Dualities and Emergent Gravity: AdS/CFT and Verlinde's Scheme

Sebastian de Haro

Amsterdam University College
University of Amsterdam
Science Park 113, 1090 GD Amsterdam, Netherlands
`s.deharo@uva.nl`
and
Department of History and Philosophy of Science
University of Cambridge
Free School Lane, Cambridge CB2 3RH, United Kingdom
`sd696@cam.ac.uk`

Abstract

In this paper I analyse two closely related examples of duality and of emergence of gravity, namely AdS/CFT and Verlinde's scheme. Based on the notion of duality introduced in Dieks et al. (2014), I here elaborate on the conditions necessary for AdS/CFT to be a duality, in particular the condition of completeness. I also address what is usually seen as a desideratum for any candidate theory of quantum gravity: the background-independence of the theory and the diffeomorphism invariance of the observables. Then I discuss Verlinde's scheme and the extent to which it gives a clear case of emergence of gravity. Finally, I give a novel derivation of the Bekenstein-Hawking black hole entropy formula based on Verlinde's scheme.

Contents

1 Introduction

1.1 Two views on the emergence of gravity

Recent developments in string theory are deeply transforming the way we think about gravity. In the traditional unification programme, gravity was a force meant to be treated on a par with the other forces: the aim was for a unified description of the four forces, and strings seemed to be of help because different vibration modes of the string give rise to different particles. With the advent of holographic ideas[1], however, a slightly different view is emerging that seems to be both more concrete and more modest in its approach. We have come to realise that gravity may be special after all, admitting a holographic reformulation that is generally not available in the absence of gravity. Hence the general goal of understanding gravity at high energies is now best conceptualised as consisting of two steps: 1) reformulate gravity (holographically) in terms of other forces; 2) use this reformulation to understand how gravity is quantised. Progress on the first step over the past seventeen years has been impressive; whether the second step is actually needed is still a matter of hot debate.

It is the latter question, on the necessity of the second step, that provides the topic of this paper. Broadly speaking, when there is a holographic duality—when gravity is

---

[1] See 't Hooft (1993) and Maldacena (1997).

dual to a quantum field theory—it makes sense to ask how to 'reconstruct' the bulk[2], including the quantum manifestations of the gravitational force. On this question, there are two main views: first, if the bulk and the boundary are exactly dual, then the duality map can be used to find out what string theory or quantum gravity look like in the bulk. Gravity may be emergent or not, but the presence of duality up to arbitrarily high energy scales guarantees that it makes sense to speak of 'quantum gravity'.

The contrasting view claims that gravity 'emerges' from the boundary, without being exactly dual to it. In this case, gravity and the bulk are the product of an *approximate* reconstruction procedure. On this view, gravity does not exist at the microscopic level, but exists only by grace of its emergence from the microscopic degrees of freedom. Hence the guiding idea here is not duality but thermodynamics. So on this view step 2) above is superfluous; indeed, nonsensical: all there is to gravity is its reformulation in terms of some deeper (non-gravitational) theory, but asking how to quantise gravity is as futile (or as useful) as asking how to quantise water waves. Of course, the claim does come with a prediction: namely, that fundamental gravitons (or fundamental closed strings, for that matter) will never be found.

It is not the purpose of this paper to assess the relative merits of these contrasting viewpoints on the quantum gravity programme[3]. Instead, the aim is to analyse them separately, thereby contrasting the different roles and conceptualisations that these approaches assign to 'duality' and 'emergence' of gravity and space-time. Regarding duality, I will concentrate on the AdS/CFT correspondence (section 2). Some salient philosophical consequences of duality for AdS/CFT were already explicated in a previous paper, herein after referred to as Dieks et al. (2014). In that paper it was also pointed out how borrowing ideas from the 'emergence' camp, gravity can be seen to emerge in AdS/CFT as well; this requires an additional ingredient, which was identified as a thermodynamic limit. In the current paper, I will analyse in some detail the conditions for the existence of duality in AdS/CFT: in particular the completeness of the observables (section 2.1), and discuss one of the requirements that are usually imposed on theories of quantum gravity: background-independence (sections 2.3 and 2.4).

Regarding the second question, that of emergence without exact duality, I will concentrate on Verlinde's scheme, already also discussed extensively in section 4 of Dieks et al. (2014). Here I will add some comments on the role of thermodynamics (section 3.1), provide a new derivation (section 3.2) of the Bekenstein-Hawking black hole entropy formula based on Verlinde's scheme, and summarise the extent to which this is a clear case of emergence (section 3.3).

Before these two main jobs, I turn in the rest of this introduction to the philosophical motivation (sections 1.2 and 1.3) and a summary of basic facts about AdS/CFT in subsection 1.4.

