

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/321374736>

Hints towards the Emergent Nature of Gravity

Article · November 2017

DOI: 10.1016/j.shpsb.2018.04.001

CITATIONS

16

READS

60

2 authors, including:



[Manus Visser](#)

University of Amsterdam

15 PUBLICATIONS 123 CITATIONS

[SEE PROFILE](#)

Hints towards the Emergent Nature of Gravity

Niels S. Linnemann^{*1} and Manus R. Visser^{†2}

¹ *Geneva Centre for Philosophy of Science, University of Geneva,
2 rue de Candolle, CH-1211 Geneva 4, Switzerland*

² *Institute for Theoretical Physics, University of Amsterdam,
Science Park 904, 1098 XH Amsterdam, The Netherlands*

November 30, 2017

Abstract

A possible way out of the conundrum of quantum gravity is the proposal that general relativity (GR) is not a fundamental theory but emerges from an underlying microscopic description. Despite recent interest in the emergent gravity program within the physics as well as the philosophy community, an assessment of the theoretical evidence for this idea is lacking at the moment. We intend to fill this gap in the literature by discussing the main arguments in favour of the hypothesis that the metric field and its dynamics are emergent. First, we distinguish between microstructure inspired from GR, such as through quantization or discretization, and microstructure that is not directly motivated from GR, such as strings, quantum bits or condensed matter fields. The emergent gravity approach can then be defined as the view that the metric field and its dynamics are derivable from the latter type of microstructure. Subsequently, we assess in how far the following properties of (semi-classical) GR are suggestive of underlying microstructure: (1) the metric's universal coupling to matter fields, (2) perturbative non-renormalizability, (3) black hole thermodynamics, and (4) the holographic principle. In the conclusion we formalize the general structure of the plausibility arguments put forward.

^{*}niels.linnemann@unige.ch

[†]m.r.visser@uva.nl

Contents

1	Introduction	3
2	What is emergent gravity?	4
2.1	Definition of emergent gravity	4
2.2	Relation to unification, emergent spacetime and EFT	10
2.3	Taxonomy of accounts: top-down vs. bottom-up	12
3	Hints towards an emergent nature	13
3.1	Universality of gravity	14
3.2	Perturbative non-renormalizability	18
3.3	Black hole thermodynamics	20
3.4	Holographic principle	24
4	Conclusion: From hints to arguments	27

1 Introduction

Gravity is notoriously harder to quantize than other interactions. It is well-known that combining general relativity (GR) with quantum mechanics poses several difficulties which are standardly traced back to the perturbative non-renormalizability of general relativity. Proponents of quantum gravity (QG) approaches proper (such as loop quantum gravity, asymptotic safety and causal dynamical triangulation) nevertheless expect that gravity is after all non-perturbatively renormalizable, and thereby possibly fundamental. In contrast, according to other approaches – like string theory and emergent gravity accounts by Sakharov (1968), Jacobson (1995), Padmanabhan (2008), Verlinde (2011), and Hu (2009) – the gravitational field cannot be fundamental but rather emerges from underlying microstructure. These approaches can be subsumed under the term ‘emergent gravity’ (EG).

The idea that gravity and spacetime originate from some underlying microscopic reality in which they do not exist is a widely held belief in, for example, the string theory community.¹ The possibility that gravitational degrees are rooted in non-gravitational degrees of freedom is often supported by reference to the AdS/CFT correspondence. Under the controversial condition that one side of the duality relation is more fundamental than the other, it indeed provides an important gateway to a theory of emergent gravity. An interesting current theme in AdS/CFT research is to connect gravitational physics to key concepts of quantum information theory. It is the philosopher of physics’ task to reflect on such current trends in quantum gravity research. In our view a particularly pressing question in this field is *why* it is widely believed that gravity is emergent – and *what* is meant by the term ‘emergent gravity’ in the first place.

As a matter of fact, it is surprising that the current philosophical literature² on EG mainly focuses on notions of emergence or possible implications for QG *rather than appraising the argument for the emergent nature of GR in the first place!* We intend to fill this gap in the literature by reconstructing and assessing arguments for the high-levelness of gravity, which are often only implicitly contained in the relevant physics literature. These arguments consist of pointing out features of gravity which are (at least typically) characteristic of systems with underlying microstructure. In other words, certain features – ranging from universal coupling to black hole thermodynamics – allow for mounting arguments from analogy to the effect that the metric field and its dynamics are emergent. Arguably, these gravitational features even remain ‘mysteries’ in a fundamental take on

¹Cf. Witten (1996), Seiberg (2006), Verlinde (2011).

²Cf. Bain (2013a), Mattingly (2014), Knox (2013), Crowther (2013, 2016), Dieks et al. (2015), de Haro (2016) and Carlip (2014). Admittedly, Carlip’s article is an appraisal of the emergent gravity program, but it only focuses on the challenges for such a program rather than the positive hints for the emergent nature of gravity.

the metric and appear to be in need of explanation, which however can be easily achieved in an emergent gravity framework.

The matter is worth exploring in particular in light of the interesting consequences a high-level view of GR could have for the conception of spacetime: it serves promoting a physical view of the metric as one field among others, as opposed to a conventional (geometric) spacetime view as advocated by Friedman (2014) and Maudlin (2012). Hence emergent gravity is very much in line with the particle physics view on GR (the ‘spin-2 approach’) advocated by Feynman et al. (2003) and Weinberg (1979), and with the dynamical approach to relativity by Brown (2005). In any case, if one agrees with the call to philosophers by Huggett and Wüthrich (2013) to engage with conceptual problems in the development of theories of QG, an assessment of the indications for the emergent nature of gravity will count as a sensible, if not necessary, contribution.

This paper is organised as follows. We start out in section 2 by clarifying our notions of emergence, microscopic structure and high-levelness. Ultimately, we provide a definition of emergent gravity which we take to capture best what is *meant* by emergent gravity in the physics literature. In section 3, we work through individual features of gravity which can at least naively be taken as suggestive of emergence. In section 4 we end by assessing the general structure of the arguments put forward.

2 What is emergent gravity?

2.1 Definition of emergent gravity

Emergent gravity is roughly taken to be the view that gravity arises due to the “collective action of the dynamics of more fundamental non-gravitational degrees of freedom” (Huggett and Wüthrich (2013)). The aim of this section is to sharpen this rough characterization of emergent gravity by providing answers to the following core questions: (i) “emergence of what?”, (ii) “emergence from what?” and (iii) “what kind of emergence?”. The first question asks what properties of gravitational physics are actually emergent. The second is concerned with the existence and nature of the underlying microstructure, and the third focuses on the emergence relation between the gravitational and non-gravitational degrees of freedom. In the next three subsections we treat these questions separately noting though that (i) and (ii) are not fully independent.

2.1.1 Emergent aspects of gravity

Which aspects of gravity are emergent? Needless to say the answer to this question depends on what we take gravity to be. A gravitational theory has for instance been defined by Thorne et al. (1973) as a spacetime theory that reproduces Kepler’s laws.³ In any case, our current best theory of gravity is GR, which formally encompasses a differentiable, 4-dimensional manifold with tensor fields on top. One of the tensor fields necessarily corresponds to a Lorentzian metric field g which has the usual chronogeometric meaning.⁴ The metric field satisfies the Einstein equations, $G_{ab} = \frac{8\pi G}{c^4} T_{ab}$, which relate the Einstein tensor G_{ab} to the stress-energy tensor T_{ab} obtained from all other tensor fields (the so-called matter fields). The matter fields themselves are subject to further (possibly coupled) dynamical equations.

One might think that ideally in an emergent model of GR all three aspects emerge together: the manifold, the metric field and its dynamics.⁵ The emergence of the manifold is, however, somewhat less important, since its physical significance is doubtful. Arguably, the (active) diffeomorphism invariance of fields in GR bereaves the manifold points of any “thisness” and renders it as a mathematical scaffolding structure to define physical fields on (cf. the hole argument of Earman and Norton (1987)).

In some approaches, such as the thermodynamic route to GR by Jacobson (1995), the metric field is first of all given ab initio, and then its (Einstein) equations of motion are derived. We think, however, that a genuine account of emergent gravity should include the emergence of both the metric field and its dynamics (rather than just the dynamics). This is not to say that Jacobson’s account is not a form of emergent gravity. In fact, in section 2.3 we categorize it as an emergent gravity scenario, because the thermodynamic interpretation of the dynamical equations associated to the metric field g naturally leads to the emergent nature of g itself.

In the next subsection, we will further see that not only the metric field and its dynamics are emergent aspects of gravity, but also the graviton – the spin-2 particle that mediates the gravitational force – should count as “gravitational” (and hence as emergent). According

³We thank Dennis Lehmkuhl for referring us to this definition. However, this definition seems to unfairly exclude modifications of GR (such as higher curvature theories and bimetric theories, like TeVeS) as gravitational theories.

⁴In particular, sufficiently accurate clocks travelling along a worldline segment γ will display a reading proportional to the worldline interval $\Delta s = \int_{\gamma} g_{ab} dX^a dX^b$. Whether the metric intrinsically contains chronogeometric significance or gains it dynamically, is up to philosophical debate. It is, however, not relevant for the characterization of emergent aspects of gravitation.

