms for the most general Lorentz-invariant linear theory of a two-index symmetric tensor field, which constitutes a generalization of the Fierz–Pauli theory describing linearized gravity. Finally we discuss two results, the well-known Weinberg–Witten theorem and a more recent theorem by Marolf, that are often invoked as no-go theorems for emergent gravity. Our analysis illustrates that, although these results pinpoint some of the particularities of gravity with respect to other gauge theories, they do not constitute an impediment for the emergent gravity program if gauge symmetries (diffeomorphisms) are emergent in the sense discussed in this paper.

Keywords: analogue gravity; emergent gravity; emergent gauge symmetries; gravitons

1. Introduction

The pursuit of a theory able to combine the principles of general relativity and quantum mechanics is one of the strongest driving forces in modern fundamental physics. Almost a century of active research on this problem has resulted in many insights [1], although a full comprehension of how both sets of principles might intertwine is still lacking. Approaches toward finding a theory of quantum gravity can be divided in emergent and non-emergent approaches. The first category corresponds to approaches in which the fundamental degrees of freedom are not taken to be the spacetime geometry itself, but other sort of microscopic degrees of freedom. The geometric structure characteristic of general relativity is assumed to be recovered in a suitable regime, for instance a weakly-coupled regime or a low-energy regime. String theory [2,3] would fall into this category, as it is a specific proposal for these microscopic degrees of freedom. More generally, one can conceive systems having another substratum for the emergence, for instance condensed-matter systems [4–7]. The second category corresponds to theories in which the geometric structure of general relativity is taken as fundamental. Different approaches in this category are set apart by the different quantization schemes applied to this structure. The most widely studied theory within this category is loop quantum gravity [8]. The distinction between the two categories above is nevertheless not a sharp one. In fact, the notion of emergence is inevitably tied up to every quantum gravitational theory, since one can show on general

grounds that there are no local observables in the Dirac sense [9] (that is, diffeomorphism invariant) for any quantum gravity theory [10].

This work is devoted to analyzing a specific kind of system within the emergent category. The motivation for this study comes from the observation that it is quite common to find analogues of gravitational fields in many physical systems, especially in condensed-matter systems [6]. However, this analogy just holds at a kinematical level, meaning that the effective geometry emerging in those systems does not display the dynamical features characteristic of general relativity. Thus, a natural question that arises is whether these analogue systems can be designed to develop the correct dynamics. This is the notion of emergent gravity that we explore in this work. Our goal is finding the conditions that make such analogue systems develop fully relativistic dynamics in some regime.

The emergence of analogue gravity usually comes associated with a low-energy regime in which the excitations can be typically described by a collection of weakly-coupled relativistic fields. They describe the collective excitations of this system [7]. In condensed-matter systems, this low-energy regime is usually characterized by the presence of Fermi points [4,5]. Around these Fermi points, we know the field content relevant for describing the low-energy regime, and we know that the fields are weakly coupled. This low-energy Lorentz invariance becomes less exact as we increase the energy. It typically manifests itself in modified dispersion relations of the form

$$E^2 = k^2 + \sum_n \zeta_n \frac{k^{2n+2}}{E_L^{2n}}, \tag{1}$$

for instance, but also on the very dissolution of the effective description. The dispersion relations by themselves offer a problem from an effective field theory rationale since they appear to generate a fine-tuning problem (see [11]). Here we do not delve into this problem, which we think arises because of the tacit assumption that the effective field description survives far in the dispersive regime, and just assume the existence of a low-energy Lorentzian regime. This ensures that we can apply the rules of effective field theory and write down the most general effective action compatible with the symmetries of the system in order to obtain a good description of the phenomena of interest [12,13]. However, these arguments cannot guarantee that a vector field A^μ or a symmetric tensor $h^{\mu\nu}$ will develop gauge symmetries. It is well known that gauge symmetries prevent the propagation of ghost states and hence eliminate the potential instabilities of such theories. The absence of gauge symmetries would leave us with a potentially unstable theory. Hence, the low-energy description needs to be endowed with a mechanism that allows the effective decoupling of the dangerous degrees of freedom. We discuss here a specific mechanism for this decoupling, first introduced by some of us in [14] for an extended version of (quantum) electrodynamics.

Here is a brief outline of this work. In Section 2, we review on general grounds the mechanism for the emergence of gauge symmetries in systems displaying no gauge symmetries ab initio. Section 3 is devoted to discussing some examples. In Section 3.1, we review how this mechanism works for an extended version of linearized Yang–Mills theory and how a non-linear completion of such theory can be obtained through a bootstrapping procedure, which was studied in detail in [15] where further details can be found. In Section 3.2, we consider the most general linear theory that can be written down for a symmetric tensor field $h^{\mu\nu}$. We study the mechanism for the emergence of gauge symmetries and find the emergence of the closest gauge symmetries to diffeomorphisms and WTDiff within this framework. We discuss the relation of this result with the emergence of linearized gravity and discuss ongoing work on the possible non-linear completions of the theory. In Section 4, we revisit some theorems, namely the Weinberg–Witten theorem in Sections 4.1 and 4.2 that are often invoked as no-go theorems for emergent gravity and discuss their limitations and interplay with our results. We conclude in Section 5. Furthermore, we have added Appendix A where we discuss some specific cases associated with particular

choices of parameters for the $h^{\mu\nu}$-theory from Section 3.2 that make the discussion of the emergence of gauge symmetries slightly different.

Notation and conventions. We work in four-dimensional spacetime and we use the signature $(-,+,+,+)$. The symbol ∇ represents the covariant derivative compatible with the flat metric of Minkowski spacetime $\eta_{\mu\nu}$. We use greek $(\mu, \nu...)$ indices for the spacetime indices and latin indices $(a, b...)$ for the internal indices of vector fields.

2. A Mechanism for the Emergence of Gauge Symmetries

The first step in the logic of our construction is to consider a non-relativistic condensed-matter-like theory that in some regime can be accurately described in terms of a set of Lorentz-invariant and weakly coupled fields over a flat background. The emergence of Lorentz invariance in this kind of systems is typically associated with the presence of Fermi-like points in the theory [4,5,7]. Here, we take for granted this first step and analyze under which conditions this effective field theoretic description develop emergent gauge symmetries.

We will work in configuration space to avoid introducing Dirac's machinery for the classification of constrained systems [9]. Thus, we need a characterization of gauge symmetries within this formulation. Such characterization is offered by Noether's theorem [16]. Generally, given a symmetry we can find a conserved current associated with it. If the symmetry is a gauge symmetry, the Noether charges associated with the corresponding current are trivial. Hence, for a gauge symmetry the current can always be written as

$$J^{\mu} = W^{\mu} + S^{\mu}, \tag{2}$$

with W^{μ} being a term that is zero on shell and $S^{\mu} = \nabla_{\nu}N^{\nu\mu}$ a superpotential, i.e., an identically conserved term. Once the theory is provided with suitable fall-off conditions, the contribution of the superpotential term to the charges identically vanishes, as it can be seen by applying the Gauss theorem [17].

This characterization of gauge symmetries is the starting point for our discussion of emergent gauge symmetries. Let us consider a field theory characterized by a set of fields, which we call Φ, with a well-posed initial value problem. For the sake of simplicity, we assume that such system is free of gauge symmetries from the beginning. There always exists a set of symmetries whose Noether charges parametrize the space of initial conditions [18], and hence the whole space of solutions which we call $\mathcal{S}$. These symmetries are in general complicated symmetries that cannot be found explicitly; they can only be found for integrable systems. Knowing that they exist is enough for our purposes.

Now let us consider a set of constraints defined by the condition $\Psi = 0$. Here, Ψ represents a set of equations that combine the fields Φ and their derivatives. Let us represent the subspace defined by this constraint as $\mathcal{U} \subset \mathcal{S}$. We demand such a space to be non-trivial in the sense of having more than one element. This means that there exist solutions for the constraint $\Psi = 0$, apart from the trivial one in which all the fields vanish, $\Phi = 0$.

Let us denote $\mathcal{Q}$ the complete set of charges that parametrizes the space of solutions $\mathcal{S}$. It is convenient to choose a parametrization of the space of charges, i.e., a set of coordinates in charge space, such that the subspace $\mathcal{U}$ is characterized by the condition that some of the charges vanish. In other words, we choose a subset of charges $\mathcal{Q}_{\Psi}^{\perp} \subset \mathcal{Q}$ such that $\Psi = 0$ if and only if $\mathcal{Q}_{\Psi}^{\perp} = 0$. Furthermore, it is useful for our purposes to choose the remaining coordinates, which we denote as $\mathcal{Q}_{\Psi}^{\parallel}$, in such a way that they leave the subspace $\mathcal{U}$ untouched. This means that the symmetries associated with the charges $\mathcal{Q}_{\Psi}^{\parallel}$ can be defined within $\mathcal{U}$ since they preserve the condition $\Psi = 0$. Now we have two possibilities depending on the values of the charges $\mathcal{Q}_{\Psi}^{\parallel}$ within the subspace $\mathcal{U}$.

Non-emergence of gauge symmetry: This corresponds to the standard situation in which the charges $\mathcal{Q}_{\Psi}^{\parallel}$ parametrize the space $\mathcal{U}$ without redundancies. In this case, we will

have that the projection onto the subspace $\mathcal{U}$ simply eliminates the degrees of freedom directly encoded in the conditions $\Psi = 0$.