---

[2] de Haro et al. (2001).

[3] The reason I will not do this is because a final answer to this question depends on the answer to physical questions that are not yet settled.

1.2 Should philosophers care?

Why should philosophers care about these specialised branches of theoretical physics in which the dust has not yet settled on several central theoretical questions? I think there are two main reasons.

First: my comments above clearly reflect different attitudes that physicists take in their approaches to the problem of quantum gravity: there are those who believe that gravity needs to be quantised, and those who believe gravity is an effective description of what in essence is a gravitation-free microscopic world. In the absence of any experiments, these assumptions are necessarily theoretical and ontological; they involve a stance about whether gravity exists at the microscopic level or whether gravity only emerges as a macroscopic phenomenon. And philosophers can help analyse those presuppositions and make sense of the different notions of 'emergence', 'duality', etc., that physicists are using. So philosophy can be helpful as a tool for physicists to sharpen their arguments and to develop heuristic schemes from which to build new theories, and occasionally philosophy can also help identify some misguided assumptions in the heuristics, as we will see in the next subsection.

The second reason I believe philosophers should care about holography is that it is here to stay. Whatever the answers to the main questions about AdS/CFT—whether the duality breaks down by some non-perturbative effect or not, or whatever answers about the ultimate nature of the objects that populate the bulk we may one day get—the result that the two sides of the duality are related by a map such as the one presented in [`AdSCFT`] (subsection 1.4 below) is robust, i.e. well established theoretically. And this result also significantly impacts concepts, such as 'duality' and 'emergence' of gravity, that philosophers of physics ought to be interested in. So it is worth taking examples such as AdS/CFT and Verlinde's scheme as case studies where these philosophical concepts can be analysed and brought to bear on the interpretations of the physics. The insight[4] that the duality of gravity does not by itself entail its emergence, and the analysis of the conditions under which emergence can take place, should by itself be indeed interesting. I believe there are other such notions that deserve philosophical scrutiny, such as background-independence.

The paper is aimed at philosophers of physics, but it is not meant as a self-contained introduction to AdS/CFT. For an introduction to this topic as well as other background on which this paper is based, I refer the reader to Dieks et al. (2014); other interesting philosophical work on AdS/CFT is Rickles (2011). There are also many technical reviews on AdS/CFT; see for instance Aharony et al. (2000).

1.3 Fallacies and heuristics

I remarked above that the importance of holography for philosophy is a two-way street: philosophers can expect their own gardens to flourish from a study of dualities and emergence of gravity. But also for physicists there is something to be gained from a critical scrutiny of the concepts they use.

---

[4] Dieks et al. (2014), sections 3.2.2 and 4.3.

I give an example here: consider the following argument, which I take from Dieks et al. (2014) but which I here write in a more abstract form:

(a) theory A and theory B are dual to one another;
(b) an exact formulation for theory A is known, but such a formulation for B is not known; therefore:
(c) theory A is more fundamental than theory B.

This argument is fallacious because, starting from the premise of duality, it appeals to the pragmatic situation that an exact formulation of B is not known, to conclude that theory B is less fundamental. But this contradicts the premise that theories A and B (rather than their currently known formulations) were dual to one another. After all, an exact duality implies a one-to-one relationship between the values of all possible physical quantities, so that it seems wrong to claim that one of the two theories is more fundamental, basic, or theoretically superior. Since (b) does not imply that a fundamental formulation of B does not exist, the inference from (b) to (c) is invalid. For all we know, a fundamental formulation of theory B might be found tomorrow!

This highlights the usefulness that philosophical analysis can have for the discussion of dualities. Physics papers do not often explicitly write out their premises or include detailed discussions of terms such as 'fundamentality' or 'duality'. Philosophers can contribute by this kind of analysis. For such fallacies make for bad heuristics for physics: if theory A is more fundamental than theory B then there is no point in looking for a fundamental description of B. And so, reliance on a mistaken conclusion might result in missing interesting physics!

1.4 Notation in the AdS/CFT Dictionary

For background details on AdS/CFT and further details on my notation, I refer the reader to section 3.1 of Dieks et al. (2014) so as to keep the paper within reasonable length limits. To fix ideas, however, and for future reference, I will now summarize some facts about AdS/CFT:

- The AdS/CFT correspondence is a duality (one-to-one mapping between states and observables of two theories, see section 2) between string theory in AdS and a conformal field theory on the conformal boundary of this space. The meaning of 'duality' is further explained in Dieks et al. (2014) (sections 3.1 and 3.2 in that paper, respectively).