⁵Lam and Wüthrich (2017) argue that GR does not necessarily have to be recovered completely in such a scenario: a functionally adequate subset of notions of GR or empirically adequate modifications of it might be sufficient already.

to the spin-2 approach the metric field and its dynamics namely arise through the self-coupling of the graviton, cf. Feynman et al. (2003) and Weinberg (1979). If we viewed the spin-2 field as non-gravitational, then GR would emerge from quantized GR, which is however against the intuition behind emergent gravity.

2.1.2 Two types of microtheories

In the EG literature, gravity is unanimously taken to be emergent from (novel) microstructure underlying the metric degrees of freedom. For instance, Carlip (2014) defines emergent gravity as:

“[...] the basic picture is that gravity, and perhaps space or spacetime themselves, are collective manifestations of very different underlying degrees of freedom.” (p. 200)

In a similar spirit, Sindoni (2012) writes:

“As a provisional definition, we will intend as emergence of a given theory as a reorganization of the degrees of freedom of a certain underlying model in a way that leads to a regime in which the relevant degrees of freedom are qualitatively different from the microscopic ones.” (p. 2)

Padmanabhan (2008), Hu (2009) and Verlinde (2011) employ similar notions of EG. They tend to compare (or even equate) the idea of gravity having underlying microstructure (sometimes referred to as “atoms of spacetime”) to the sense in which thermodynamic or hydrodynamic systems have underlying microstructure.

However, despite the apparent consensus in the literature, this definition of emergent gravity is problematic.⁶ In general, the **received definition of emergent gravity** in terms of *very different* underlying degrees of freedom, or *novel* microstructure simply suffers from vagueness. It is vague both with respect to qualifiers like “very” or “novel”, and the characterization of what there actually is supposed to be over and above gravity (“different degrees of freedom” or “microstructure”). In the following, we give three major instances of changes in degrees of freedom for which it is not a priori clear whether they should be counted as “very different” degrees of freedom or not.

Firstly, the discretization of a theory of gravity changes its degrees of freedom in some sense. After all, the continuous version can only be reproduced upon coarse-graining again. But, according to the EG proponent, the degrees of freedom of a discretized theory should not be “very” or “qualitatively” different from the degrees of freedom of the original theory.

⁶Admittedly, some authors like Sindoni (2012) do not even attempt to give a fully worked-out definition.

If that were the case, loop quantum gravity and causal set theory would count as instances of emergent gravity, which would make the definition of EG too broad.

Secondly, quantization of a theory does to some extent change the theory’s degrees of freedom. As quantization generically leads to discretization of classically continuous quantities – such as energy, angular momentum, and possibly even spacetime structure – coarse-graining is typically part of restoring the classical limit (together with suppressing superposition effects). However, quantization does not *necessarily* entail a change of degrees of freedom, at least not in a Bohmian interpretation of quantum mechanics. Whether or not quantization leads to “qualitatively different” degrees of freedom depends on the specific interpretation of quantum mechanics. Preferably the definition should not depend upon the interpretation of quantum mechanics at play.

Thirdly, in a quantum field theory (QFT) setting, the renormalization group flow describes a change of theories under a coarse-graining operation. Coarse-graining here corresponds to integrating out high energy modes. If one starts with a renormalizable theory which is well-behaved up to arbitrarily high energies, then coarse-graining will produce effective field theories which are only valid up to a certain level of energy. Despite a change of its degrees of freedom under renormalization (energy modes are integrated out), the resulting theory will not necessarily involve “qualitatively” different degrees of freedom. Generally speaking, the renormalization group flow only changes the coupling constants of a quantum field theory. In the context of gravity, asymptotic safety claims that quantum general relativity (quantum GR⁷) is subject to a renormalization group flow such that it eventually runs into a UV-fixed point in theory space. The only difference in the degrees of freedom of low-energy and high-energy versions of quantum GR is then encoded in the UV cut-off of the corresponding theory.

The three examples above show that the received definition of EG is problematic, since it is unclear whether discretization, quantization, and renormalization lead to “qualitatively different” degrees of freedom. This issue is important because we want to distinguish EG from other quantum gravity accounts that arise from discretizing, quantizing or renormalizing GR. One could attempt to sharpen the characterization of the degrees of freedom underlying gravity by calling them “non-gravitational” (see also Huggett and Wüthrich (2013)), but this remains unsatisfactory as long as one has no clear notion of “gravitational”. We already alluded in the previous subsection to the problem that we cannot straightforwardly give a definition of “gravitational” in terms of GR alone, without run-

⁷By quantum GR, we denote any form of quantization approach turning GR into a quantum theory, including perturbative quantum gravity, canonical quantum gravity, loop quantum gravity, causal dynamical triangulation, and more; but not, for instance, string theory and causal set theory.

ning the risk of calling theories underlying GR non-gravitational, which are however not at all instances of emergent gravity. For example, the spin-2 approach, LQG, and causal set theory are possible quantum theories underlying GR that we would like to exclude as cases of emergent gravity.

We now propose to circumvent this issue by introducing a distinction between two types of microtheories, i.e. two types of theories underlying a theory T :⁸

We call a theory M_1 underlying a theory T a **type I microtheory** to T if and only if it is inspired⁹ from T (for instance through discretization, quantization or renormalization). An underlying theory M_2 to theory T is called a **type II microtheory** if and only if it is not directly inspired or motivated¹⁰ from T .

The structure linked to a microtheory is called **microstructure**.

A theory is called **emergent** if and only if there exists an underlying type II microtheory to it.

A type I microtheory to GR can now coherently be counted as gravitational, whereas a type II microtheory to GR is by definition non-gravitational since gravity emerges from it. Another advantage of this **new definition of emergent gravity** in terms of type II microtheories is that the microstructure associated to these theories is automatically novel, because it cannot be inspired from gravity. Examples of possible type I microtheories to GR are causal set theory, loop quantum gravity and other quantum gravity approaches proper. This is in line with the general idea of emergent gravity being something over and above quantum gravity proper (see Carlip (2014), Padmanabhan (2008), Jacobson (1995)). Furthermore, string theory would count as an instance of type II microtheory to GR.¹¹

As a final note we emphasize that, although proper quantum gravity and emergent gravity approaches are defined in a different way (as originating from type I and type II microtheories, respectively) they are not necessarily in conflict with one another.¹² They

⁸We necessarily require microtheories to T to reduce to T in an appropriate limit, such as a classical or low-energy limit.

⁹We would like to acknowledge Jeremy Butterfield for suggesting this term to us, and thank Claus Beisbart and Sam Fletcher for pressing us on our definition.

¹⁰One might argue that notions as “inspired” or “motivated” from are as vague as the notions “qualitatively” or “very different”, criticized above. However, despite their granted vagueness, the former notions allow for a more precise distinction between what is referred to as EG in the literature and what is not, as we demonstrate below.

¹¹Arguably, perturbative string theory counts as a type I microtheory, as its target space with a dynamic background metric is inspired from GR, whereas non-perturbative string theory is a type II microtheory underlying GR, since it does not involve any kind of Lorentzian metric.

¹²An often expressed sentiment by proponents of emergent gravity is that gravity should not be quantized if gravity is emergent (cf. Jacobson (1995), Padmanabhan (2008), Verlinde (2011)). One can, however, *prima facie*

are trivially compatible as any physical theory (and so even the most successful theory of quantum gravity proper) could in principle turn out to be underlied by a more fundamental theory.¹³ This general skepticism becomes most interesting if it is connected to concrete approaches which are standardly believed to be complete, but are then shown to have underlying structure nevertheless. Quantized gravity approaches like loop quantum gravity or causal dynamical triangulation (instances of QG) could in fact be underlied by string theory (an instance of EG) *even if* they are UV-complete. An appraisal of emergent gravity should take this general skeptic’s point into account, if it does not want emergent gravity to sound like a cheap option always available.

2.1.3 Different notions of emergence

With respect to the third question – “what kind of emergence?” – it is important to point out that philosophers and physicists use the word ‘emergent’ differently, and themselves not at all coherently either.¹⁴ *Philosophers* tend to define emergence as the failure of reduction. A theory (property) is then called emergent from another more fundamental theory (property) if the former is not derivable or explainable from the latter.

In contrast, *philosophers of physics* mostly work with Butterfield’s notion of emergence in terms of supervenience on the one hand, and novelty and robustness on the other hand. Supervenience of a theory (property) on another (standardly more fundamental) theory (property) means that there can be no change in the former theory (property) without a change in the latter theory (property). The problem with this notion of emergence, of course, lies in its deference of the difficulty in defining emergence to that of defining robustness and novelty (similar to how the philosopher’s sense of emergence seems to ultimately depend on the definition of reducibility). Following Butterfield (2011), we take ‘novel’ properties to mean features which are not contained in the microstructure, whereas ‘robust’ properties are properties that are not sensitive to (and somehow independent from) the specifics of the underlying structure.

Lastly, *physicists* often use the word ‘emergence’ in the sense of reducibility, that is derivability or explainability of a certain structure from another one. Take a string

equally quantize high-level as well as low-level theories. Whether to quantize or not to quantize, is a question of the theory’s formal nature, not of the level of reality we apply it to. For instance, quantization of condensed matter field theories is unproblematic and is well-known to yield quasi-particles, like phonons.

¹³This skepticism is quite conventional within a quantum field theory setting. In lack of knowledge of possible higher energy modes at play in a perturbatively considered QFT, a naively fundamental (i.e. renormalizable) QFT might equally be just effective (see Cao and Schweber (1993)).