Emergence of gauge symmetry: This case corresponds to the situation in which some of the charges $\mathcal{Q}_\Psi^\parallel$ vanish when we restrict to $\mathcal{U}$. In this case, not all the solutions in $\mathcal{U}$ can be distinguished by just using operations within $\mathcal{U}$. Hence, the solutions related by a finite transformation associated with the charges $\mathcal{Q}_\Psi^\parallel$ can be interpreted as describing a single physical configuration. A pictorical representation of this situation can be found in Figure 1.
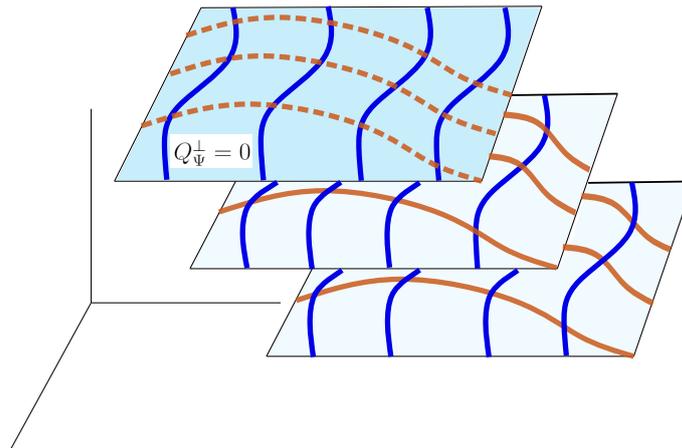

Figure 1. We represent the space of solutions $\mathcal{S}$ and the set of surfaces characterized by constant values of $\mathcal{Q}_\Psi^\perp$. Among them, there is a privileged one represented in darker blue characterized by $\mathcal{Q}_\Psi^\perp = 0$, i.e., it represents the subspace $\mathcal{U}$ from the discussion above. The solid blue and brown curves represent curves of constant $Q_\Psi^\parallel$ within the surfaces of $\mathcal{Q}_\Psi^\perp = $ constant. The dashed brown curves within the subspace $\mathcal{U}$ represent the emergent gauge orbits, i.e., they connect points that cannot be distinguished by performing operations within $\mathcal{U}$ and are characterized by the same vanishing value of some of the charges $Q_\Psi^\parallel$ in $\mathcal{U}$.

Interpreting these configurations as a single physical configuration means that these finite transformations would correspond to emergent gauge symmetries. Furthermore, we would interpret the orbits generated by these transformations as emergent gauge orbits. This translates into the elimination of entire dynamical degrees of freedom, each corresponding to a pair of initial conditions or charges that parametrize $\mathcal{S}$. Notice that this mechanism relies on the naturalness of the condition $\Psi = 0$ in some regime of the theory, e.g., a low-energy regime or a weakly-coupled one. We will come back to this point when considering specific examples below.

This mechanism has been implemented successfully for Maxwell [14] and linearized Yang–Mills theories [15]. To extend this mechanism to non-linear theories, in [15] we demanded the existence of a bootstrapping procedure that determines the self-coupling of the linear theory, thus connecting smoothly the linear and non-linear theories. In this paper, we start exploring the application of this mechanism to gravity.

3. Two Examples: Yang–Mills Theories and Linearized Gravity

3.1. Yang–Mills Theories

Let us present here an extended version of Yang–Mills theory that will allow us to illustrate the emergence of gauge symmetries. Our starting point is the most general quadratic and Lorentz-invariant Lagrangian giving rise to second-order differential equations that can be written for a set of N vector fields A_μ^a, with the latin indices $(a, b, c...)$ running from 1 to N:

$$S_2 = \int d^4x \sqrt{-\eta} \left[-\frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} + \frac{1}{2} \xi_{ab} (\nabla_\mu A^{a\mu})(\nabla_\nu A^{b\nu}) - \frac{1}{2} M_{ab} A^{a\mu} A_\mu^b + A_{a\mu} j^{a\mu} \right], \quad (3)$$

up to a boundary term which we skip here. We have introduced the notation $F^a_{\mu\nu} = 2\nabla_{[\mu}A^a_{\nu]}$, where ∇ is the affine connection compatible with $\eta_{\mu\nu}$, and we have introduced the non-degenerate matrix ξ_{ab} and the mass matrix M_{ab}. The case which is especially interesting for us is the massless case $M_{ab} = 0$, since it is a generic feature that field theories emerging around Fermi points in condensed-matter systems give rise to massless excitations [5]. Masses can be acquired in a second step through a Higgs-like mechanism. However, for the moment, we will keep the M_{ab} term since it is instructive to study the massive and massless cases as representatives of the two possible situations described in the previous section regarding the mechanism for the emergence of gauge symmetries. The equations of motion can be derived straightforwardly from action (3) by taking a simple functional derivative,

$$\nabla_\mu F^{\mu\nu}_b - \xi_{ab}\nabla^\nu\nabla_\mu A^{a\mu} - M_{ab}A^{a\nu} = j^\nu_b. \tag{4}$$

It is possible to identify a set of symmetries for this theory associated with the following transformations:

$$A^a_\mu \to A^a_\mu + \nabla_\mu\chi^a, \qquad j^{a\mu} \to j^{a\mu}, \tag{5}$$

where χ^a are functions that need to obey

$$(\xi_{ab}\Box + M_{ab})\chi^a = 0. \tag{6}$$

The currents associated with these symmetries, once evaluated on shell, can be written as follows [15]:

$$J^\mu|_{\text{on shell}} = -\nabla_\nu(F^{a\mu\nu}\chi_a) + \xi_{ab}\left(\nabla_\nu A^{a\nu}\nabla^\mu\chi^b - \chi^b\nabla^\mu\nabla_\nu A^{\nu a}\right) \tag{7}$$

Notice that the first part of this equation is a superpotential and hence it does not contribute to the charges. Furthermore, the second term vanishes if we restrict ourselves to the subspace $\mathcal{U}$ of solutions of Equation (4) satisfying $\nabla_\mu A^{a\mu} = 0$. Hence, projecting onto the subspace $\mathcal{U}$ makes the current acquire the form of a gauge current encapsulated in Equation (2). Notice that although Equation (7) is associated with a transformation whose generators χ^a are spacetime-dependent local functions, it does not constitute a gauge transformation a priori. Gauge transformations are associated with local transformations whose generators are arbitrary functions of spacetime. However, in our case they are constrained to obey Equation (6). Thus, the transformation we are discussing does not constitute a gauge transformation from the beginning although it is generated by local functions. A further discussion of this point with a concrete example can be found in Sections 2.1 and 2.2 from [14].

The projection introduced above is quite natural once we consider the dynamical equations that the scalar fields $\varphi^a = \nabla_\mu A^{a\mu}$ obey, which can be obtained by taking the divergence in Equation (4):

$$\left[\Box\delta^c_a + \left(\xi^{-1}\right)^{cb}M_{ab}\right]\varphi^a = 0. \tag{8}$$

We have a set of sourceless Klein–Gordon equations for these scalars φ^a, and hence it is quite natural to restrict ourselves to the subspace $\mathcal{U}$ defined by $\varphi^a = 0$, since there are no Lorentz-invariant sources that might generate excitations of these scalars. From this point onwards, there are two possible situations which we discuss below.

Massive case $M_{ab} \neq 0$. In this case no gauge symmetries emerge, since the putative emergent gauge symmetries from Equations (5) and (6) do not leave $\mathcal{U}$ invariant. Actually, they translate into the following map for φ^a

$$\varphi^a \to \varphi^a - \left(\xi^{-1}\right)^{ab}M_{bc}\chi^c, \tag{9}$$

which takes $\varphi^a = 0$ to a non-vanishing value for $\varphi^a \neq 0$.

Massless case $M_{ab} = 0$. In this case gauge symmetries emerge naturally, since the transformations from Equations (5) and (6) become gauge symmetries by leaving the scalar fields φ^a invariant (note that the last term of (9) now vanishes).

It is better to pause at this point to summarize what we have obtained so far and what are we pursuing for the non-linear theory. We began with the most general Lorentz-invariant vector field theory for a set of N vector fields A_μ^a. If these vector fields represent massless excitations, we have shown that the coupling to a conserved current leads to a set of sourceless Klein–Gordon equations for the scalar sector of the theory. Since no excitations can be produced within that subspace, we have removed the scalar sector by projecting onto the subspace $\varphi^a = 0$. The resulting effective field theory that we have found displays emergent gauge symmetries given by Equation (5), with χ^a satisfying $\Box \chi^a = 0$. Notice that these transformations are just the residual gauge symmetries of linearized Yang–Mills theory after the Lorenz gauge is chosen.

The next natural step would be to study non-linear vector-field theories. Among the possible non-linear vector-field theories, Yang–Mills theories are special since their gauge symmetry avoids the propagation of ghosts when one linearizes them. Furthermore, they are such that they can be derived from the linear theory through a bootstrapping procedure in which one analyzes the self-coupling of the theory [19]. However, such procedure does not lead to a unique theory by itself due to the ambiguities inherent to the definition of a Noether current [20,21], at least not with further specifications [22]. The combination of the mechanism introduced here for the emergence of gauge symmetries with the bootstrapping procedure allowed us to make an incursion on the non-linear vector-field theories that are free from instabilities. We present here the main idea and refer the reader to [15] for further technical details.