- Fields in the bulk are generically denoted $\phi(r,x)$, where r is the radial coordinate and x are coordinates along the boundary directions. The dimension of the boundary is d; the bulk has dimensionality $d+1$.

- The conformal boundary is at $r=0$. A boundary condition for the bulk field ϕ is of the type: $r^{\Delta-d}\,\phi(x,r)|_{r=0}=\phi_{(0)}(x)$, where $\phi_{(0)}(x)$ is fixed.

- In the boundary theory, $\phi_{(0)}(x)$ is interpreted as a source that couples to an operator $\mathcal{O}(x)$. The scaling dimension of this operator is denoted Δ.

- The radial coordinate r is dual to the renormalization group (RG) scale (see section 3.1) in the boundary CFT.

- Generically, neither is the bulk pure AdS nor is the boundary theory exactly conformal. It is sufficient that the bulk be asymptotically locally AdS and that the QFT on the boundary have a fixed point. However, I will continue to use the common phrase 'AdS/CFT correspondence'.

Given the above, the AdS/CFT correspondence now relates the string theory partition function in the bulk, with a certain choice of boundary conditions for the field ϕ, to the generating functional of connected correlation functions for the corresponding operator $\mathcal{O}(x)$ in the CFT:

$$Z_{\text{string}}\left(r^{\Delta-d}\phi(x,r)\Big|_{r=0}=\phi_{(0)}(x)\right)=\left\langle e^{\int \mathrm{d}^dx\,\phi_{(0)}(x)\,\mathcal{O}(x)}\right\rangle_{\text{CFT}}\,. \qquad \texttt{[AdSCFT]}$$

In section 2 I will be concerned with the question of how precise and general this mapping is, as well as with its interpretation.

2 Conditions for AdS/CFT Duality

A duality is a one-to-one mapping of observables and states preserving certain structures (Dieks et al. 2014; see also Butterfield (2014)). Let us now unpack the notion so as to see what reasons could lead to AdS/CFT failing to be a duality. Two conditions must be met in order for this bijection to exist. The (observable sub-) structures of the two theories should be:

i. *Complete* (sub-) structures of observables, i.e. no observables can be written down other than those in `[AdSCFT]`: this sub-structure of observables contains what the two theories on each side of the duality regard as all the 'physical observables'.

ii. Identical, i.e. the (sub-) structures of observables on either side are identical to each other. In other words, the duality is *exact*.

If condition (ii) is not met, we have a weaker form of the correspondence: a relation that is non-exact (this was discussed extensively in Dieks et al. (2014)): only an *approximate* duality[5]. This would be the case if for instance the observables agree with each other only in some particular regime of the coupling constants. But duality could also fail for another important reason—the failure of the relation `[AdSCFT]` to capture *all* the observables in one theory (or in both of them), i.e. a failure of (i). I will address (i) in the the next subsection. In sections 2.2 and 2.3 I will discuss a specific condition that (i) imposes on the bulk theory if it is to qualify as a theory of gravity, namely whether the bulk observables satisfy usual requirements of diffeomorphism invariance. Finally, in 2.4, assuming that (i)-(ii) are met, I will discuss two possible interpretation of the duality.

---

[5] See Aharony et al. (2000) p. 60 for a discussion of weak forms of AdS/CFT.

2.1 Completeness of observables

In this subsection I discuss the first desideratum for a bijection—completeness, independently on each side. On the CFT side condition (i) is rather uncontroversial: even though the notion of an S-matrix is not well defined in a CFT, a set of observables can be defined—namely, as the set of all correlation functions of renormalizable, local Hermitian operators that preserve the symmetries of the theory. This entire set of correlation functions can be obtained by functional differentiation of [`AdSCFT`].

Some field theorists have argued that the lack of an S-matrix is problematic from a physical perspective about what good observables are supposed to be like[6]. The fact that particles in a CFT cannot be properly isolated seems to prevent our applying the notion of a 'free particle', which is crucial in formulating both quantum mechanics and quantum field theories, as well as constraining the kinds of experimental set-ups that can be achieved: if the world were described by a CFT, the experimenter would be unable to create the conditions of isolation and control of the system that are critical in experiments such as the ones carried out at CERN. These conditions are also instrumental in ensuring reproducibility of measurements.

Scale invariance thus makes conformal field theories very different from ordinary field theories: the lack of a length scale implies that the concept of 'asymptotic states' (i.e. roughly: quantum states of particles that are far apart from each other and hence are non-interacting) cannot be defined, since points separated by arbitrarily large distances can always be brought arbitrarily close to each other by a coordinate transformation, hence can never be separated in a CFT. Thus a particle excitation can never be fully isolated. Asymptotic states are crucial for defining an S-matrix in ordinary quantum field theory; CFT's therefore do not possess an S-matrix and thus the relevant observables must be taken to be the expectation values of operators.