¹⁴We would like to thank Karen Crowther for sharing an unpublished note, in which she distinguished between the philosopher’s, the physicist’s and the philosopher of physics’ usage of the word ‘emergence’. This distinction goes back to Nickles (1973).

theorist claiming that he/she derived general relativity from perturbative string theory in a low-energy limit: he/she will probably call this an instance of emergent gravity. More concretely, Seiberg (2006) takes space and time to be emergent if they “will not be present in the fundamental formulation of the theory and will appear as approximate semiclassical notions in the macroscopic world” (p. 1). Here, Seiberg identifies emergence of spacetime with its derivability (through semiclassical approximation) from a microscopic structure.

Which notion of emergence is applicable to emergent gravity? We should not forget that the notion of emergent gravity is introduced by physicists – not by philosophers. It is thus plainly wrong to presume that the notion of emergence in EG should be interpreted in the philosopher’s sense. Rather, it is the physicist’s notion which seems to apply here – as physicists expect GR to be reducible to more fundamental microscopic theories – thereby excluding the philosopher’s notion from the start.

Still, we do not take the physicist’s reduction to exhaustively specify the relationship between the metric and its (putative) underlying microstructure. There is arguably an aspect of novelty and robustness in the emerging metric structure which need to be accounted for.¹⁵ First, the non-gravitational degrees of freedom (as defined in the previous section) are qualitatively different from (and hence *novel* with respect to) the gravitational degrees of freedom, as the former are not inspired from the latter. Secondly, albeit more specifically, in thermodynamic approaches by Jacobson, Verlinde and Padmanabhan, gravity is viewed as *robust* with respect to the underlying microstructure, in the sense that it only depends on coarse-grained notions such as entropy and temperature. Thus, the notion of emergence in emergent gravity seems most suitably characterized by the philosopher of physics’ sense. This also seems to capture best how advocates of EG models use the word ‘emergence’.

2.2 Relation to unification, emergent spacetime and EFT

In the previous section we tried to get hold of the notion of emergent gravity – as used in practice – by characterizing it through a microtheory which is underlying GR but not inspired from it (type II microtheory). We will now compare the notion of emergent gravity to related ones.

Unification. The quest for the unification of gravity with the other fundamental interactions of Nature is logically compatible with emergent gravity, as defined above. If gravity is emergent, and unification is realized at a certain energy scale, then also the

¹⁵Some physicists might also have this notion of emergence in mind. The previous distinction between the three senses of emergence should be understood as a general labeling, not specific to the context of QG. We thank Christian Wüthrich for pressing us on this point.

other interactions might have type II microstructure. In a sense the idea of emergent gravity then loses its pre-eminent status. For example, in the standard grand unification scenario of string theory all interactions are on a par with each other and are equally emergent from the underlying microstructure of strings and D-branes. In other emergent gravity approaches such as those of Verlinde (2011, 2017) and Jacobson (1995, 2016), typically only gravity is viewed as emergent whereas the other forces are treated as fundamental. This, however, does not exclude the scenario that all forces eventually emerge from a joint microstructure. But the approach to such a unified description could be different from standard grand unification, in the sense that gravity is expected to emerge prior and in a qualitatively different way from the other forces when going up the energy scale.

Emergent spacetime. In the current GR era ‘gravity’ is often used synonymously with ‘spacetime’ so that, at first sight, ‘emergent spacetime’ and ‘emergent gravity’ seem to be synonyms. The usage of the term ‘emergent spacetime’, however, comprises many different senses in which typical aspects of spacetime – such as dimensionality, causal and chronogeometric structure, continuity of space and time, and more – turn out to be non-fundamental. For example, Huggett and Wüthrich (2013) use the term for both the emergence of classical spacetime from their quantum versions, as in LQG, *and* the emergence of spacetime from another spatiotemporal structure, as in condensed matter approaches, or non-spatiotemporal structure, as in non-commutative geometry. On our account the former type of emergence is not an instance of EG, whereas the latter is.¹⁶

Effective field theory. Within a particle physics or quantum field theory approach to GR the Einstein-Hilbert action is standardly rendered as a low-energy effective action (cf. Burgess (2004)). Similarly, perturbative quantum GR is then simply understood as a low-energy effective field theory (EFT) (cf. Donoghue (2012)). The EFT view and emergent gravity only make contact if the gravitational low energies degrees of freedom are qualitatively different from the high-energy degrees of freedoms (for example if the spin-2 particle is a composite particle). Mere effectiveness of a field theory, however, does not yet entail that its higher energetic versions feature qualitatively different degrees of freedom. For example, in the asymptotic safety approach to gravity the gravitational action changes with the energy scale while the metric degrees of freedom remain fundamental.

¹⁶Emergence of spacetime does not necessarily entail the emergence of matter fields. Only without a straightforward rule to assign the currently known matter fields to the underlying degrees of freedom of spacetime, the emergence of spacetime implies the emergent nature of matter as well.

2.3 Taxonomy of accounts: top-down vs. bottom-up

In the previous sections we have given a broad definition of emergent gravity and demarcated this notion from related ones. In this section we give a quick overview of the several models of emergent gravity and different approaches towards finding such models.

One could taxonomize the emergent gravity program in (at least) two natural ways: either by qualitatively distinguishing different models from each other, or by differentiating between research strategies for establishing the emergent nature of gravity. Regarding the first taxonomy, Bain (2013b) and Carlip (2014) have classified different models in terms of the nature of the underlying microstructure. In some models the underlying structure is spatiotemporal, whereas in others any sense of ordinary spacetime itself is emergent and hence the microstructure is necessarily non-spatiotemporal.

In this paper we are more concerned with the second kind of taxonomy. We distinguish between two different strategies towards emergent gravity: (1) bottom-up accounts that start from the underlying microstructure and work their way upwards towards GR; and (2) top-down accounts that start from the emergent gravitational structure and try to find universal features of the underlying microstructure.¹⁷ Since finding a detailed microscopic model of emergent gravity is generally quite hard, some authors just study analogies between gravitational features of GR (or QFT in curved spacetime) and standardly emergent structures (such as thermodynamics or hydrodynamics). Such top-down accounts do not establish gravity to be emergent, but should often be taken as hints towards the emergence of gravity. Conversely, bottom-up accounts typically start from microscopic theories such as condensed matter physics, quantum information theory or string theory, and try to derive GR.

Below we list different top-down and bottom-up accounts (see Carlip (2014) for more EG accounts):

Bottom-up accounts. The archetypical example is induced gravity by Sakharov (1968), and variants thereof (see Visser (2002) for a review), in which the Einstein-Hilbert action appears as a term in the one-loop effective action of matter fields. Other examples include: non-perturbative string theory, such as the matrix model for M-theory by Banks et al. (1997); some version of the AdS/CFT correspondence, where the conformal field theory is viewed as fundamental and the bulk spacetime as emergent (cf. Berenstein (2006), Horowitz and Polchinski (2006)); and possibly analog models of gravity, such as acoustics in a moving fluid (cf. Unruh (1995), Barcelo et al. (2005)).

¹⁷‘Bottom’ (‘top’) here refers to UV (IR) physics, which is opposite the convention of Hu (2009).

Top-down accounts. Typical examples in this category are derivations of the Einstein equations from thermodynamics or hydrodynamics, such as by Jacobson (1995), Padmanabhan (2008) and Verlinde (2011). Recently, Jacobson (2016) proposed a new derivation of the Einstein equations from the entanglement properties of the vacuum. Another approach by Hu (2009) is to extend the semi-classical field equations by adding fluctuation terms on the energy/matter side.

3 Hints towards an emergent nature

In this section we will construct and criticize arguments to the effect of an emergent gravity paradigm. Naively speaking, there are two strategies for arguing that gravity is emergent: (1) to look for indications in our best theories of gravity (GR, its higher derivative extensions or semi-classical gravity) for an emergent paradigm; and (2) to directly present a model for the supposed microstructure from which gravity emerges.

The first strategy is to look for analogies between (X) gravitational theories and (Y) other theories said to have microstructure, such as thermodynamics or hydrodynamics. One then hopes to demonstrate the emergent nature of gravity by establishing a sufficiently strong analogy between (X) and (Y) in a relevant sense. Certain features of gravity – mainly its universality, perturbative non-renormalizability, its thermodynamic features and its holographic nature – allow for mounting such arguments from analogy. Arguably, these gravitational features even remain ‘mysteries’ in a fundamental take on the metric and appear to be in need of explanation, which however can be easily achieved in an emergent gravity framework. We call this the *top-down strategy* (similarly to the already mentioned top-down accounts).

But how could a top-down strategy ever do more than establishing the plausibility of an emergent nature? Strictly speaking, it cannot do more. But philosophy of physics, among other things, has to make sense of the conceptual aspects of a theory, and read the signals that come from these conceptual aspects. If there are thus signals calling for an emergent reading of gravity, then the philosopher of physics has to make them public. The real question is, rather, how strong these signals are, i.e. how to construct arguments from them and consequently assess their relative strengths.

The second strategy at worst only gives a proof-of-concept for the emergence of gravity. At best, it provides a genuine candidate for the microstructure underlying gravity. String/M-theory as well as condensed matter systems from which gravity emerges are examples of such candidates. We call this the *bottom-up strategy* (similarly to the above mentioned bottom-up accounts).