The idea of the bootstrapping procedure is the following. Consider a linear field theory displaying a symmetry and hence having a conserved current which we call $J_1^{a\mu}$. We assume that our theory couples perturbatively to such current in such a way that we recover a smooth limit for the free theory when we take the coupling constant to be zero. We call such a coupling constant g. Thus, schematically we will have equations of motion of the form

$$\frac{\delta S_2}{\delta A_{a\mu}} = g J_1^{a\mu} + \mathcal{O}\left(g^2\right). \tag{10}$$

To derive the right-hand side from an action principle, we would need to add a suitable interacting term S_3 to the action such that

$$\frac{\delta S_3}{\delta A_{a\mu}} = J_1^{a\mu}. \tag{11}$$

However, notice that S_3, which needs to be also symmetric under the transformations that we introduce for the bootstrapping procedure, would contribute to the conserved current due to Noether's theorem. Hence, we would obtain a contribution to the current of the form $g J_2^{a\mu}$ which we would want to derive from an action principle, from S_4. In this way, we would obtain an infinite set of constraints for the successive actions S_n and S_{n+1}. More concretely, we would have

$$\frac{\delta S_{n+1}}{\delta A_{a\mu}} = J_{n-1}^{a\mu}, \tag{12}$$

where $J_{n-1}^{a\mu}$ is obtained from S_n through Noether's theorem. In this way, it is possible to obtain a potentially infinite set of equations which we would need to solve in order to find the non-linear theory described by the action

$$S = \sum_{n=2}^{\infty} g^{n-2} S_n. \tag{13}$$

There are two points we want to stress. First, we need to remember that conserved currents are defined up to the addition of identically conserved quantities. Hence, at each order in the bootstrapping procedure we have ambiguities that need to be taken into account. For instance, in the gravitational context they are crucial to recover general relativity from Fierz–Pauli theory through this procedure [23]. Second, at each order we need to ensure that the generated contribution to the action S_{n+1} is still invariant under the symmetry transformations that were used to define the current $J^{a\mu}_{n-1}$. This imposes non-trivial constraints on the partial actions [15].

For the case of emergent Yang–Mills theories, we need to consider the following quadratic action as the starting point,

$$S_2 = \int d^4x \sqrt{-\eta} \left[-\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a + \frac{\lambda}{2} (\nabla_\mu A^\mu_a)(\nabla_\nu A^{a\nu}) \right], \tag{14}$$

which is the action from Equation (3) with $M_{ab} = 0$, as we demand for the emergence of gauge symmetries. We have omitted the material content which we will discuss later, and we have restricted ourselves to the case $\xi_{ab} = \lambda \delta_{ab}$ for the sake of simplicity. The symmetries that we can identify to begin the bootstrapping procedure are the following rigid transformations:

$$A^{a\mu} \to A^{a\mu} + f^{abc} A^\mu_b \zeta_c, \tag{15}$$

where ζ_c is the arbitrary real constant and f^{abc} is a constant tensor antisymmetric in the first two indices, i.e., obeying $f^{abc} = f^{[ab]c}$.

It turns out that the bootstrapping procedure imposes two non-trivial constraints that must be fulfilled to be self-consistent [15]. On the one hand, the constants f^{abc} must be the structure constants of a semi-simple compact Lie Algebra for the symmetry in Equation (15) to be also a symmetry of S_3. On the other hand, we need to impose the constraint $\nabla_\mu A^\mu_a = 0$ to avoid breaking the bootstrapping procedure since, if we do not impose it, the equations of motion cannot be derived from an action principle. The resulting action from the bootstrapping procedure can be written as

$$S = \int d^4x \sqrt{-\eta} \left[-\frac{1}{4} \mathcal{F}^a_{\mu\nu} \mathcal{F}^{\mu\nu}_a + \frac{\lambda + g\vartheta}{2} (\nabla_\mu A^\mu_a)(\nabla_\nu A^{a\nu}) \right], \tag{16}$$

where g is the coupling constant introduced above, ϑ is a Lagrange multiplier that enforces the constraints $\nabla_\mu A^\mu_a = 0$, and we have introduced the object

$$\mathcal{F}^a_{\mu\nu} = 2\nabla_{[\mu} A^a_{\nu]} + g f^{bca} A_{b\mu} A_{c\nu}. \tag{17}$$

This resulting theory has the properties that we were looking for. Namely, its $g \to 0$ limit is the free emergent linearized Yang–Mills theory discussed above. The same comment applies to the infinitesimal emergent non-linear gauge transformations, given by

$$A^a_\mu \to A^a_\mu + \nabla_\mu \chi^a + g f^{abc} A_{b\mu} \chi_c \tag{18}$$

with the fields χ^a obeying

$$\Box \chi^a + g f^{abc} A^\mu_b \nabla_\mu \chi_c = 0. \tag{19}$$

These reduce to the emergent gauge symmetries of the linear theory from Equations (5) and (6) in the $g \to 0$. These transformations are such that they preserve the subspace $\mathcal{U}$ defined by the constraint $\nabla_\mu A^{a\mu} = 0$. Finally, we notice that the inclusion of matter is straightforward, since the bootstrapping procedure in the matter sector is disentangled from the bootstrapping procedure in the gauge sector that we have sketched here. The self-consistency constraint that appears in the matter sector is that the fields need to transform under a representation of the gauge group whose associated Lie Algebra has the structure constants f^{abc} that we introduced above. For scalar fields, this procedure

needs two iterations, like for the Yang–Mills case, whereas for fermionic fields it requires just one iteration to be completed.

This has been just a brief summary of how an emergent Yang–Mills theory can be obtained by a suitable combination of the mechanism introduced in Section 2 and the bootstrapping procedure. For further details, see [15]. In the following we will focus on a situation that has not been studied previously.

3.2. Linear Graviton Physics

The discussion of the case for a symmetric tensor field $h^{\mu\nu}$ is completely parallel to the above section. The first step is to write down the most general quadratic Lorentz-invariant action for such a field. This means that the $h^{\mu\nu}$-field is a good variable to describe our effective field theory. For instance, it can represent the deformations of a given condensed-matter-system near a Fermi point, as we have explained in the introduction. Therefore, the resulting effective theory does not need be gauge-invariant: this is definitely asking too much since everything we know about it at this stage is that it can be described by a field $h^{\mu\nu}$. Explicitly, this action has the form

$$S = \int d^4x \sqrt{-\eta} \left[-\frac{1}{4} \nabla_\mu h^{\alpha\beta} \nabla^\mu h_{\alpha\beta} + \frac{1}{2}(1 + \xi_1) \nabla_\mu h^{\mu\alpha} \nabla_\nu h^\nu_\alpha \right.$$
$$\left. -\frac{1}{2}(1 + \xi_2) \nabla_\mu h^{\mu\nu} \nabla_\nu h + \frac{1}{4}(1 + \xi_3) \nabla_\mu h \nabla^\mu h - \frac{1}{4} m_1^2 h_{\mu\nu} h^{\mu\nu} - \frac{1}{4} m_2^2 h^2 \right], \quad (20)$$

where indices are raised and lowered with the flat spacetime metric $\eta_{\mu\nu}$, we have introduced the trace of the field $h^{\mu\nu}$ as $h = h^{\mu\nu} \eta_{\mu\nu}$, and we have introduced the dimensionless parameters ξ_i and the mass parameters m_i.

We can also include a coupling to matter by adding a term of the form $h^{\mu\nu} T_{\mu\nu}$ to the previous Lagrangian, with the properties of $T_{\mu\nu}$ being discussed in more detail below (for the moment we will just assume that it is a symmetric tensor). The equations of motion can be computed by performing a variational derivative, which results in

$$\frac{1}{2} \Box h_{\mu\nu} - (1 + \xi_1) \eta_{\beta(\mu} \nabla_{\nu)} \nabla_\alpha h^{\alpha\beta} + \frac{1}{2}(1 + \xi_2) \left(\nabla_\mu \nabla_\nu h + \eta_{\mu\nu} \nabla^2 \cdot h \right)$$
$$- \frac{1}{2}(1 + \xi_3) \eta_{\mu\nu} \Box h - \frac{1}{2} m_1^2 h_{\mu\nu} - \frac{1}{2} m_2^2 \eta_{\mu\nu} h = T_{\mu\nu}, \quad (21)$$

where we have introduced the notation $\nabla^2 \cdot h = \nabla_\mu \nabla_\nu h^{\mu\nu}$ and we will keep using it throughout this section.