This objection, however, seems to relate to the *pragmatics* of describing our actual world, rather than to a lack of conceptual consistency in the structure of conformal field theories. Important as such intuitions may be as heuristic guiding principles, criticism based on them does not seem to address the AdS/CFT proposal, which was not advanced as an actual description the world—where, after all, the cosmological constant seems to be positive rather than negative—but as a model for structurally relating quantum gravity theories and quantum field theories[7]. The question that is relevant for an assessment of AdS/CFT by its own lights is therefore not whether the observables satisfy standards of current scientific practice, but whether they could give a candidate physical representation of a world with a *negative* cosmological constant (viz. asymptotic conformal invariance) or whether some observables are fundamentally missing.

Next I turn to the bulk observables to see whether they satisfy (i), i.e. whether they are well defined and form a complete set. There have been suggestions in the literature that completeness of the bulk description requires some non-local observables to be defined

---

[6] Such criticisms were made by Gerard 't Hooft at the Spinoza Meeting on Black Holes (discussion with Juan Maldacena, 1999): 't Hooft (1999).

[7] An important criticism advanced by 't Hooft is that string theory (including some of the CFT's involved) has not been given a formulation of a mathematical rigour comparable to that of the standard model, where at least lattice approaches are available ('t Hooft (2013)).

as well. Specifically, [`AdSCFT`] only envisions observables defined near the boundary, and proposals such as Heemskerk and Polchinski (2011) suggest that observables need to be defined in the interior of AdS as well. In that case equation [`AdSCFT`] would be incomplete. Whether condition (i) would still be met, if such proposals are correct and necessary, would thus depend on whether there is an equation that generalizes [`AdSCFT`] and relates the new observables on both sides of the duality. The proposal Heemskerk and Polchinski (2011) is certainly in this spirit: it is assumed that [`AdSCFT`] can be *extended* to include such operators, hence preserving (i).

2.2 Background-independence of the theory

A constraint that is usually imposed on the formulation of theories of quantum gravity is that they ought to be ‘background-independent’[8]. The main reason for this seems to be that in Einstein’s theory of general relativity, the metric is a dynamical quantity that is determined by the equations of motion rather than set to a fixed value from the outset, so that in particular there is no preferred class of reference frames[9]. A minimalist approach to background-independence is the requirement that a theory is background-independent if it is generally covariant and in its formulation does not make reference to a fixed metric or other fixed fields. So in particular the metric is determined dynamically from the equations of motion. This minimalist notion is actually sufficient for the current discussion because, as we will see, a stronger sense would not be compatible with Einstein’s equations with a negative cosmological constant.

In the minimalist sense, classical gravity in AdS is fully background-independent: its equations of motion are Einstein’s equations with a negative cosmological constant, which are generally covariant and have no dependence on a background metric whatsoever. In fact, they can be derived from a generally covariant action, invariant under diffeomorphisms that leave the asymptotic form of the line element fixed. In the full quantum version of AdS/CFT, quantum corrections manifest themselves as higher-order corrections in powers of the curvature to Einstein’s action. These are generally covariant as well. Hence, in this minimalist sense, also a quantum version of AdS/CFT—which would include quantum corrections as an infinite series with increasing powers of the curvature—is generally covariant.

Can the background-independence of the theory in the minimalist sense be changed by the study of particular solutions of Einstein’s equations? The equations of motion do not determine the boundary conditions, which need to be specified additionally[10]. Notice that this is not a restriction on the class of solutions considered; the equations of motion simply do not contain the information about the boundary conditions, and the latter have

---

[8] For a seminal discussion of background-independence, see e.g. Anderson (1964) and (1967).

[9] The concept of ‘background-independence’ is not a very precise one with a fixed meaning. For a discussion of this, see Belot (2011). According to Belot, the concept is relative to an interpretation of the theory. Between a background dependent theory and a completely background-independent theory there are a range of possibilities: background-independence comes in degrees. See also Giulini (2007).