In the current paper we only follow the top-down strategy. Below we analyse four hints for the emergent nature of gravity: the universality of gravity (subsection 3.1), the perturbative non-renormalizability (subsection 3.2), black hole thermodynamics (subsection 3.3), and the holographic principle (subsection 3.4). We will see that the latter three hints are closely related.

3.1 Universality of gravity

Universal attraction is, of course, a well-known property of gravity, but first we want to clarify what is exactly meant by this type of universality, both in Newtonian gravity and in general relativity. We will then be sufficiently equipped to assess whether the universal nature of gravity hints at its emergence.

In Newtonian gravity, one can distinguish between two aspects of the universality of the gravitational force: its universal scope and its universal strength. The first universal aspect is that all particles have a gravitational mass, and hence attract each other as described by Newton’s law of gravitation

$$F = G \frac{m_1 m_2}{r^2} .$$

Hence, Newton’s law of gravitation is *universal in scope* as it applies to any two Newtonian objects of a given mass. The other fundamental forces (electromagnetic, weak and strong force) do not have such a general scope as gravity. The electromagnetic force, for example, only applies to particles with electric charge. A related important difference between gravitation and electromagnetism is that there exists only one type of gravitational charge, whereas there are two types of electric charges in nature.

The second universal feature of Newtonian gravity is *its universal strength*: gravity is ‘equally strong’ for different types of freely falling bodies. More precisely, two test bodies in a gravitational field which are not acted upon by other forces, such as air resistance, move with the same acceleration irrespective of their mass and internal composition. In the literature this property is known as the ‘weak equivalence principle’ (WEP) or as the ‘universality of free fall’, and it holds as a consequence of the proportionality between inertial and gravitational mass.¹⁸ Within the ontology of Newtonian theory, WEP is an extra assumption on top of the three laws of Newtonian mechanics, which is based on empirical evidence (in contrast to GR, where it is built into the formalism). Taken together, the universal scope of gravity and its universal strength imply that *all* test bodies in a gravitational field fall equally fast in the absence of other forces.

In general relativity, or any sound relativistic theory of gravity, both of these universal

¹⁸Cf. Will (2006).

features are of course present, albeit in a different guise. First, the universal scope is extended to everything that carries energy: the gravitational field in GR acts on *all* sources of stress-energy.¹⁹ Second, the WEP is also an indisputable principle in general relativity, and it even forms one of the foundations of the theory. One of the main differences, though, between Newtonian gravity and general relativity is that freely falling bodies in GR are moving inertially with respect to a local inertial frame, and hence no gravitational force is acting on them (in contrast to the Newtonian picture of gravity as a force). In fact, the trajectories of freely falling test bodies in GR are identified with timelike geodesics of the metric. Hence, the WEP does not follow in GR from the equality between the inertial and gravitational mass – since these are not fundamental notions in GR – but it is rather embodied by the fact that geodesic motion is independent of the particular constitution of test bodies.²⁰

Furthermore, in standard treatments of GR the term ‘universality’ is often attributed to the coupling of the metric to matter fields. Will (2006), for example, defines the universal nature of this coupling as follows:²¹

“The property that all non-gravitational fields should couple in the same manner to a single gravitational field is sometimes called ‘universal coupling’.” (p. 68)

In this definition both of the universal features mentioned above are apparent: the coupling has a universal scope since it applies to *all* non-gravitational fields, and it has universal strength as it always occurs *in the same manner*. Admittedly, the characterization of universal strength is rather vague, but what is probably meant here by “the same manner” is that the coupling between matter fields and the one gravitational field is always *minimal* (which excludes curvature-matter couplings in the equations of motion of the matter fields). It should be noted, though, that the notion of universal coupling actually goes beyond the two universal features considered above. Universal coupling is namely a central aspect of the Einstein equivalence principle (EEP), which is a stronger claim than the weak equivalence principle.²² The EEP states that in any local Lorentz frames the non-gravitational laws of physics must take on their familiar special relativistic forms (cf. Misner et al. (1973)). This principle ensures that the outcome of any non-gravitational experiment is independent of

¹⁹In GR (or any other metric theory of gravity) the metric field does not only act on sources of stress-energy, but it is also acted upon by these sources. Hence, the metric is a dynamical field in such theories of gravity.

²⁰Following Brown and Read (2016).

²¹This definition of universal coupling dates back to work by Thorne et al. (1973). They have also proven that a Lagrangian-based relativistic theory of gravity is universally coupled if and only if it is a metric theory.

²²It is conjectured, though, by Leonard Schiff that any complete and self-consistent gravitation theory that obeys WEP must also obey EEP. See Will (2006) for an overview of the plausibility arguments in favour of Schiff’s conjecture.

the spacetime location, and is hence more powerful than the WEP.²³

Now, the universality of gravity has been put forward by some authors as a hint towards the emergence of gravity. We will analyze their arguments, and specifically pay attention to which aspect of universality (universal scope or strength) is being used in the arguments. In our analysis we take ‘universal coupling’ to be the most precise definition of universality of gravity in GR.

To begin with, Verlinde (2011) argues for emergent gravity as follows in his introduction:

“Of all forces of Nature gravity is clearly the most universal. Gravity influences and is influenced by everything that carries an energy, and is intimately connected with the structure of space-time. The universal nature of gravity is also demonstrated by the fact that its basic equations closely resemble the laws of thermodynamics and hydrodynamics. So far, there has not been a clear explanation for this resemblance.” (p. 1)

Admittedly, Verlinde at first sight seems to conflate (or at least directly equate) two different notions of universality, namely the universal coupling of the gravitational interaction and the universal applicability of thermodynamics and hydrodynamics (also to GR). However, we take him to promote the view that universal coupling of gravity, on the one hand, and the applicability of thermodynamics to GR, on the other hand, find a common cause explanation in some kind of coarse-graining process. He subsequently points out that these universal aspects of gravity would be explained by the emergence of gravity.

The idea that the universal attraction of gravity hints at underlying (non-trivial) microstructure already dates back (at least) to Feynman. In section 1.5 of the Lectures on Gravitation Feynman et al. (2003) suggested the following:

“The fact of a universal attraction might remind us of the situation in molecular physics; we know that all molecules attract one another by a force which at long distances goes like $1/r^6$. This we understand in terms of dipole moments which are induced by fluctuations in the charge distributions of molecules. That this is universal is well known from the fact that all substances may be made to condense by cooling them sufficiently. Well, one possibility is that gravitation may be some attraction due to similar fluctuations in something, we do not know just what, perhaps having to do with charge.” (p. 15)

²³As pointed out by Read et al. (2017), the EEP in its usual form – contra common belief – does not hold strictly. The standard Maxwell wave equations in curved spacetime, for instance, do feature curvature couplings, and thus provide an example of equations of motion in the general relativistic context which – even at a point – do not take their special relativistic form. As a result, the EEP is actually only realized effectively in GR, namely in the case that local curvature effects are negligible.

Feynman raises the interesting possibility that gravity might be a high-level interaction, or in his words “that gravitation is a consequence of something that we already know, but that we have not calculated correctly” (p. 15). To make this point, he suggests that the universal coupling property of gravity might equally be a result of coarse-graining as the dipole expansion for a composition of molecules. *All* molecules attract each other with the same force – given by the $1/r^6$ -dipole expansion – when seen from a sufficiently large distance. Since the same applies to gravity – all particles attract each other with the same $1/r^2$ -force – one could think that this is also the result of a multipole expansion, induced by fluctuations in some underlying microstructure. Finally, note that Feynman uses both aspects of universality in his argument: the universal scope (*all* molecules attract each other) and the universal strength (the force between them is always *the same*).

Now, based on these ideas of Verlinde and Feynman we can construct the following argument from universality towards emergent gravity. We will use the similarity between hydrodynamics and gravitation to make the argument more concrete. First of all, hydrodynamics and gravity are both universal in scope and in strength. Also in hydrodynamics of uncharged fluids the motion of *all* test bodies in a fluid is in a sense *the same*, i.e. their trajectory does not depend on their internal structure or constitution. In hydrodynamics, the universal scope and strength are the result of the coarse-graining of the underlying microstructure. Different microstructures (such as different types of atoms) lead to the same smooth hydrodynamic description. In a similar fashion, the universal coupling of gravity could also be the result of coarse-graining. This argument for EG can be formalized as follows:

1. Universality in hydrodynamics implies coarse-graining.
2. Coarse graining implies type II microstructure.
3. Universal coupling of gravity is similar to universality in hydrodynamics.

-
4. \therefore Universal coupling of gravity implies type II microstructure.

We take premise (1) to be largely uncontroversial. Premise (2) is, however, easy to attack as renormalization group theory provides a straightforward counterexample to it. Coarse-graining could also follow from the renormalization group flow, and hence be due to type I microstructure (as discussed in section 2.1.2). Premise (3) can be attacked from a geometrical view²⁴ of the metric field: the metric field in some sense plays the role of spacetime and, therefore, the universal coupling between the metric field and matter fields is to be expected. According to this view, the universality of gravity is a natural consequence of

²⁴See for example Maudlin (2012).

the fact that it relates to the general framework of spacetime, and is hence not similar to the universality in hydrodynamics. However, this specific criticism on premise (3) is moot if the geometric view of the metric field in GR is moot. A rebuttal of the geometric view can, for instance, be based on the dynamical approach to relativity by Brown (2005).