Before moving forward, it is better to pause and discuss the values of the parameters ξ_i and m_i for which the system has a gauge symmetry from the beginning, in order to avoid those cases in the discussion of emergent gauge symmetries. We will follow the discussion of [24]. On the one hand, for $\xi_1 = m_1 = 0$, the equations of motion are invariant under the group of linearized transverse diffemorphisms, whose infinitesimal transformations are generated by a divergenceless vector field χ^μ as follows:

$$h^{\mu\nu} \to h^{\mu\nu} + 2 \nabla^{(\mu} \chi^{\nu)}, \qquad \nabla_\mu \chi^\mu = 0. \quad (22)$$

In addition, for $m_2 = 0$, $\xi_2 = \xi_3 = 0$ we can relax the condition on the divergence for the vector field $\nabla_\mu \chi^\mu = 0$ and recover the full group of linearized diffeomorphisms, i.e., one recovers the Fierz–Pauli theory [25], i.e., transformations generated by arbitrary vector fields χ^μ as

$$h^{\mu\nu} \to h^{\mu\nu} + 2 \nabla^{(\mu} \chi^{\nu)}. \quad (23)$$

We could have considered a set of transformations that is slightly more general than the ones introduced here, namely transformations generated also by a vector field χ^μ as

$$h_{\mu\nu} \to h_{\mu\nu} + 2 \nabla_{(\mu} \chi_{\nu)} + k \eta^{\mu\nu} \nabla_\alpha \chi^\alpha \quad (24)$$

However, these transformations do not add anything new. To understand this, we first notice that when we put the masses equal to zero $m_1 = m_2 = 0$, a field redefinition of the form

$$h^{\mu\nu} \rightarrow h^{\mu\nu} + \lambda\eta^{\mu\nu}h, \tag{25}$$

(with $\lambda \neq -1/4$ in order for it to be invertible) leaves the functional form of the Lagrangian invariant at the expense of changing the parameters ξ_2 and ξ_3 as [24]

$$\xi_2 \rightarrow \xi_2 + 4\lambda + \lambda(4\xi_2 - 2), \tag{26}$$

$$\xi_3 \rightarrow \xi_3 + 2\lambda(2 + 4\xi_3 - \xi_2) + \lambda^2(16\xi_3 - 8\xi_2 + 6). \tag{27}$$

Had we begun with the case $\xi_1 = \xi_2 = \xi_3 = 0$, for which the theory is invariant under the whole set of diffeomorphisms, we would have ended up with new parameters ξ_2 and ξ_3

$$\xi_2 = 2\lambda, \tag{28}$$

$$\xi_3 = 4\lambda + 6\lambda^2. \tag{29}$$

Furthermore, notice that within this new parametrization of field space, the theory is not invariant under the standard transformations of the form (23) but transformations of the form (24) with $k = 2\lambda$. Thus, it is clear that the more general set of transformations of the form (24) is included in the set of transformations of the form of ordinary linear diffeomorphisms (23) upon, possibly, a field redefinition. Hence, we will focus only on the set of transformations (23) since the transformations (24) do not add anything new.

On the other hand, if $m_1 = m_2 = 0$ and the parameters ξ_i satisfy the relations $-2 + \xi_2 - 4\xi_3 = 0$ and $1 - \xi_1 + 2\xi_2 = 0$, we have the following Weyl transformations, generated by a scalar field ϕ, as gauge transformations:

$$h^{\mu\nu} \rightarrow h^{\mu\nu} + \frac{1}{2}\eta^{\mu\nu}\phi. \tag{30}$$

This Weyl symmetry can be combined with transverse diffeomorphisms to build the so-called linearized WTDiff (Weyl transverse-diffeomorphism) invariant gravity by choosing $\xi_1 = 0$, which guarantees that transverse diffeomorphisms are gauge symmetries as explained above. This forces $\xi_2 = -1/2$ and $\xi_3 = -5/8$ (this corresponds precisely to the case of considering the theory with $\xi_1 = 0$ and performing a putative field redefinition of the form (25) with the limiting value $\lambda = -1/4$ for which the transformation becomes singular. Thus, it is not a true field redefinition since one can not recover the trace of h from the transformed field [24]).

The Fierz–Pauli theory (with the whole linearized group of diffeomorphisms) and WTDiff (with linearized Weyl and transverse diffeomorphism transformations) are the two largest possible gauge symmetry groups for a tensor field $h^{\mu\nu}$ [24].

Once we have discussed the choices of parameters for which the theory displays gauge symmetries, let us go back to the discussion of emergent gauge symmetries. We will be assuming that the parameters ξ_i are such that the theory is *generic*, in the sense that the coefficients of the different terms that will appear from now in equations of motion, conserved currents, etc. do not vanish. Furthermore, we will assume that they take generic values different from the ones that endow action (20) with gauge symmetries from the beginning. The discussion of special choices of the parameters has been moved to Appendix A since it just involves the analysis of different particular cases and does not affect the main point of this section.

Following the discussion in Section 2, the first step is to identify a set of physical symmetries in our system that can become potential gauge symmetries. The key observation is that the following transformations

$$h^{\mu\nu} \rightarrow h^{\mu\nu} + 2\nabla^{(\mu}\chi^{\nu)}, \tag{31}$$

with the field χ_μ obeying the following conditions,

$$-\xi_1 \Box \nabla_{(\mu} \chi_{\nu)} + (\xi_2 - \xi_1) \nabla_\mu \nabla_\nu (\nabla^\alpha \chi_\alpha)$$
$$+ (\xi_2 - \xi_3) \eta_{\mu\nu} \Box \nabla_\alpha \chi^\alpha - m_1^2 \nabla_{(\mu} \chi_{\nu)} - m_2^2 \eta_{\mu\nu} \nabla_\alpha \chi^\alpha = 0. \tag{32}$$

are physical symmetries of the action defined in Equation (20). This can be seen by computing their Noether currents. The previous expression can be written in a more illustrative form as

$$\nabla_{(\mu} V_{\nu)} + \xi_4 \eta_{\mu\nu} (\nabla_\alpha V^\alpha) + \left(\frac{\xi_4}{\xi_1} m_1^2 - m_2^2 \right) \eta_{\mu\nu} (\nabla_\alpha \chi^\alpha) = 0,$$
$$V_\nu = \Box \chi_\nu + \left(1 - \frac{\xi_2}{\xi_1} \right) \nabla_\nu (\nabla_\sigma \chi^\sigma) + \frac{m_1^2}{\xi_1} \chi_\nu, \tag{33}$$

with $\xi_4 = (\xi_3 - \xi_2)/(2\xi_1 - \xi_2)$. As we are interested in physical symmetries that can become emergent gauge symmetries, let us concentrate on the massless case. Then we have

$$\nabla_{(\mu} V_{\nu)} + \xi_4 \eta_{\mu\nu} (\nabla_\alpha V^\alpha) = 0,$$
$$V_\nu = \Box \chi_\nu + \left(1 - \frac{\xi_2}{\xi_1} \right) \nabla_\nu (\nabla_\sigma \chi^\sigma). \tag{34}$$

Taking the trace of the equation on the first line (contracting with $\eta^{\mu\nu}$), we obtain that $\nabla^\alpha V_\alpha = 0$ and thus $\nabla_{(\mu} V_{\nu)} = 0$. Imposing also that χ^μ goes to zero at infinity, we deduce that $V^\mu = 0$.

At this stage and as an aside, it is interesting to realize that the condition $V^\mu(\chi) = 0$ for a given vector field χ^μ can be extracted by requiring that the infinitesimal transformations from Equation (31) leave the subspace of $h^{\mu\nu}$ invariant, obeying the following (putative gauge) condition of the form

$$\nabla^\mu h_{\mu\nu} - \frac{\xi_2}{2\xi_1} \nabla_\nu h = 0. \tag{35}$$

Let us now pass to explicitly analyze the structure of the Noether charges. Applying Noether's theorem to the symmetries satisfying Equation (31) and suitably rearranging the terms, we get the following expression for the current with $V_\mu = 0$,

$$J^\mu = W^\mu + 2\nabla_\nu N^{[\mu\nu]} + M^\mu + D^\mu, \tag{36}$$

where the terms in the previous expressions read

$$W^\mu = -\Box h^{\mu\nu} \chi_\nu + 2T^{\mu\nu} \chi_\nu, \tag{37}$$

$$N^{\mu\nu} = \nabla^\nu h^{\mu\rho} \chi_\rho + h^{\nu\rho} \nabla^\mu \chi_\rho, \tag{38}$$

$$M^\mu = -\bigg[(1 + \xi_2) \nabla_\alpha h^{\mu\alpha} + (1 + \xi_3) \nabla^\mu h - (1 + \xi_1) h^{\mu\alpha} \nabla_\alpha$$
$$+ \frac{1}{2} (1 + \xi_2) h \nabla^\mu + (1 + \xi_2) h^{\mu\nu} \nabla_\nu - (1 + \xi_3) h \nabla^\mu \bigg] \nabla_\rho \chi^\rho, \tag{39}$$

$$D^\mu = -\nabla^\alpha h_{\alpha\beta} \nabla^\mu \chi^\beta + \nabla^\mu \nabla^\alpha h_{\alpha\beta} \chi^\beta + (1 + \xi_1) \nabla_\beta h^\beta_\nu \nabla^\mu \chi^\nu + (1 + \xi_1) \nabla_\beta h^\beta_\nu \nabla^\nu \chi^\mu$$
$$- \frac{1}{2} (1 + \xi_2) \nabla_\nu h \nabla^\mu \chi^\nu - \frac{1}{2} (1 + \xi_2) \nabla_\nu h \nabla^\nu \chi^\mu + (1 + \xi_2) h \Box \chi^\mu. \tag{40}$$

The charges associated with these currents J^μ do not identically vanish as long as we avoid the particular choices of parameters discussed above. We can also check that by restricting to a subspace of solutions satisfying the condition (35) one is still not able to make all Noether charges associated with the symmetry $V_\mu = 0$ to vanish with full generality.