[10] This elementary point has subtleties regarding the relation between the bulk and boundary metrics in the application to AdS/CFT, which are discussed as an essential ingredient of the ‘dictionary’ between the bulk and the boundary in de Haro et al. (2001).

to be supplied in addition to them. As in classical mechanics, two initial conditions—the asymptotic values of the metric and of its conjugate momentum—are required in order to fully solve the equations of motion. Rather than viewing this as an instance of explicit breaking of background-independence, it seems we should regard this as a case of *spontaneous breaking* of the symmetry: the symmetry is only broken by a choice of a particular solution because this entails a choice of asymptotic background metric. Further quantities one may wish to compute, such as the observables [`AdSCFT`], will carry some dependence on this choice of solution. These observables, however, do exist (and match, in the semi-classical approximation between both sides of [`AdS/CFT`]) for any background—even when the CFT is on a background that is not even locally Minkowski (de Haro et al. 2001). The observables mentioned in [`AdSCFT`] are scalars. In section 2.3 I will discuss the tensor quantities defined by functional differentiation of those in [`AdSCFT`].

I have thus argued that the dependence on the boundary metric introduced by the need to impose boundary conditions should not be seen as a failure of background-independence in this minimalist sense, but as a case of spontaneous symmetry breaking, in the innocuous sense of 'choosing a particular solution which happens to lock a certain symmetry possessed by the equations of motion'. This does not seem to present a problem for AdS/CFT as a theory of quantum gravity, since there is no stronger form of background-independence that is compatible with Einstein's equations in a space with negative cosmological constant.

2.3 Diffeomorphism invariance of the observables [`AdSCFT`] (l.h.s.)

The bulk partition function on the left-hand side of [`AdSCFT`] is diffeomorphism invariant for diffeomorphisms that preserve the asymptotic form of the metric up to a conformal transformation. Restriction to such diffeomorphisms is required by the derivation of the equations of motion from Einstein's action. Two solutions that differ by a 'large' diffeomorphism—one that changes the conformal class of the asymptotic metric—are physically inequivalent solutions and, hence, not related by a symmetry of the theory.

Further observables are derived from [`AdSCFT`] by taking further derivatives with respect to the boundary metric. These observables transform as tensors under the diffeomorphisms discussed in the last pargraph. Here, however, there is a difference between the cases of even vs. odd boundary dimension d:

- For odd d, the observables are covariant if they carry space-time indices, and invariant if they don't. This is precisely what one expects from tensor observables in a generally covariant theory, hence diffeomorphism invariance is preserved by the observables derived from [`AdSCFT`] in the case of odd d.
- For even d, the bulk diffeomorphisms that yield conformal transformations of the boundary metric are broken due to IR divergences. This is called the 'holographic Weyl anomaly', Henningson and Skenderis (1998). In the rest of this subsection, I will give some details (for more, cf. Henningson and Skenderis (1998)). We will see that this holographic Weyl anomaly exactly mirrors the breaking of conformal invariance by quantum effects in the CFT—the well-known 'conformal anomaly'.

For even d, the asymptotic conformal invariance is broken by the regularization of the large-volume divergence. The partition function [`AdSCFT`] now depends on which conformal structure is picked for regularization. As a consequence, observables such as the stress-energy tensor derived from [`AdSCFT`], no longer transform covariantly, but pick up an anomalous term. The conformal anomaly is well-known in two-dimensional CFT's (see e.g. sections 5.4.1-5.4.2 of Di Francesco et al., 1996), where the new term is the Schwarzian derivative. Such a term is also present in four and six dimensions.

The conformal anomaly of the CFT is a quantum effect, proportional to $\hbar$. On the gravitational side, the anomaly is inversely proportional to Newton's constant G and is a classical effect resulting from the large-distance divergence of the action. As with anomalies in field theory, there is no regularization of the boundary action that respects the asymptotic conformal symmetry. Furthermore, the anomaly is robust and is not cancelled by higher-order effects: it is fully non-linear and can be derived without relying on classical approximations.

In string theory the conformal anomaly needs to vanish in order for the theory to be consistent and this is in fact how both the required number of space-time dimensions and Einstein's equations are derived in string theory (for a discussion, see e.g. Vistarini (2015)). The reason the conformal anomaly needs to vanish there is that conformal invariance is part of the local reparametrization invariance of the world-sheet metric. The world-sheet metric is an auxiliary field that has been introduced, but it is integrated out and the theory cannot depend on it. This is not the case in a generic CFT: the metric is a fixed background and the theory is not diffeomorphism invariant. The form of the fields is such that the theory has classical conformal symmetry, but there is no a priori reason why this symmetry should remain at the quantum level.

2.4 Interpreting dualities

Given the symmetry between the two ends of the duality relation—the one-to-one mapping of observables and states allows us to replace each quantity on one side by a quantity on the other—is there a way to decide which theory is more empirically adequate? In Dieks et al. (2014) a distinction was made between two views:

(i) An **external point of view,** in which the meaning of the observables is externally fixed. For instance, in the context of AdS/CFT, the interpretation fixes the meaning of r as the 'radial distance' in the bulk theory, whereas in the boundary theory the meaning of r (or whatever symbol it corresponds to in the notation of the boundary theory) is fixed to be 'renormalization group scale', as reviewed in section 1.4.