Thus, we conclude that the universal coupling property of gravity is with no doubt striking. But whether it can be taken as a valuable hint for the high-level nature of gravity is far from clear.

3.2 Perturbative non-renormalizability

A generic quantum field theory contains divergences. In a perturbative series expansion of a QFT, these divergences manifest themselves in loop terms, which involve an integration over arbitrarily high momenta. As a sort of technical cure, these divergences can be absorbed into a proper redefinition of the coupling constants of the QFT. A theory is commonly called *renormalizable* if only a finite number of coupling constants need to be redefined to remove the mentioned divergences.²⁵

A still influential tradition in quantum field theory, dating back to Bethe, regards only renormalizable quantum field theories as good candidates for fundamental theories (cf. Cao (1998), section 8.3).²⁶ Such a view builds on the presuppositions that fundamental theories should be predictive, and that only renormalizable theories can be predictive in arbitrary energy regimes. The latter presupposition is based on the strict definition of predictivity that a theory possesses predictive power if it contains a finite number of parameters – which makes (non-)renormalizable theories, by definition, (non-)predictive.

This traditional view has important consequences for the (non)-fundamentality of GR. It is generally believed that GR, viewed as a quantum field theory, is perturbatively non-renormalizable. This is often presented as a consequence of the negative mass dimension of Newton’s constant (gravity’s coupling constant).²⁷ Although the power-counting argument is generically a good criterion to judge whether a theory is renormalizable or not; it is rather heuristic, and neither a necessary nor a sufficient criterion for renormalizability (see

²⁵The definition of renormalizability is independent of the perturbative/non-perturbative distinction. In particular, a perturbatively renormalizable QFT is generically still divergent in the sense that the perturbative series is a divergent but asymptotic series. Note though that a perturbative series may still define a QFT uniquely.

²⁶The renormalizability of a theory alone cannot completely establish its fundamentality. This follows from the (physical) possibility that there exist higher, but strongly suppressed, energy modes which are not captured by the QFT under consideration. Hence, it might be the case that the QFT is the result of coarse-graining from yet another higher energetic QFT. In this respect, Cao and Schweber (1993) rightly advocate (1) an epistemological anti-foundationalism and (2) a plurality in QFT ontology.

²⁷The mass dimension of Newton’s constant is $[G] = 2 - d$, where d is the number of spacetime dimensions, which means that a gravitational theory is power counting non-renormalizable for $d > 2$.

for instance Peskin and Schroeder (1995), section 10.1). More precisely, quantum GR has been established to be non-renormalizable, up to 2-loop order (Goroff and Sagnotti (1986)). It is only an expectation, albeit not a theorem, that quantum GR is also divergent at all orders in a perturbative expansion.

Thus, according to the traditional view, GR's (expected) non-renormalizability makes it a non-fundamental theory and thus hints at an emergent nature of the theory. The argument can be formalized as follows:

1. Non-renormalizability implies non-predictivity.
2. Non-predictivity implies non-fundamentality.
3. (Quantum) GR is non-renormalizable.
-
4. \therefore (Quantum) GR is non-fundamental.

Let us now assess the validity of the three assumptions that feature in this argument. Premise (1) is true by definition. We take this definition of (non-)predictivity to be rather uncontroversial, with the caveat, of course, that GR can be predictive at a certain low-energy scale (say in the solar system) but non-predictive at the Planck scale. Premise (2) only seems acceptable if one regards physical theories to be solely constructed for predictive purposes, and nothing more. Only then can we straightforwardly categorize a theory which is not predictive to arbitrary high energies as non-fundamental. Standard accounts of physical theories, though, also value other explanatory features than mere predictivity, such as their ability to explain structures of the world. Therefore, a more generous – in other words, less narrow – account of theories than that of Bethe can simply accept that a fundamental theory of the world turns out to be unpredictable in certain regimes, as long as it sufficiently satisfactory in other aspects.

Whereas premise (2) is subject to philosophical controversy, premise (3) is simply an open question within the current state of physics. Quantum GR has so far only been convincingly established to be perturbatively non-renormalizable, and there exist after all paradigm cases of theories which are perturbatively non-renormalizable, but non-perturbatively renormalizable, such as the Parisi (1975) model. Whereas many quantizers of gravity like to stress this possibility of the non-perturbative renormalizability, in particular string theorists nevertheless keep up the analogy between non-renormalizable gravity and non-renormalizable Fermi 4-theory – which was replaced by a renormalizable higher energy theory – in demand for new, underlying structure to GR.²⁸

²⁸In this tradition, Zee (2010) for instance pathetically asks: “Fermi’s theory cried out, and the new physics turned out to be the electroweak theory. Einstein’s theory is now crying out. Will the new physics turn out to be string theory?” (p. 157)

Arguments for the non-perturbative renormalizability of gravity have been put forward building on holography in one sense or another. The entropy argument by Shomer (2008) – originally going back to Aharony and Banks (1999) – for instance tries to establish that d -dimensional GR cannot be turned into a d -dimensional renormalizable quantum field theory through cashing out an incompatibility of the Bekenstein-Hawking entropy formula with a generic entropy formula for a renormalizable theory. However, arguably the semi-classical presuppositions behind this reasoning, such as the violent extrapolation of the validity of the Bekenstein-Hawking formula to high energies, render the argument problematic.²⁹

Going beyond a mere stance on whether GR is renormalizable or not, Verlinde (2011) even claims that today’s standard framework for renormalization, the Wilsonian framework, is inapplicable to gravity. He points out that (1) Wilsonian renormalization rests on the insensitivity of large scale physics from low scale physics, while (2) GR *is* sensitive to short distance physics as it ‘knows’ about the (number of) fundamental degrees of freedom underlying gravity. The latter statement is based on holographic considerations, which imply that there is a limited amount of underlying degrees of freedom associated to each region of space. It is in this sense that gravity knows about the more fundamental level.

3.3 Black hole thermodynamics

Thermodynamic phenomena are standardly seen to be underlied by non-trivial microstructure.³⁰ For instance, temperature is viewed as a measure of the averaged motion of underlying particles, and in the Boltzmannian view entropy is a measure of the amount of underlying microstates for a given macroscopic system. In this section we will explore in how far the thermodynamic properties of black holes (and those of causal horizons in general)³¹ can, as a consequence, be taken to hint at underlying microstructure of the gravitational field. The structure of the argument can be formalized as follows:

1. Thermodynamic systems have underlying type II microstructure.
2. The laws of black hole mechanics are similar to those of thermodynamics.
-
3. \therefore Black hole spacetimes have underlying type II microstructure.

Note that not only the relationship between thermodynamics proper and black hole thermodynamics, but also the linkage between thermodynamics proper and type II microstructure

²⁹Cf. Doboszewski and Linnemann (2016).

³⁰Whether thermodynamics is reducible – on the level of theories – to statistical mechanics is a slightly different and highly intricate question. This is a long-lasting debate in the philosophy community.

³¹Cf. Jacobson and Parentani (2003).

is far from uncontroversial, as we will point out in the following. However before we do so, we mention a stronger version of the analogical argument, which is not based on black hole thermodynamics but on the thermodynamic properties of local Rindler causal horizons through each spacetime point. In the latter setup, Jacobson (1995) has derived the field equations in GR from a thermodynamic first law that governs these local horizons, and thereby he interpreted the Einstein equation as an equation of state. This seminal derivation suggests the following analogical argument:

1. Thermodynamic systems have underlying type II microstructure.
 2. At each point in spacetime the Einstein equation can be interpreted as a thermodynamic equation of state.
-
3. \therefore The Einstein equation is due to underlying type II microstructure of GR.

Although this argument is more far-reaching than the first argument above, in the sense that it applies to the full Einstein equation instead of a specific solution to it, the thermodynamics of local causal horizons is also less well established than that of black holes. Therefore, in the rest of this section we will mostly focus on black hole thermodynamics.

Black hole thermodynamics consists of a group of laws, originally derived in general relativity by Bardeen et al. (1973), which are analogous to (variants of) the standard laws of thermodynamics (see Table 1). Upon results from semi-classical calculations – especially the derivation of black hole temperature by Hawking (1975) – the temperature and entropy featuring in the laws of black hole mechanics were accepted as proper thermodynamic quantities. As a consequence, black hole thermodynamics as such was recognized by the majority of the physics community as thermodynamics proper.³²

Table 1: Overview of the analogies between standard and black hole laws of thermodynamics (largely taken from Kiefer (2007), p. 202)

	Standard thermodynamics	Black hole thermodynamics
Zeroth law:	T constant on a body in thermal equilibrium	κ constant on the horizon of a black hole
First law:	$dE = TdS - pdV + \mu dN$	$dM = \frac{\kappa}{8\pi G}dA + \Omega_H dJ + \Phi_H dQ$
Second law:	$dS \geq 0$	$dA \geq 0$
Third law:	$T = 0$ cannot be reached	$\kappa = 0$ cannot be reached

Contra the received view, Dougherty and Callender (2016) identify two potential problems with black hole thermodynamics. First, they argue that the putative thermodynamic laws of black holes only represent impoverished versions of the actual thermodynamic laws.

³²See the textbook by Kiefer (2007) or review by Wald (2001).