However, we can focus our attention on a reduced set of symmetries, defined as transformations of the form (31) with the generators obeying

$$\Box\chi^\nu = 0,$$
$$\nabla_\nu(\nabla_\sigma\chi^\sigma) = 0. \tag{41}$$

Notice that these conditions automatically ensure that $V^\mu(\chi) = 0$, since they correspond to independently putting to zero the two terms entering the definition of V^μ in (33). It is not difficult to realize that the corresponding Noether charges all vanish when restricted to a subspace of solutions of the theory, which we call $\mathcal{U}$, characterized by the conditions

$$\nabla_\mu h^{\mu\nu} = 0, \qquad \nabla_\nu h = 0. \tag{42}$$

The conditions from Equation (41) for the transformations (31) introduced above precisely correspond to the subset of such transformations that preserve the conditions from Equation (42). We will come back to this point shortly. Furthermore, notice that the conditions defining this subspace are Lorentz-invariant themselves.

To see why when we restrict ourselves to the subspace of solutions $\mathcal{U}$ the Noether charges of the symmetries satisfying Equation (41) identically vanish, it is useful to notice that we have rearranged the conserved current (36) in a sum of four pieces, where:

1. The first term, W^μ, vanishes on shell within the subspace $\mathcal{U}$;
2. The second term, given by the divergence of $N^{\mu\nu}$, is a superpotential;
3. The third term, M^μ, vanishes identically if we take the transformations such that $\nabla_\alpha(\nabla_\mu\chi^\mu) = 0$;
4. The fourth term, D^μ, is such that it also identically vanishes within the subspace $\mathcal{U}$.

This discussion allows us to conclude that if we define a projection onto the subspace $\mathcal{U}$, defined by the constraints $\nabla_\mu h = 0$ and $\nabla_\mu h^{\mu\nu} = 0$, of the massless theory, these physical symmetries become emergent gauge symmetries within that subspace. We will discuss in a moment how this subspace could be selected dynamically.

In addition to the diffeomorphism-like transformations from Equation (31) already discussed, we can also ask whether the Weyl transformations introduced in Equation (30) can become emergent gauge symmetries using the same mechanism. The first step is identifying the following Weyl-like physical symmetries of the theory, given by the transformations $h_{\mu\nu} \to h_{\mu\nu} + \eta_{\mu\nu}\phi/2$ where the scalar field ϕ needs to obey the following equation:

$$\left[\left(-\frac{1}{2} + \frac{1}{4}\xi_2 - \xi_3\right)\Box - \frac{1}{4}m_1^2 - m_2^2\right]\eta_{\mu\nu}\phi + \left[\frac{1}{2} + \xi_2 - \frac{1}{2}\xi_1\right]\nabla_\mu\nabla_\nu\phi = 0. \tag{43}$$

Taking the trace of this equation and considering the case $m_1 = m_2 = 0$, we find the condition

$$\Box\phi = 0, \tag{44}$$

and plugging it back in (43) we find the condition

$$\nabla_\mu\nabla_\nu\phi = 0, \tag{45}$$

for the Weyl transformations to be symmetries of the theory. The solutions to this equations are of the following form

$$\phi(x) = A + B_\mu x^\mu \qquad A, B_\mu \in \mathbb{R}. \tag{46}$$

Now, we can compute the Noether currents associated with these symmetries which read

$$J^\mu = \frac{1}{4}(2 - \xi_2 + 4\xi_3)(h\nabla^\mu\phi - \phi\nabla^\mu h), \tag{47}$$

where we recall that we have set the masses to zero $m_1 = m_2 = 0$. To ensure that these transformations correspond to gauge symmetries, we need to ensure that they have trivial Noether charges. For that purpose, we have two options.

First case: $B_\mu \neq 0$. If we take $B_\mu \neq 0$ we need to restrict ourselves to the subspace $h = 0$ in order to have a trivial current. However, notice that these transformations do not preserve such subspace and hence we would be in the first situation described in Section 2, namely the non-emergence of gauge symmetry.

Second case: $B_\mu = 0$. In this second case we can relax the condition on h to be $\nabla_\mu h = 0$ in order to have a trivial current. Notice that these transformations preserve the subspace $\nabla_\mu h = 0$ and hence this second option corresponds to the second situation described in Section 2, i.e., the emergence of gauge symmetry. More explicitly, to have WTDiff emergent transformations, we need to restrict ourselves to the same subspace $\mathcal{U}$ introduced above. We recall that such space is defined by the conditions (42) and that we need to choose it because the emergent Weyl gauge symmetries preserve them and the Noether currents restricted to such subspace give identically vanishing charges. Hence, we conclude that it is possible to also find emergent Weyl symmetries with the mechanism introduced in Section 2.

It is interesting to realize that the space $\mathcal{U}$ in which we find the emergence of gauge symmetries is the same independently of whether we are trying to find the closest theory to Fierz–Pauli theory or to linearized WTDiff within this emergent paradigm. In fact, a (non-transverse) diffeomorphism generated by a vector field with constant divergence $\nabla_\alpha \chi^\alpha = \kappa \in \mathbb{R}$, such as the ones we have considered above, is equivalent to the composition of a transverse diffeomorphisms generated by the transverse part of such vector field χ^μ and a Weyl transformation generated by a constant scalar field with $\phi = 4\kappa$. Both of them can be seen to produce constant shifts in h and hence preserve the subspace $\mathcal{U}$. In other words, when we understand the conditions defining subspace $\mathcal{U}$, i.e., $\nabla_\mu h^{\mu\nu} = \nabla^\nu h = 0$, as gauge fixing conditions for either Fierz–Pauli or linearized WTDiff, the residual gauge symmetries that leave such subspace invariant for both theories are exactly the same. Thus, whether one decides to understand the emergent gauge theory we have built as having a Fierz–Pauli flavor or a linearized WTDiff flavor is inconsequential for the discussion at the linear level.

Now, let us pass to discuss under which conditions the space $\mathcal{U}$ is selected naturally by the dynamics of the theory. If we assume that the tensor $T^{\mu\nu}$ is conserved, taking the divergence of the equations of motion uncovers a structural constraint:

$$-\frac{1}{2}\xi_1 \Box \nabla_\mu h^{\mu\nu} + \frac{1}{2}(\xi_2 - \xi_1)\nabla^\nu \nabla^2 \cdot h + \frac{1}{2}(\xi_2 - \xi_3)\nabla^\nu \Box h = \nabla_\mu T^{\mu\nu} = 0. \tag{48}$$

Taking an additional divergence we find

$$\Box \left[(\xi_2 - 2\xi_1)(\nabla^2 \cdot h) + (\xi_2 - \xi_3)\Box h \right] = 0. \tag{49}$$

Given this structural equation, which is a wave equation without a source, we can argue as we have already done before that the following solutions are selected dynamically,

$$(\xi_2 - 2\xi_1)(\nabla^2 \cdot h) + (\xi_2 - \xi_3)\Box h = 0. \tag{50}$$

Plugging this condition into Equation (48) we obtain

$$\Box \left[\nabla_\mu h^{\mu\nu} + \frac{(\xi_2 - \xi_3)}{(\xi_2 - 2\xi_1)}\nabla^\nu h \right] = 0, \tag{51}$$

and applying again the argument that the lack of sources selects naturally the trivial solution, we conclude that

$$\nabla_\mu h^{\mu\nu} + \frac{(\xi_2 - \xi_3)}{(\xi_2 - 2\xi_1)} \nabla^\nu h = 0. \tag{52}$$

Notice that this condition does not imply that the two terms in the previous expression are zero independently. For this to happen one needs to impose further constraints.

One possibility is to further require that the trace of $T^{\mu\nu}$ is a constant over spacetime. Taking the trace of the equations of motion for the field $h^{\mu\nu}$ and introducing the following notation $T = \eta^{\mu\nu} T_{\mu\nu}$, we find

$$\frac{1}{2}\left[(-2 + \xi_2 - 4\xi_3)\Box h + (2 - 2\xi_1 + 4\xi_2)\nabla^2 \cdot h\right] = T. \tag{53}$$

Taking a derivative, using the assumption of constant trace for the current, i.e., $\nabla_\mu T = 0$, and using also (50) we obtain the following condition (assuming that the parameters do not take specific values, which we discuss in Appendix A):

$$\Box(\nabla_\nu h) = 0. \tag{54}$$

Again, for a generic system, and applying the argument that the lack of sources leads naturally to the trivial solution, we obtain the condition $\nabla_\nu h = 0$. Inserting this condition in Equation (52) we precisely see that the subspace $\mathcal{U}$ introduced in Equation (42) is naturally recovered. Thus, under the assumption of coupling to a conserved current $\nabla_\mu T^{\mu\nu} = 0$ with constant trace, $\nabla_\mu T = 0$, we have found that the system develops emergent gauge symmetries that can either be understood as emergent diffeomorphisms or emergent Weyl-transverse diffeomorphisms.