More generally, on the external view the interpretative apparatus for the entire theory is fixed on each side. Since the two interpretations are different, the physics they describe are mutually exclusive. Hence only one of the two sides of the duality provides a correct description of empirical reality. On this interpretation there is only a formal/theoretical, but no empirical, equivalence between the two theories, as they clearly use different physical quantities; only one of them can adequately describe the relevant experiments. In short: the 'exact symmetry' between the two theories, expressed by the duality relation, is broken by the different interpretation given to the symbols.

(ii) An **internal point of view:** if the meaning of the symbols is not fixed beforehand, then the two theories, related by the duality, can describe the same observations. Indeed, on this view we would normally say that here we have two formulations of one theory, not two theories. For example: for what might intuitively be interpreted as a 'length', a reinterpretation in terms of 'renormalization group scale' is now available. In other words, one remains undogmatic about the intuitive meaning of the symbols and derives their interpretation from their place and relation with other symbols in the theoretical structure[11].

I believe the latter point of view is correct, assuming (i) we are considering theories of the whole world, and (ii) some form of structural realism is adopted. In this case it is impossible, in fact meaningless, to decide that one formulation of the theory is superior, as both theories are equally successful by all epistemic criteria (for more discussion of this point, see Butterfield (2014)).

The strong equivalence between the theories, according to the internal view, may of course be broken by pragmatic factors. These may make one formulation superior to the other—for instance, if it allows certain computations to be easily done in a particular regime in one theory whereas they are hard to do in the other. This is certainly epistemically relevant and may in fact reveal some objective features of the underlying physical system, in the same way that the applicability of classical thermodynamics reflects some objective features of the system—it is large, it is near equilibrium, etc.

In general, the difference between the external and internal points of view comes down to a distinction between theories regarded as complete descriptions of the world (on the internal view) or only as partial descriptions of empirical reality (on the external view). If a theory only gives a partial description of empirical reality, then there may be other relevant factors, in other words a *context* external to the theory, that fixes the meaning of the symbols. Those relevant factors could be a strong form of metaphysical realism, but might also be more mundane; such as the limited applicability of the theory as a description of the actual world. On the other hand, if the theory describes the entire world, or if the theory gives an idealized description of the world and it can be argued that the elements of the world that the idealization neglects are irrelevant to the ontology of the theory, then—in the absence of strong realist assumptions—the internal point of view seems much more natural. It is also better suited to a 'science first' position on metaphysics.

In order to exhibit the difference between the two positions and to explain why in the case of complete theories (as specified in the last paragraph) I think the internal view is more natural, let me discuss a familiar example. Position-momentum duality in quantum mechanics, despite its differences with the dualities discussed here, may be a useful analogy[12]. With the advent of quantum mechanics one might have insisted that a description of atoms in terms of position or in terms of momentum have different physical interpretations. In particular, the meanings of p and of x are different because the kinds of experimental set-ups that measure position and momentum are different. However,

---

[11] This is reminiscent of the "conceptual role semantics" tradition in the philosophy of language.

[12] For more on this duality as a case study, see Fraser (2015) and Bokulich (2015).

we have gotten used to the idea that the position and momentum descriptions really are different representations of the same theory—with Fourier transformation playing the role of the duality map. Given that all the observables have the same values in both representations, we should simply say that we have not two theories but one[13]. Of course, quantum mechanics is not a description of the whole world: it does not describe gravity. Yet we do not expect that the incorporation of gravity will modify our conclusion about the equivalence of the position and momentum descriptions at the atomic scale. So position-momentum duality in quantum mechanics, on a structuralist interpretation, appears as a clear case of the internal view, despite the fact that quantum mechanics does not give a complete description of the world. In fact, this duality is usually seen as teaching us something new about the nature of reality: namely, that atoms are neither particles, nor waves. By analogy, it is to be expected that dualities in string theory such as AdS/CFT teach us something about the nature of space-time and gravity.

3 Emergent Gravity in Verlinde's Scheme

There have been a proliferation of papers on 'emergent space-time' and 'emergent gravity' in the context of dualities in recent years[14]. The intuitive idea is this: if theory F and theory G are dual to each other, and theory F is some kind of field theory with no gravity whereas theory G is a theory of gravity, then there must be some suitable sense in which gravity can be said to emerge in theory G from theory F, where the latter is regarded as more fundamental.