Second, they rightly point out that the information-theoretic interpretation of black hole entropy stands in conflict with the standard non-epistemic interpretation of entropy in thermodynamics. According to a non-epistemic view on thermodynamics, the description of thermodynamic processes, such as heat cycles in engines, should not be subject to an agent’s knowledge. However, proponents of black hole thermodynamics, such as Bekenstein or Wald, largely maintain a purely agent-based, information-theoretic notion of entropy. Bekenstein (1972), for example, justifies the attribution of entropy to the black hole horizon by stating that otherwise information is lost for an agent when something has fallen through the event horizon. There are, of course, ways to attack such an information-theoretic reasoning towards entropy, and an agent-based view on thermodynamics more generally (see also Wüthrich (2017) for an appraisal of Bekenstein’s information theoretic argument).

Rather than taking a general stance in this debate, we are only interested in whether it would have any effect on the analogical argument mentioned above. It seems to us that, generally speaking, the analogical reasoning can still go through even if not all black holes laws are proper thermodynamic laws. What seems most relevant to us is that the analogy holds sufficiently with respect to the notions of entropy implemented in black hole thermodynamics and thermodynamics proper. Hence, for the evaluation of the analogical argument the decisive questions are (1) whether black hole entropy is a proper thermodynamic entropy, and (2) whether thermodynamic entropy is linked to underlying type II microstructure.

Let us first consider question (1). Against Dougherty and Callender (2016), recently Prunkl and Timpson (2017), Curiel (2017) and Wallace (2017b) have argued for a standard thermodynamic take on black hole entropy. Prunkl and Timpson (2017) mention that black holes can feature in thermodynamic Carnot cycles just like any other thermodynamic system. Moreover, Wallace (2017b) stresses that in all relevant respects black hole thermodynamics is not at all conceptually worse off than thermodynamics proper (granting of course that there is no empirical corroboration of black hole thermodynamics at the moment, nor likely to be had in the near future). We do not want to repeat their specific arguments here, but we take them to convincingly establish that black hole entropy is thermodynamic entropy (in accordance with the consensus in the physics community).

With regard to question (2), Chirco et al. (2014) acknowledge that thermodynamic theories – as far as we know – have microstructure, but claim that the underlying degrees of freedom need not be novel or “hidden”. Thermal radiation, for example, has an underlying microstructure which allegedly arises from quantization of the electromagnetic field, and is hence not novel. Chirco et al. (2014) state that, by analogy, the microstructure underlying black hole thermodynamics (or horizon thermodynamics) can arise from the quantization of

the metric field alone. We can defuse the sort of analogical reasoning at play here through usage of our terminology: type I microstructure corresponds to no “hidden degrees of freedom” in the language of Chirco et al. (2014), whereas type II microstructure corresponds to hidden degrees of freedom. Thermal radiation is underlied by electrodynamic fields, and they in turn can be seen to be underlied by quantum electrodynamic structure. Quantum electrodynamic structure is, thus, microstructure of type I with respect to classical electrodynamic structure, but type II microstructure with respect to thermal radiation. There is namely no sense in which electrodynamics can be seen to be inspired from radiation thermodynamics. This contradicts the claim of Chirco et al. (2014) that thermal radiation has type I microstructure (or no hidden degrees of freedom in their terminology). So their analogical reasoning cannot even take off.

However, one could argue from LQG (as Chirco et al. (2014) do as well) or already from effective quantum field theory that black hole spacetimes (or the gravitational metric structure per se) have type I microstructure *simpliciter*. Take the latter case: in an effective quantum field theory view on GR, the Bekenstein-Hawking formula for entropy is derivable from the partition function instantiated by the QFT path integral at play (see Gibbons and Hawking (1977)). The calculation is formally in no way different from usual derivations of entropy from the partition function in statistical mechanics (see for instance Wallace (2017a) for a pedagogical introduction). But conceptually speaking, it must strike one as odd that a *mere* formal analogy between QFT and statistical mechanics allows for calculating a meaningful statistical mechanical property (i.e. entropy) from a quantum field theoretic approach, *without* there being any meaningful quantum field theoretical analogon to this property. However, if one accepts the partition function formally obtained from a quantum field path integral as a true statistical mechanical object, one should interpret the summation over different metric and matter fields in the partition function as the usual summation over all possible microstates. This would mean that the quantum structure obtained from quantizing the metric field of GR is sufficient microstructure to explain its thermodynamic behaviour.

This quantum field theory approach is, however, in conflict with the typical statistical mechanical derivation of thermodynamic entropy. The latter, namely, is based on type II microstructure, such as the positions of particles for a gas. If black hole entropy is a standard thermodynamic entropy then you might say that it also requires a standard statistical mechanical explanation in terms of type II microstructure. String theory indeed provides such a typical statistical mechanical derivation, because the Bekenstein-Hawking formula has been derived by counting certain string microstates or D-brane configurations (see Horowitz and Polchinski (2006) for a review). In other words, black hole entropy is

interpreted in string theory as Boltzmann entropy, and hence is due to underlying type II microstructure. But, of course, we can only concur that it remains a physical possibility that black hole entropy arises due to type I microstructure, such as entanglement (see also Bianchi and Myers (2014) for a discussion on this topic).

3.4 Holographic principle

The holographic principle is generally viewed as an important guiding principle in the search for a theory of quantum gravity. It puts a precise limit on the number of fundamental degrees of freedom associated to spacetime regions, by stating that the maximal amount of information inside a given spacetime region cannot exceed the area of its boundary (measured in Planck units). In this section we are particularly interested in whether the holographic principle entails anything about the nature of a putative microstructure of GR. Before we explain how the principle might be viewed as a hint for type II microstructure – and hence for emergent gravity – we first give a more precise definition of it.

The *holographic principle*, originally proposed by 't Hooft (1993) and followed up by Susskind (1995), is based on the following simple argument. Consider a spatial spherically symmetric volume V in an asymptotically flat spacetime that is bounded by a surface B . Now the entropy and number of degrees of freedom inside B are maximized if the volume contains a black hole. This is because, if one tries to increase the entropy by putting in more and more matter, at some point it will collapse into a black hole. As we saw in the previous section, thermodynamical arguments suggest that black holes carry an entropy proportional to their horizon area $A(B)$. If one were to throw more matter into the black hole, then its size would grow and it would not fit inside the surface B any more. Therefore, by dropping the specific assumption of spherical symmetry and asymptotic structure, 't Hooft concluded that the number of degrees of freedom contained in any spatial region will not exceed the area of the region's boundary.

Later Bousso (2002) gave the holographic principle a covariant formulation. Firstly, the holographic principle in its most general form does not hold for spacelike volumes, but for light sheets and their associated spacelike boundary. Secondly, a rendering of the holographic principle in terms of degrees of freedom already amounts to an interpretation of (black hole) entropy in a general-relativistic setting. Whether entropy in a gravitational context is due to a Boltzmannian-type counting procedure, due to entanglement across the surface or even a third option, is however up to debate (string theorists typically promote the first, LQG proponents favour the second one or even both at once, see Rovelli (2004)). The interpretation-neutral version of the covariant entropy bound now reads as follows:

“The entropy on any light sheet of a surface B will not exceed the area of B ”³³

$$S[L(B)] \leq \frac{k_B c^3}{\hbar} \frac{A(B)}{4G}.$$

Here L is a light sheet “constructed by following light rays that emanate from the [spacelike] surface B , as long as they are not expanding”. The covariant entropy bound can be shown to reduce to the original version of the holographic principle by ’t Hooft and Susskind for spatial volumes and their associated boundary surfaces.

The holographic principle has led to important progress in quantum gravity research. Already ’t Hooft put forward a much stronger statement than the entropy bound, namely that all the fundamental degrees of freedom inside a region can be described by a quantum mechanical theory living on the region’s boundary (which is now known as a *holographic duality*). Such holographic duality relations have indeed been found, most notably the AdS/CFT correspondence, and play a central role in quantum gravity research nowadays.³⁴ We want to stress though that the holographic principle (or, more precisely, the covariant entropy bound) should not be conflated with holographic duality relations. The latter is a conjecture about the equivalence of two, at first sight, unrelated theories, whereas the former is a bound on the gravitational entropy.

Physicists often interpret the holographic relation as an emergent relation, in the sense that a theory of gravity in a given spacetime emerges from a quantum theory without gravity on the boundary. However, Dieks et al. (2015) and de Haro (2016) among others³⁵ worked out that emergence (also in our sense of underlying microstructure) and holographic duality are *prima facie* independent notions. Only if the holographic duality is approximate at a certain level, it can turn out that either the bulk or the boundary is more fundamental. In the latter case, the bulk system can be seen to have underlying microstructure in terms of the (microscopic) boundary system. The judge is still out, though, on whether the AdS/CFT correspondence is an emergence relation or not, because it is still only proven in a large N limit (for semi-classical bulk gravity).

We take it, however, that the holographic principle itself (and not just the specific notion of holographic duality) already contains a hint for emergent gravity which may be formalized roughly as follows (cf. the introduction of Bousso (2002) for a similar argument):

³³Bousso (2002), p. 19.

³⁴See Maldacena (1999) for the original derivation of the AdS/CFT correspondence, and Aharony et al. (2000) for a comprehensive review.

³⁵See also the earlier work by Rickles (2013) and Teh (2013).