Let us summarize the discussion concerning the emergence of gauge symmetries for the two-index symmetric tensor field $h^{\mu\nu}$ up to this point. We have managed to prove that for the most general Lorentz-invariant theory constructed from such tensor field, the mechanism presented in Section 2 works and provides us with emergent gauge symmetries. At the linear level, the gauge transformations that emerge can be either understood as the manifestation within this emergent framework of either linearized diffeomorphisms or linearized Weyl and transverse diffeomorphisms transformations. To put it explicitly, consider any generic Lorentz-invariant theory of $h^{\mu\nu}$ which fulfills the following conditions:

1. It describes massless excitations, namely the masses from the Lagrangian in Equation (20) are equal to zero ($m_1 = m_2 = 0$).
2. It couples to a two-index symmetric source $T^{\mu\nu}$ that is conserved, i.e., divergenceless $\nabla_\mu T^{\mu\nu} = 0$.
3. The source has constant trace, namely $\nabla_\mu T = 0$.

Our analysis implies that at the linear level, such a theory has a natural truncation that is indistinguishable from either Fierz–Pauli theory or linearized WTDiff theory in the gauge described by (42). Notice that, instead of a projection to a given dynamical subspace of the theory, for these theories we understand the conditions $\nabla_\mu h^{\mu\nu} = \nabla^\nu h = 0$ as gauge fixing conditions. Actually, we emphasize that Fierz–Pauli theory and linearized WTDiff are indistinguishable in this specific gauge, in the sense that the residual gauge transformations that emerge are of the same form. This is the reason for understanding this emergent gauge theory as the manifestation of either WTDiff and Fierz–Pauli theory in the emergent framework: both theories are indistinguishable in this specific gauge.

The extension to the non-linear regime of the gravitational case through a bootstrapping procedure is expected to be much more involved than for the vector-field case discussed in Section 3.1 and it is left for future work. It constitutes an ongoing project on which we expect to report soon. Here we simply advance that the source of this difficulty arises, apart from the obvious technical complications appearing due to the presence of

much more terms in the Lagrangian, from the nature of the non-linear terms appearing in the series generated through the bootstrapping procedure. Whereas for vector fields there appears a finite series that finishes after two iterations, for gravity we need to deal with an infinite "formal" series to obtain the non-linear theory [22,26].

4. No-Go Theorems for Emergent Gravity Reconsidered

The Weinberg–Witten theorem [27] is often used to make a case against emergent gravity, attempting to claim that any research program pursuing to describe gravitational dynamics as an emergent phenomenon is condemned to fail. Marolf's recent arguments seem to present also an important obstruction towards the emergent gravity paradigm [28], at least from the perspective of constraining the required characteristics of emergent approaches. In this section, we will review the Weinberg–Witten theorem and also revisit Marolf's argument and point out why the possibility that gauge symmetries may be emergent cannot be ignored when attempting to use these theorems in this way. Although our results for gravity are still for the linear theory, there is a clear interplay between our results and these obstructions as we will discuss. Let us note that these theorems assume a quantum character for the gravitational field. Our results for linear fields, although derived for classical fields, can be straightforwardly extended to the quantum level through a quantization procedure, see Section 3.4 of Ref. [14]. The issue of the non-linear theory is much more subtle, since we do not have a non-perturbative definition of quantum field theories that does not involve a lattice regularization. Hence, a careful analysis of that point would be required, although our main conclusions below do not depend crucially on this issue.

4.1. Weinberg–Witten Theorem

The Weinberg–Witten theorem [27] states that a theory that allows the construction of a conserved Lorentz-covariant energy-momentum tensor cannot contain gravitons, neither fundamental nor composite in the spectrum of asymptotic states (in fact, the theorem applies to particles with spin $j > 1$). The content of the theorem can be rephrased as saying that gravitons are so special that one cannot build a Lorentz-covariant and gauge-invariant energy-momentum tensor for them. This distinctive feature already appears in the Fierz–Pauli theory for linear gravitons. One can certainly build different Lorentz-invariant energy-momentum tensors for $h_{\mu\nu}$; however, none of them happens to be invariant under the (linear) gauge transformation generated by a vector field χ^μ that take the form $\delta h^{\mu\nu} = 2\nabla^{(\mu}\chi^{\nu)}$, with ∇ representing the connection compatible with flat spacetime metric, as before.

From the perspective of emergent gravity the theorem implies that, if excitations with the properties of gravitons arise in a suitable regime, these excitations cannot have a Lorentz-invariant energy-momentum tensor that is gauge invariant. This is often taken as an indication that obtaining gravitons from systems akin to condensed-matter systems is not possible. However, our discussion of emergent gauge symmetries above indicates that such an interpretation is pulling too far, or misunderstanding, the significance of the theorem. Let us discuss this more explicitly.

The key aspect is the emergent nature of gauge symmetries and what this implies for observables, in particular for the energy-momentum tensor. Our starting point is a system akin to a condensed-matter system which develops excitations with the properties of gravitons at low energies. We have seen that this requires certain conditions on the source to which the field $h^{\mu\nu}$ couples which, when satisfied, are equivalent to the decoupling of some degrees of freedom leading to the natural restriction to the subspace $\mathcal{U}$ described in previous sections. As the original high-energy theory has no gauge symmetries, all configurations are different and, therefore, in general will have different values for most observables (in this section, this word will always refer to the high-energy theory), which includes the fields themselves and the energy-momentum tensor as specific examples. However, the development of emergent gauge symmetries implies the emergence of

equivalence classes for which the values of certain observables do not change, which are then denoted as gauge invariant. The specific observables that change or not within a given equivalence class are determined by the structure of the theory. The basic fields A_a^μ and $h^{\mu\nu}$ are examples of observables of the high-energy theory that change under the emergent gauge symmetries. On the other hand, for vector fields A_a^μ the energy-momentum tensor is gauge-invariant [14,15], while for gravity it is not.

In summary, that certain observables of the high-energy theory turn out to be non gauge invariant does not pose a problem for emergence. From the perspective of the high-energy theory, all observables retain their meaning and have well-defined values regardless of the emergence of gauge symmetries. It is certainly not required and even unreasonable to assume that all observables of the high-energy theory become gauge invariant. In general, the process of emergence separates observables in two categories, depending on their behavior under gauge transformations. However, whether the set of gauge-invariant observables includes the energy-momentum tensor, or any other observable for that matter, is inconsequential for the rationale of emergence.

We have discussed the Weinberg–Witten theorem from a purely conceptual point of view and highlighted that, although it points out a particular property of gravity, it does not forbid the emergence of gravity. From a technical point of view, it applies to gravity when understood as a perturbative quantum field theory built on top of flat spacetime having a well-defined notion of gravitons as asymptotic states. Let us also mention that, within this framework, several technical aspects of the theorem have been criticized throughout the years [29–33]. These criticisms further limit the application of the theorem to realistic physical situations. Furthermore, a comprehensive and systematic discussion of the ways in which the Weinberg–Witten theorem can be avoided can be found in [34].

4.2. Marolf's Theorem

The very interesting work of Marolf [28] contains a theorem that seems to constrain the kind of theories that might give rise to an emergent gravitational description. According to Marolf [28], the only theories that might give rise to such a description are the ones that he calls *kinematically non-local* (we will introduce their definition of kinematical non-locality in a moment). From our perspective, although Marolf's theorem pinpoints the particularities of constructing an emergent theory of gravity from a *kinematically local* theory, it does not automatically imply that such theories cannot have an effective gravitational description in some regime. Let us now introduce Marolf's theorem and carefully analyze it as well as its implications. Quoting it literally, Marolf's theorem states [28]:

Marolf's theorem: *"Consider any limit where the effective description of a local theory is a gravitational theory with universal coupling to energy, the same notion of time evolution, and a compatible definition of locality. In this limit all local observables away from the boundary become independent of time."*

The theorem above contains several definitions that need to be clarified to make it intelligible, which we review here for the benefit of the reader. On the one hand, Marolf restricts their study to what he calls gravitational theories with *universal coupling to the energy*, namely theories for which the Hamiltonian can be written as the integral over the boundary of space at each time of some local function, which he calls gravitational flux. Furthermore, such gravitational flux needs to be a gauge-invariant function of the gravitational field and its derivatives. All diffeomorphism-invariant theories fall into this category (non-Lorentz-invariant theories like for instance Hořava-Lifshitz gravity [35] do not in a strict sense, although Marolf suggests the existence of a straightforward generalization of this theorem that would apply to such theories). Massive gravity theories, like dRGT theory [36], and scalar gravities, like the so-called Nordström gravity [37], do not fall into this category since they do not have universal coupling to the energy.

On the other hand, by locality he assumes a version of microcausality which he calls *kinematic locality*. According to their definition, a kinematically local theory is one for which every pair of local bosonic gauge-invariant Heisenberg operators evaluated at spacelike points commute.