In section 1 I reviewed a possible obstruction to this statement: if F and G are dual descriptions of the whole world then there does not seem to be a sense in which one theory can be more fundamental than the other. They are different formulations of the same theory describing the same physics, even if their interpretation may be rather different—this is the internal point of view reviewed in section 2.4. In sections 3.2.2-3.2.3 of Dieks et al. (2014) it was further argued that only when gravity appears after some approximation scheme, such as a thermodynamic limit, are we entitled to talk about the emergence of gravity. This approximation scheme could involve either one of two disparate situations:

(i) The duality breaks down at some order in perturbation theory, i.e. there is no duality at the microscopic level. In this case there is certainly room for emergence of gravity from the boundary theory, in a natural and uncontroversial sense of emergence.

(ii) The duality is exact but gravity appears in theory G after some approximation scheme is applied. In this case, since a microscopic theory of G exists, one needs to explain what one means by theory G not containing gravity at the microscopic level while recovering it

---

[13] Bohr's position seems to contradict this because he invokes the context of the measurement as determining which concept—position or momentum—is applicable. On Bohr's view, there is only one possible description for a given measurement context. Bohr's view, however, does not satisfy the requirements of completeness in my previous paragraph, for it invokes a macroscopic context that significantly affects the interpretation.

[14] See e.g. Carlip (2012), Teh (2012).

at the macroscopic level, i.e. a precise definition is needed of what is meant by 'a theory of gravity'.

The analysis in section 4.3 of Dieks et al. (2014) showed that Verlinde's scheme is best described as a case of (i), and therefore does not face this challenge. That is: there is no microscopic theory G and gravity appears as a holographic reformulation of the degrees of freedom of theory F in the thermodynamic limit.

In the next subsection I review Verlinde's derivation of Newton's law based on holography and thermodynamics and in section 3.2 I will give a novel derivation of the Bekenstein-Hawking entropy formula based on Verlinde's scheme.

3.1 Verlinde's scheme

Consider a 'system' of n bits coupled to a reservoir at temperature T and with a total energy $E = \frac{1}{2}NkT$. $N \gg n$ is the number of bits in the reservoir; and knowing the total energy of the reservoir and the number of bits one knows the temperature. Therefore the reservoir is characterized by two numbers, E and N. In this composite system an entropic process takes place by which the 0s and 1s tend to an equilibrium. It is this entropic process that will give rise to gravity.

The quantities that characterize this system, together with its reservoir, are therefore E, N, n: the first two for the reservoir, n for the system itself. To set up a holographic relationship between the composite system and the bulk, we have to imagine the system as being embedded on the 'boundary' of some space, more precisely on a spherical screen of area A enclosing a total mass M, in a flat space outside[15]. The bulk and the reservoir are mapped to each other through the following relations:

$$\begin{aligned} N &= \frac{A\,c^3}{G\hbar} \\ \frac{E}{c^2} &= M\ . \end{aligned} \tag{1}$$

I stress that at this point this is the *definition* of the holographic map between the bulk and the reservoir. The numerical constants are chosen for dimensional reasons, and, as we will see later in this section, their numeric value is irrelevant.

Further, let us imagine that the reservoir is itself a part of the system with so many degrees of freedom that a thermodynamical description is valid. The bits N and n are thus of the same nature, only N is so large that a description in terms of equipartition and temperature is applicable. For instance, we can think of an Ising model starting at some fiducial temperature and ending up at temperature T after a number of block-spin transformations—of averaging over spins and rescaling (see e.g. Chandler (1987)).

In the case at hand, this RG flow will be represented by a trajectory in the space labelled by a parameter x. x labels the level of coarse-graining in the boundary theory and can now be used to distribute the system with its degrees of freedom N radially, as a series of concentric spheres related by coarse graining operations.

More precisely, consider a particle approaching the screen; as it falls in, it disappears from the system and becomes part of the reservoir. The microscopic details of the particle

---

[15] n will correspond to the mass just outside, or just on the screen, as we will see later.

are irrelevant for macroscopic physics—its relevant effects are captured by the increase in the entropy. Throwing particles in amounts to removing particles from the system and adding them to the reservoir, thus removing some information from the system, wich results in a change of thermodynamical state. Removing particles from the system and adding them to the reservoir is like the operation of 'integrating out' in the Ising model and RG.

Now we should recall that in general an RG flow involves two operations: 1) lowering the cutoff together with integrating out degrees of freedom above that cutoff, 2) rescaling of variables so that the new cutoff plays the same role as the old one. This has been discussed in some detail in Dieks et al. (2014). In the context of Verlinde's scheme, I have discussed integrating out, i.e. step 1), in the previous paragraph, where a particle was thrown in. I now turn to rescalings, operation 2). When a particle is added to the reservoir, N increases by one. The area (1), however, can be held fixed if a simultaneous rescaling of $\hbar$ is carried out. The identification between bulk and boundary quantities (1) indeed involves the constant $\hbar$, which is introduced for dimensional reasons. This proportionality constant is part of the *definition* of the holographic map—of how the boundary theory is embedded in the bulk hence it is not a physical quantity but a choice of length scale. Indeed $\hbar$ drops from the final formula for the force.