1. In a local QFT the total entropy of a spatial region scales with its volume.
2. The holographic principle implies that the total entropy associated to a spatial region does not scale with its volume, but with the area of the boundary of the region.
-
3. \therefore The microtheory underlying gravity is non-local.
-
4. \therefore Non-local degrees of freedom are typically an example of type II microstructure.

Let us expand on this. We defined emergent gravity in section 2 by means of type II microstructure. Now the holographic principle is not only concerned with the *number* of fundamental degrees of freedom of quantum gravity, but it is also suggestive of the *kind* of degrees of freedom that underly GR. As is well known, the holographic principle is in conflict with the conventional wisdom in statistical mechanics or quantum field theory that the number of degrees of freedom scales with the volume size of a system. Conventional quantum field theories are interaction-local.³⁶ Since degrees of freedom are linked to each spatial point, the information content of a spatial region grows with the volume. The holographic principle, on the other hand, implies that the entropy is bounded by the area of spacelike surfaces. Hence, the holographic principle casts doubt on the locality of the fundamental degrees of freedom of gravity, and this might mean that the underlying microstructure of GR is ‘non-trivial’ (or of type II).

Still, the non-locality of the underlying degrees of freedom does not need to imply that the microstructure is of type II. It is true that in string theory, for example, black hole entropy is accounted for by non-local fundamental objects such as strings and D-branes. However, LQG (which is arguably a form of quantum GR) does for instance feature interaction non-locality but is based on quantization of the standard gravitational degrees of freedom. So whether the non-locality of the fundamental degrees of freedom is a hint towards emergent gravity is not clear.

³⁶By interaction locality, we mean what is more usually dubbed ‘micro-causality’, i.e. that spacelike separated operator field values do not commute. For a scalar field ϕ , this means that $[\phi(x), \phi(y)] = 0$ provided that x and y are spacelike separated points.

4 Conclusion: From hints to arguments

We have assessed four different hints for the emergent nature of gravity: (1) universal coupling, (2) non-renormalizability, (3) black hole thermodynamics, and (4) the holographic principle. Whether you take any of these hints to be strong indications in favour of an emergent paradigm or not, depends to a good degree on (not fully uncontroversial) background assumptions. If you are committed to the geometric view, universal coupling is just a consequence of the fact that gravity relates to the geometry of spacetime; whereas if you favour a physical view of the metric field, universal coupling *is* suggestive of emergent gravity. Likewise, whether quantum GR is renormalizable or not, or whether a proper quantum gravity approach like LQG is able to cope with the holographic principle are open issues. String theorists and quantum general relativity proponents are likely to take these issues to be settled, although their answers will differ. Whether black hole thermodynamics on its own is a hint, depends on whether black hole entropy can be explained through quantized GR, or only by a type II microtheory such as string theory. And although Jacobson's reinterpretation of GR in terms of horizon thermodynamics provides a clear case of how gravity can be due to underlying type II microstructure, this reinterpretation might be contested. Having said this, we take it to have established why under acceptable background assumptions the hints make a case for emergent gravity.

However, we do not want to limit ourselves to presenting a critical account of individual suggestive features for emergent gravity, but rather strive forward by mounting exemplary analogical arguments based on these hints all together. For instance, the features of GR we referred to as possible hints in the previous sections can star in an argument from example:

Argument from example

Models of hydrodynamics have features F_1, \dots, F_n .

Whenever a model of hydrodynamics applies to a domain of the world, a model of a corresponding type II microtheory applies as well. In particular, there is type II microstructure.

Models of thermodynamics have features F_1, \dots, F_n .

Whenever a model of thermodynamics applies to a domain of the world, a model of a corresponding type II microtheory applies as well. In particular, there is type II microstructure.

Models of GR have features F_1, \dots, F_n .

\therefore Whenever a model of GR applies to a domain of the world, a model of a corresponding type II microtheory applies as well. In particular, there is type II microstructure.

$\{F_1, \dots, F_n\}$ is a set of features including (1) thermodynamic aspects and (2) universality. Another form of analogical argument which can be used in favour of emergent gravity is

Argument from likeness

Models of hydrodynamics and GR are similar with respect to features $\{F_1, \dots, F_n\}$.

Models of hydrodynamics and GR are dissimilar with respect to features $\{G_1, \dots, G_m\}$.

The features $\{F_1, \dots, F_n\}$ weigh out the features $\{G_1, \dots, G_m\}$.

Whenever a model of hydrodynamics applies to a domain of the world, a model of a corresponding type II microtheory applies as well. In particular, there is type II microstructure.

—

\therefore The models of hydrodynamics and GR are decisively similar. Hence, whenever a model of GR applies to a domain of the world, a model of a corresponding type II microtheory applies as well. In particular, there is type II microstructure.

where the set $\{G_1, \dots, G_m\}$ could include, for instance, background-independence.³⁷

So, rather than assessing the strength of individual features, one could argue that several features together make a stronger case than they would on their own. To assess the strength of such holistic arguments, one standardly considers several criteria such as: the number of similarities involved; the overall structural or syntactical similarity; common causes or general laws featuring in analogies (strong causal ties/law-like relations in the target domain and the source domain, respectively, are desired); usage of generalizations (as little as possible); and the material analogy (concrete, observable similarities between the source and the target domain are desirable).³⁸ Applying these criteria to the argument of likeness above, the partial lack of material analogy for certain similarities between GR and hydrodynamics will strike some as problematic. Whereas, for instance, local thermodynamic features of hydrodynamic theories have been observed (and are observable in a straightforward sense), the same cannot be said about GR (black hole thermodynamics is still awaiting *direct* empirical tests.) On the other hand, in particular the mere amount of similarities between gravity and hydrodynamics, as well as their very nature – local

³⁷GR is background-independent – see Belot (2011) for an explication of this notion – while (standard) hydrodynamics is not.

³⁸See Bartha (2016).

thermodynamics and universality are all law-like relations, albeit on a formal level – speak in favour of the argument.

In our discussion of hints we left out two putative challenges standardly mentioned in the context of emergent gravity: (1) the Weinberg-Witten theorem, and (2) the difficulty of emergence of diffeomorphism invariant systems from non-diffeomorphism invariant ones. These challenges have been sufficiently dealt with – or rather dismissed – elsewhere in the literature (see Carlip (2014) for a dismissal of the Weinberg-Witten theorem as a challenge). That diffeomorphism invariant systems can emerge out of non-diffeomorphism invariant systems, has been demonstrated both by the spin-2 account and analogue gravity models as a proof of principle.

Concluding, we have taken up the challenge of making explicit arguments in favour of an emergent gravity paradigm which are referred to in the literature or simply roam around in the community. Needless to say, the depiction of the arguments at play cannot give sufficient credit to all possible variants. Let us anticipate another possible objection, namely that it is unfair to appraise arguments which are only meant as mere heuristics or intuitions in the first place. However, we take this point to be ill-founded. That arguments can only be plausibility arguments at the heuristic level, does not mean that they are immune to scrutiny and critical assessment tout court. The philosopher’s of physics job in the process of discovery of quantum gravity – so we believe – should amount to providing exactly this kind of assessments. The current paper is our contribution to this effort.

Acknowledgments

We are grateful to Harvey Brown, Karen Crowther, James Read, Jeroen van Dongen, Sebastian de Haro, Niels Martens, Kian Salimkhani, Erik Verlinde, and Christian Wüthrich for useful comments on the paper and/or general discussions on emergent gravity. We would also like to thank the History and Philosophy of Gravity group at the University of Amsterdam, and the audiences of the EPSA conference in Exeter and the ‘Thinking about Space and Time’ conference in Bern for useful feedback (both held in summer 2017). The work of M.V. was supported by the Spinoza Grant and by the Delta Institute for Theoretical Physics, both financed by the Dutch Science Organisation (NWO). N.L. is thankful for financial support from the Swiss Science Foundations (SNF) (project number 105212_165702).

References

- Ofer Aharony and Tom Banks. Note on the quantum mechanics of M theory. *Journal of High Energy Physics*, 03:016, 1999.
- Ofer Aharony, Steven S. Gubser, Juan M. Maldacena, Hirosi Ooguri, and Yaron Oz. Large N field theories, string theory and gravity. *Phys. Rept.*, 323:183–386, 2000.
- Jonathan Bain. The emergence of spacetime in condensed matter approaches to quantum gravity. *Studies in History and Philosophy of Modern Physics*, 44(3):338–345, 2013a.
- Jonathan Bain. Effective field theories. In B. Batterman, editor, *The Oxford Handbook of Philosophy of Physics*, pages 224–254. Oxford University Press, New York, 2013b.
- Tom Banks, Willy Fischler, Stephen H. Shenker, and Leonard Susskind. M theory as a matrix model: A Conjecture. *Physical Review D*, 55:5112–5128, 1997.
- Carlos Barcelo, Stefano Liberati, and Matt Visser. Analogue gravity. *Living Reviews in Relativity*, 8:12, 2005.
- James M. Bardeen, Brandon Carter, and Stephen W. Hawking. The four laws of black hole mechanics. *Comm. Math. Phys.*, 31(2):161–170, 1973.
- Paul Bartha. Analogy and Analogical Reasoning. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.
- Jacob D. Bekenstein. Black holes and the second law. *Lettere Al Nuovo Cimento (1971–1985)*, 4(15):737–740, 1972.
- Gordon Belot. Background-independence. *General Relativity and Gravitation*, 43(10):2865–2884, 2011.
- David Berenstein. Large N BPS states and emergent quantum gravity. *Journal of High Energy Physics*, 01:125, 2006.
- Eugenio Bianchi and Robert C. Myers. On the architecture of spacetime geometry. *Classical and Quantum Gravity*, 31:214002, 2014.
- Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825, 2002.