Then, he states that if there were a kinematically local (non-gravitational) theory that in some regime gave place to an effective description in terms of a gravitational theory with universal coupling to energy, all of its local bulk observables would freeze out unless there were a change on the notion of time evolution or locality between the gravitational and non-gravitational descriptions (these are notes of caution). Technically speaking, Marolf's observation is that given the gravitational flux which is a gauge-invariant operator, it should be possible to write it in the microscopic underlying theory as an integral over gauge-invariant operators supported near the boundary. Otherwise, it would correspond to a theory in which the notions of locality between the microscopic and effective gravitational theory would be very different. Assuming this *compatible-locality property* among both descriptions, Marolf reaches the conclusion that, given a kinematically local underlying theory and having expressed the Hamiltonian as an operator in that theory just supported near the boundary, all the local (or bulk) observables of the theory need to freeze out in the regime in which the effective gravitational description applies. This is because the Hamiltonian (as it is itself another observable) will commute with any local observable away from the boundary.

Given the theorem, he then concludes that it is difficult to obtain an emergent gravitational description starting from microscopic kinematically local theories, without changing abruptly the notion of locality. Otherwise, the emergence would entail a rather strange behaviour on the non-gravitational description. Thus, he interprets the theorem as saying that one would rather start looking for emergence from microscopic theories which are already kinematically non-local at a fundamental level, as it is done for instance in the context of gauge-gravity duality [38,39] which, to some extent, can be interpreted as emergent theories for the bulk gravity theory. However, this conclusion is based on the additional assumption that all observables in the UV completion of the theory must freeze at low energies or, equivalently, that all these observables must be gauge invariant, which may not be the case in an emergent framework. If at least one observable of the UV completion is not gauge invariant, then the notion of locality can be preserved. This would be generally the case if gauge symmetries are emergent.

If we take our discussion of the linear setting above and assume the existence of a suitable non-linear extension along the lines realized in the discussed Yang–Mills case, we can make the following observation. When an effective gauge invariance is developed in the low-energy regime of the theory, the observables of the high-energy theory can be separated into those that are gauge invariant from the perspective of the emergent gauge symmetry and those that are not. As an intermediate stage of emergence, we started with a generic massless Lorentz-covariant system described by action (20) for an object $h^{\mu\nu}$. Without further considerations, this system does not describe gravity, as for instance it contains additional degrees of freedom, and, therefore, does not satisfy the conditions of Marolf's theorem. However, we have discussed that, if the field $h^{\mu\nu}$ couples to a conserved and constant-trace two-index symmetric current, there is a natural decoupling of certain degrees of freedom and a simultaneous emergence of gauge symmetries. Interestingly, this decoupling makes the construction compatible with Marolf's conditions for the theorem. Then, it is true that all the local gauge-invariant observables away from the boundary need to freeze out. However, non-gauge-invariant observables, which are still meaningful from the point of view of the high-energy UV-completion of the theory, do not need to freeze out. In fact, our example shows that actually they (i.e., the $h_{\mu\nu}$) do not freeze out. As we have pointed out above, these observables are not constrained to be frozen by the theorem. The freeze out only affects a subset of the observables of the high-energy theory, namely those observables that are invariant under the emergent gauge transformations. In this way, the high-energy theory retains an "interesting enough" bulk dynamics.

Another aspect that we think is worth keeping in mind is the effect of having a background metric. As our underlying theory has a (flat) background metric the degrees of freedom are defined with respect to this background and are perfectly local, as for electrodynamics and other quantum field theories. In the same vein, our theory has a well-defined local Hamiltonian density and does not have a fundamental gauge invariance. However, one could decide—and this is a choice—to analyze the theory from a gauge perspective, arguing that the background metric is not operationally measurable. Then, for consistency one has to change the notion of degree of freedom as now one cannot use the background metric as an external reference. The degrees of freedom must now be gauge-invariant variables defined with respect to the metric, variables which in general relativity are highly non-local. Now, the former Hamiltonian density is not gauge invariant and therefore one can no longer think of it as locally defining an energy density. Instead, from the gauge perspective the Hamiltonian appears as a gauge-invariant and non-local (holographic) quantity. The curvatures in the bulk of spacetime certainly evolve, but even those do not constitute gauge-invariant degrees of freedom as we do not have an independent notion of what is a point (or event). If the notion of what is a degree of freedom did not change from the underlying theory to the gravitational emergent theory, then Marolf's theorem would be true in asserting that any local (gauge-invariant!) observable would need to freeze in the effective gravitational regime. However, the change in the notion of degree of freedom permits that local degrees of freedom in the bulk of the underlying theory have a well-defined local kinematics.

Therefore, a decoupling from the high-energy perspective has as consequence a freeze out of certain non-local observables. While we have not provided a complete non-linear model satisfying the criteria that guarantee the decoupling, the latter are general enough to make us think that such models should exist (assuming that the extension of our theory into the non-linear regime is completely parallel to the one we have already done with Yang–Mills theory [15]). At this stage we can safely conclude that searching for models satisfying these criteria is a well-motivated endeavour.

5. Discussion and Conclusions

In this work, we have reviewed a mechanism proposed by some of the authors for the emergence of gauge symmetries in field theories based on the natural decoupling of some of their degrees of freedom. We have also reviewed how this mechanism for the emergence of linear gauge theories combined with a suitable bootstrapping procedure linking the linear and non-linear regimes can give rise to emergent gauge symmetries resembling Yang–Mills theories. This example serves as a warm-up exercise for the more convoluted case of finding an emergent gravitational theory.

We have analyzed in detail the case of linear gravitons and found that it is possible to give rise to the gauge symmetries characteristic of linear gravity through this mechanism. Although we have left the analysis of the non-linear extensions through a bootstrapping procedure for future work, we expect no substantial differences from the standard bootstrapping for gravity. The main difficulties actually come from the problems that are present already for the Fierz–Pauli theory: namely the infinite nature of the series, the inherent ambiguities present in the construction, etc.

Additionally, we have discussed some results in the literature that are often invoked as no-go theorems and discussed their interplay with our work. These results are indeed technically correct and pinpoint some differential features of gravity and gravitons with respect to other theories and gauge particles. However, using these results as no-go theorems implies making additional assumptions about the possible UV-completions, in particular regarding the preservation of gauge symmetries at high energies. As we have discussed, these theorems cannot rule out emergent frameworks in which gauge symmetries are emergent themselves.

We finish this work by mentioning that, although many technical details remain to be understood and analyzed within this framework, it seems to provide a successful approach

towards building an emergent gravitational description of a certain theory. Given that a theory acquires an effective (typically we think of it as a low-energy regime) Lorentz-invariant description for which a field $h^{\mu\nu}$ describes adequately the excitations of the theory, we have found that the decoupling of some of its degrees of freedom (which is quite natural from a dynamical point of view) automatically gives rise to an emergent gauge symmetry. The remaining extension into the non-linear regime through a bootstrapping procedure would allow us to constrain the possible non-linear completions of such linear theory. As a result, we would be able to build emergent gravitational description of systems, by simply developing a low-energy Lorentz-invariant regime and having a $h^{\mu\nu}$ field in their low-energy spectrum. We expect to report soon on these issues.

Another future line of research that is worth exploring is understanding the emergence of gauge symmetries in alternative descriptions of gravity. For instance, the reformulation of gravity as a more standard gauge theory of the local Poincaré group [40] seems to be a promising place to test the generality of the applicability of our mechanism.

Author Contributions: All the authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Special Cases of Parameters for Linear Graviton Fields

We have omitted the discussion of special cases of the parameters (ξ_1, ξ_2, ξ_3) for which some of the terms in the equations from Section 3.2 vanish. The unique special case that gives new insights is the case in which the three of them are equal: $\xi = \xi_1 = \xi_2 = \xi_3$. For the remaining cases, for example the case for which some of the terms in Equation (53) vanish, namely $-2 + \xi_2 - 4\xi_3 = 0$ or $1 - \xi_1 + 2\xi_2 = 0$, one can see that the same arguments presented Section 3.2 for the selection of the subspace $\mathcal{U}$ straightforwardly apply and hence we do not need to discuss them separately.

The case $\xi = \xi_1 = \xi_2 = \xi_3$ where the masses have been put to zero is interesting because the condition (32) that the generators χ^μ of the transformations (31) need to verify reduces to

$$\nabla_\mu \Box \chi_\nu = 0. \tag{A1}$$

Thus, it seems that for this specific choice of parameters the transformations of the form

$$h^{\mu\nu} \to h^{\mu\nu} + 2\nabla^{(\mu}\chi^{\nu)}, \qquad \Box\chi^\mu = 0, \tag{A2}$$

have a chance to become emergent gauge symmetries. Actually, for this choice of parameters the terms of the associated current can be recast in the form (36) without the M'^μ. This means that we have a current of the form

$$J'^\mu = W'^\mu + 2\nabla_\mu N'^{\mu\nu} + D'^\mu, \tag{A3}$$

with W'^{μ} a term that vanishes on shell, $N'^{\nu\mu}$ a superpotential and D'^{μ} a term that identically vanishes within the subspace $\nabla_\mu h^{\mu\nu} = 0$. Their explicit form is:

$$W'^{\mu} = -\Box h^{\mu\alpha}\chi_\alpha - (1+\xi)\nabla_\nu\nabla^\mu h\chi^\nu + (1+\xi)\Box h\chi^\mu + 2T^{\mu\nu}\chi_\nu, \tag{A4}$$

$$N'^{\mu\nu} = \nabla^\nu h^{\mu\rho}\chi_\rho + h^{\nu\rho}\nabla^\mu\chi_\rho + \frac{1}{2}(1+\xi)h\nabla^\nu\chi^\mu + (1+\xi)\nabla^\mu h\chi^\nu, \tag{A5}$$

$$D'^{\mu} = -\nabla^\alpha h_{\alpha\beta}\nabla^\mu\chi^\beta + \nabla^\mu\nabla^\alpha h_{\alpha\beta}\chi^\beta + (1+\xi)\nabla_\beta h^\beta_\nu\nabla^\mu\chi^\nu + (1+\xi)\nabla_\beta h^\beta_\nu\nabla^\nu\chi^\mu. \tag{A6}$$

Thus, if we define a projection onto the subspace $\tilde{\mathcal{U}}$ defined by the condition

$$\nabla_\mu h^{\mu\nu} = 0, \tag{A7}$$

the transformations of the form (A2) become emergent gauge symmetries.

The naturalness of such construction relies again on how natural it is to restrict the dynamics to the subspace $\tilde{\mathcal{U}}$. We want to see if that subspace for this particular choice of parameters is selected dynamically. For this particular choice of parameters, the divergence of the equations of motion assuming we couple the theory to a conserved source automatically gives

$$\Box\nabla_\mu h^{\mu\nu} = 0, \tag{A8}$$

as it can be seen by putting all the ξ_i parameters equal to the same value ξ in Equation (48). Thus, for this choice of parameters, the transformations of the form (A2) naturally emerge as gauge symmetries since the subspace onto which we need to project for them to become emergent gauge symmetries is naturally selected from a dynamical point of view.

Although this construction seems to be closer to Fierz–Pauli theory since it is equivalent to Fierz–Pauli theory in the partial gauge fixing given by the only condition $\nabla_\mu h^{\mu\nu} = 0$, it does not seem to be natural due to the fine-tuning among the parameters that it requires. For instance, we expect that, at the quantum level, radiative corrections may tend to rapidly break the equality among parameters. However, it is possible that the emergent gauge symmetries themselves are enough to protect this equality among the parameters and render them radiatively stable. We leave this question for future work, when we extend this construction to the whole non-linear and quantum regime, in which these radiative corrections enter and may play a relevant role.

References

1. Carlip, S. Quantum gravity: A Progress report. *Rep. Prog. Phys.* **2001**, *64*, 885. [CrossRef]
2. Polchinski, J. *String Theory. Volume 1: An Introduction to the Bosonic String*; Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK, 2007. [CrossRef]
3. Polchinski, J. *String Theory. Volume 2: Superstring Theory and Beyond*; Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK 2007. [CrossRef]
4. Volovik, G.E. Emergent physics: Fermi point scenario. *Philos. Trans. R. Soc. Lond. A* **2008**, *366*, 2935–2951. [CrossRef]
5. Volovik, G. *The Universe in a Helium Droplet*; International Series of Monographs on Physics; Oxford University Press: Oxford, UK, 2009.
6. Barceló, C.; Liberati, S.; Visser, M. Analogue gravity. *Living Rev. Relativ.* **2005**, *8*, 12. [CrossRef] [PubMed]
7. Barceló, C.; Carballo-Rubio, R.; Garay, L.J.; Jannes, G. Electromagnetism as an emergent phenomenon: A step-by-step guide. *New J. Phys.* **2014**, *16*, 123028. [CrossRef]
8. Thiemann, T. *Modern Canonical Quantum General Relativity*; Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK, 2007. [CrossRef]
9. Dirac, P. *Lectures on Quantum Mechanics*; Belfer Graduate School of Science, Monograph Series; Dover Publications: Mineola, NY, USA, 2001.
10. Carlip, S. Challenges for Emergent Gravity. *Stud. Hist. Phil. Sci. B* **2014**, *46*, 200–208. [CrossRef]
11. Collins, J.; Perez, A.; Sudarsky, D.; Urrutia, L.; Vucetich, H. Lorentz invariance and quantum gravity: An additional fine-tuning problem? *Phys. Rev. Lett.* **2004**, *93*, 191301. [CrossRef]
12. Polchinski, J. Effective field theory and the Fermi surface. Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles. 1992. Available online: http://xxx.lanl.gov/abs/hep-th/9210046 (accessed on 17 Septermber 2021).
13. Manohar, A.V. Introduction to Effective Field Theories. *arXiv* **2018**, arXiv:1804.05863.

14. Barceló, C.; Carballo-Rubio, R.; Di Filippo, F.; Garay, L.J. From physical symmetries to emergent gauge symmetries. *J. High Energy Phys.* **2016**, *10*, 084. [CrossRef]
15. Barceló, C.; Carballo-Rubio, R.; Garay, L.J.; García-Moreno, G. Emergent gauge symmetries: Yang–Mills theory. *arXiv* **2021**, arXiv:2101.12188.
16. Henneaux, M.; Teitelboim, C. *Quantization of Gauge Systems*; Princeton Paperbacks, Princeton University Press: Princeton, NJ, USA, 1994.
17. Julia, B.; Silva, S. Currents and superpotentials in classical gauge invariant theories. 1. Local results with applications to perfect fluids and general relativity. *Class. Quant. Grav.* **1998**, *15*, 2173–2215. [CrossRef]
18. Aldaya, V.; Guerrero, J.; López-Ruiz, F.F.; Cossío, F. $SU(2)$ particle sigma model: the role of contact symmetries in global quantization. *J. Phys. A* **2016**, *49*, 505201. [CrossRef]
19. Deser, S. Self-interaction and gauge invariance. *Gen. Relativ. Gravit.* **1970**, *1*, 9–18. [CrossRef]
20. Ortin, T. *Gravity and Strings*, 2nd ed.; Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK, 2015. [CrossRef]
21. Padmanabhan, T. From gravitons to gravity: Myths and reality. *Int. J. Mod. Phys. D* **2008**, *17*, 367–398. [CrossRef]
22. Barceló, C.; Carballo-Rubio, R.; Garay, L.J. Unimodular gravity and general relativity from graviton self-interactions. *Phys. Rev. D* **2014**, *89*, 124019. [CrossRef]
23. Butcher, L.M.; Hobson, M.; Lasenby, A. Bootstrapping gravity: A Consistent approach to energy-momentum self-coupling. *Phys. Rev. D* **2009**, *80*, 084014. [CrossRef]
24. Alvarez, E.; Blas, D.; Garriga, J.; Verdaguer, E. Transverse Fierz–Pauli symmetry. *Nucl. Phys. B* **2006**, *756*, 148–170. [CrossRef]
25. Fierz, M.; Pauli, W. On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. *Proc. R. Soc. Lond. A* **1939**, *173*, 211–232. [CrossRef]
26. Barceló, C.; Carballo-Rubio, R.; Garay, L.J. Absence of cosmological constant problem in special relativistic field theory of gravity. *Ann. Phys.* **2018**, *398*, 9–23. [CrossRef]
27. Weinberg, S.; Witten, E. Limits on Massless Particles. *Phys. Lett. B* **1980**, *96*, 59–62. [CrossRef]
28. Marolf, D. Emergent Gravity Requires Kinematic Nonlocality. *Phys. Rev. Lett.* **2015**, *114*, 031104. [CrossRef]
29. Sudarshan, E.C.G. Massless Particles of High Spin. *Phys. Rev. D* **1981**, *24*, 1591–1594. [CrossRef]
30. Kugo, T. Limitations on the Existence of Massless Composite States. *Phys. Lett. B* **1982**, *109*, 205–208. [CrossRef]
31. Flato, M.; Sternheimer, D.; Fronsdal, C. Difficulties With Massless Particles. *Commun. Math. Phys.* **1983**, *90*, 563. [CrossRef]
32. Lopuszanski, J.T. On Charges of Massless Particles. *J. Math. Phys.* **1984**, *25*, 3503. [CrossRef]
33. Lopuszanski, J. Extended Weinberg–Witten Theorem. *J. Math. Phys.* **1988**, *29*, 1253–1257. [CrossRef]
34. Bekaert, X.; Boulanger, N.; Sundell, P. How higher-spin gravity surpasses the spin two barrier: No-go theorems versus yes-go examples. *Rev. Mod. Phys.* **2012**, *84*, 987–1009. [CrossRef]
35. Horava, P. Quantum Gravity at a Lifshitz Point. *Phys. Rev. D* **2009**, *79*, 084008. [CrossRef]
36. de Rham, C.; Gabadadze, G.; Tolley, A.J. Resummation of Massive Gravity. *Phys. Rev. Lett.* **2011**, *106*, 231101. [CrossRef]
37. Deruelle, N. Nordstrom's scalar theory of gravity and the equivalence principle. *Gen. Relativ. Gravit.* **2011**, *43*, 3337–3354. [CrossRef]
38. Banks, T.; Fischler, W.; Shenker, S.H.; Susskind, L. M theory as a matrix model: A Conjecture. *Phys. Rev. D* **1997**, *55*, 5112–5128. [CrossRef]
39. Maldacena, J.M. The Large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.* **1998**, *2*, 231–252. [CrossRef]
40. MacDowell, S.W.; Mansouri, F. Unified Geometric Theory of Gravity and Supergravity. *Phys. Rev. Lett.* **1977**, *38*, 739; Erratum in **1977**, *38*, 1376. [CrossRef]