- Harvey R. Brown. *Physical relativity: Space-time structure from a dynamical perspective*. Oxford University Press, 2005.
- Harvey R. Brown and James Read. Clarifying possible misconceptions in the foundations of general relativity. *American Journal of Physics*, 84(5):327–334, 2016.
- Clifford P. Burgess. Quantum gravity in everyday life: General relativity as an effective field theory. *Living Reviews in Relativity*, 7:5–56, 2004.
- Jeremy Butterfield. Emergence, reduction and supervenience: a varied landscape. *Foundations of Physics*, 41(6):920–959, 2011.
- Tian Yu Cao. *Conceptual developments of 20th century field theories*. Cambridge University Press, 1998.
- Tian Yu Cao and Silvan S. Schweber. The conceptual foundations and the philosophical aspects of renormalization theory. *Synthese*, 97(1):33–108, 1993.
- Steven Carlip. Challenges for emergent gravity. *Studies in History and Philosophy of Modern Physics*, 46:200–208, 2014.
- Goffredo Chirco, Hal M. Haggard, Aldo Riello, and Carlo Rovelli. Spacetime thermodynamics without hidden degrees of freedom. *Physical Review D*, 90(4):044044, 2014.
- Karen Crowther. Emergent spacetime according to effective field theory: From top-down and bottom-up. *Studies in History and Philosophy of Modern Physics*, 44(3):321–328, 2013.
- Karen Crowther. *Effective Spacetime: Understanding Emergence in Effective Field Theory and Quantum Gravity*. Springer, Heidelberg, 2016.
- Erik Curiel. A Primer on Energy Conditions. In G. Schieman D. Lehmkuhl and E. Scholz, editors, *Towards a Theory of Spacetime Theories*, volume 13 of *Einstein Studies*, pages 43–104. Birkhauser, 2017.
- Sebastian de Haro. Dualities and emergent gravity: Gauge/gravity duality. *Studies in History and Philosophy of Modern Physics*, 2016.
- Dennis Dieks, Jeroen van Dongen, and Sebastian de Haro. Emergence in holographic scenarios for gravity. *Studies in History and Philosophy of Modern Physics*, B52:203–216, 2015.

- Juliusz Doboszewski and Niels S. Linnemann. How not to establish the non-renormalizability of gravity. *Unpublished draft*, 2016.
- John F. Donoghue. The effective field theory treatment of quantum gravity. *AIP Conf. Proc.*, 1483:73–94, 2012.
- John Dougherty and Craig Callender. Black Hole Thermodynamics: More Than an Analogy? *Unpublished draft*, 2016.
- John Earman and John Norton. What Price Spacetime Substantivalism? The Hole Story. *The British Journal for the Philosophy of Science*, 38(4):515–525, 1987.
- Richard P. Feynman, Fernando B. Morinigo, and William G. Wagner. *Feynman lectures on gravitation*, 2003.
- Michael Friedman. *Foundations of space-time theories: Relativistic physics and philosophy of science*. Princeton University Press, 2014.
- Gary W. Gibbons and Stephen W. Hawking. Action Integrals and Partition Functions in Quantum Gravity. *Physical Review D*, 15:2752–2756, 1977.
- Marc H. Goroff and Augusto Sagnotti. The ultraviolet behavior of Einstein gravity. *Nuclear Physics B*, 266(3):709–736, 1986.
- Stephen W. Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.
- Gary T. Horowitz and Joseph Polchinski. Gauge/gravity duality. In D. Oriti, editor, *Approaches to quantum gravity: Toward a New Understanding of Space, Time and Matter*, pages 169–186. Cambridge University Press, 2006.
- Bei Lok Hu. Emergent/quantum gravity: macro/micro structures of spacetime. *J. Phys. Conf. Ser.*, 174:012015, 2009.
- Nick Huggett and Christian Wüthrich. Emergent spacetime and empirical (in) coherence. *Studies in History and Philosophy of Modern Physics*, 44(3):276–285, 2013.
- Ted Jacobson. Thermodynamics of Spacetime: the Einstein Equation of State. *Physical Review Letters*, 75(7):1260, 1995.
- Ted Jacobson. Entanglement Equilibrium and the Einstein Equation. *Physical Review Letters*, 116(20):201101, 2016.

- Ted Jacobson and Renaud Parentani. Horizon entropy. *Foundations of Physics*, 33(2): 323–348, 2003.
- Claus Kiefer. *Quantum Gravity*, volume 136. Oxford University Press, 2 edition, 2007.
- Eleanor Knox. Effective spacetime geometry. *Studies in History and Philosophy of Modern Physics*, 44(3):346–356, 2013.
- Vincent Lam and Christian Wüthrich. Spacetime is as spacetime does. *Unpublished Draft*, 2017.
- Juan M. Maldacena. The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.*, 38:1113–1133, 1999. [Adv. Theor. Math. Phys. 2, 231 (1998)].
- James Mattingly. Unprincipled microgravity. *Studies in History and Philosophy of Modern Physics*, 46:179–185, 2014.
- Tim Maudlin. *Philosophy of physics: Space and time*. Princeton University Press, 2012.
- Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation*. Macmillan, 1973.
- Thomas Nickles. Two concepts of intertheoretic reduction. *The Journal of Philosophy*, 70 (7):181–201, 1973.
- Thanu Padmanabhan. Emergent gravity and Dark Energy. *arXiv:0802.1798 [gr-qc]*, 2008.
- Giorgio Parisi. The theory of non-renormalizable interactions: The large N expansion. *Nuclear Physics B*, 100(2):368–388, 1975.
- Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- Carina Prunkl and Christopher Timpson. Black Hole Entropy is Entropy and not (necessarily) Information. *Unpublished draft*, 2017.
- James Read, Harvey R. Brown, and Dennis Lehmkuhl. Two miracles of general relativity. *Unpublished draft*, 2017.
- Dean Rickles. AdS/CFT duality and the emergence of spacetime. *Studies in the History and Philosophy of Modern Physics*, B44:312–320, 2013.
- Carlo Rovelli. *Quantum gravity*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2004.

- Andrei D. Sakharov. Vacuum quantum fluctuations in curved space and the theory of gravitation. *Sov. Phys. Dokl.*, 12:1040–1041, 1968. [Gen. Rel. Grav. 32, 365 (2000)].
- Nathan Seiberg. Emergent spacetime. In *The Quantum Structure of Space and Time: Proceedings of the 23rd Solvay Conference on Physics. Brussels, Belgium. 1 - 3 December 2005*, pages 163–178, 2006.
- Assaf Shomer. A Pedagogical explanation for the non-renormalizability of gravity. *arXiv:0709.3555 [hep-th]*, 2008.
- Lorenzo Sindoni. Emergent models for gravity: An overview of microscopic models. *Symmetry, Integrability and Geometry: Methods and Applications*, 8, 2012.
- Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36:6377–6396, 1995.
- Gerard 't Hooft. Dimensional reduction in quantum gravity. In *Salamfest 1993:0284-296*, pages 0284–296, 1993.
- Nicholas J. Teh. Holography and Emergence. *Studies in History and Philosophy of Modern Physics*, 44(3):300–311, 2013.
- Kip S. Thorne, David L. Lee, and Alan P. Lightman. Foundations for a theory of gravitation theories. *Physical Review D*, 7(12):3563, 1973.
- William G. Unruh. Sonic analogue of black holes and the effects of high frequencies on black hole evaporation. *Physical Review D*, 51:2827–2838, Mar 1995.
- Erik P. Verlinde. On the origin of gravity and the laws of Newton. *Journal of High Energy Physics*, 2011(4):1–27, 2011.
- Erik P. Verlinde. Emergent gravity and the dark universe. *SciPost Phys.*, 2(3):016, 2017.
- Matt Visser. Sakharov’s induced gravity: A Modern perspective. *Mod. Phys. Lett.*, A17: 977–992, 2002.
- Robert M. Wald. The thermodynamics of black holes. *Living Reviews in Relativity*, 4:6, 2001.
- David Wallace. The case for black hole thermodynamics, part ii: statistical mechanics. *arXiv:1710.02725 [gr-qc]*, 2017a.

- David Wallace. The case for black hole thermodynamics, part i: phenomenological thermodynamics. *arXiv:1710.02724 [gr-qc]*, 2017b.
- Steven Weinberg. Ultraviolet divergencies in quantum theories of gravitation. In S. Hawking and W. Israel, editors, *General relativity, an Einstein Centenary survey*, pages 790–831. Cambridge University Press, 1979.
- Clifford M. Will. The confrontation between general relativity and experiment. *Living Reviews in Relativity*, 9(1):3, 2006.
- Edward Witten. Reflections on the fate of space-time. *Physics Today*, 49N4:24–30, 1996.
- Christian Wüthrich. Are black holes about information? Forthcoming in R. Dawid, R. Dardashti, and K. Thébault, editors, *Epistemology of Fundamental Physics*, Cambridge University Press, 2017.
- Anthony Zee. *Quantum field theory in a nutshell*. Princeton University Press, 2010.