I now review the calculation that in Verlinde's scheme fleshes out these ideas and leads to Newton's law. Consider Figure 1 below, which is a rehearsal of Bekenstein's argument. When the particle advances by a distance $\Delta x = \frac{\hbar}{mc}$ (the particle's Compton wavelength) from the red to the green screen, it disappears behind the red screen and information is lost from the bulk. The amount of information lost is estimated by Bekenstein to be roughly one bit of information, in conveniently chosen units:

$$\Delta I_{\text{lost}} = -2\pi k_{\text{B}} \ . \tag{2}$$

Using Shannon's connection between loss of information and increase in entropy, the entropy increases by an amount $\Delta S = -\Delta I_{\text{lost}}$:

$$\Delta S = 2\pi k_{\text{B}} = 2\pi k_{\text{B}} \frac{mc}{\hbar} \Delta x \ . \tag{3}$$

From the bulk point of view, x appears here as the position of the particle falling in; from the point of view of the boundary it is a bookkeeping device that records the level of coarse graining and hence the increase in entropy as more information is lost. The force associated with this increase in entropy is given by the second law of thermodynamics: $F = T\,\frac{\Delta S}{\Delta x} = 2\pi k\,\frac{mc}{\hbar}\,T$. Solving for T from (1) and writing the area of the sphere $A = 4\pi R^2$, we get Newton's law:

$$F = G\,\frac{Mm}{R^2} \ . \tag{4}$$

I will say more about the physics of this equation in section 3.2.

3.2 The Bekenstein-Hawking Entropy Formula

In this section I will give a novel derivation of the black hole entropy formula based on Verlinde's arguments. In the discussion so far in section 3, we have been concerned with

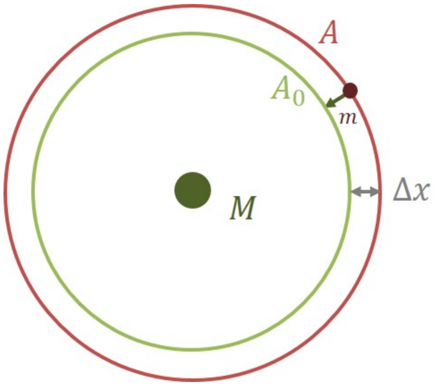

Figure 1: A test mass falls into the screen.

the reservoir and its thermodynamial properties. Let us now describe the system itself in thermodynamical terms. In analogy with (1), we have:

$$m = \frac{1}{2c^2}\, n\, k_{\rm B} T \tag{5}$$

as well as:

$$\mathbf{a} = -\nabla\Phi \ , \tag{6}$$

where $\Phi = -\frac{GM}{r}$ is Newton's potential, the variable that measures the level of coarse graining. For small displacements $\Delta r < 0$ (inward fall) we also have $\Delta\Phi < 0$, hence the acceleration points inward and

$$\frac{\Delta S}{n} = k_{\rm B}\, \frac{\Delta\Phi}{2c^2} > 0 \tag{7}$$

increases as the particle falls in. Thus the direction of increasing coarse graining is indeed the direction of decreasing Newtonian potential, $\Phi < 0$.

Let us consider what happens when we reach the maximum potential, the potential at the horizon of a black hole: $r = R_{\rm S} = \frac{2GM}{c^2}$. In that case:

$$\Phi(R_{\rm S}) = -\frac{c^2}{2} \tag{8}$$

which gives:

$$-\frac{\Phi(R_{\rm S})}{2c^2} = \frac{1}{4}\, k_{\rm B} \ . \tag{9}$$

In particular, consider a particle falling in from infinity $\Phi = 0$ towards the horizon. Its change in entropy is:

$$\Delta S = \frac{1}{4}\, k_{\rm B}\, n \ . \tag{10}$$

Let us now assume that the entire entropy of the black hole was built up by the repetition of such a process: particles coming in from infinity, each of them adding an amount of $\frac{1}{4}\,k_{\rm B}$ per bit. Using (1) for these N particles, we get exactly the Bekenstein-Hawking entropy formula:

$$S=\frac{1}{4}\,k_{\rm B}\,N=\frac{1}{4}\frac{k_{\rm B}A\,c^3}{G\hbar}\ . \tag{11}$$

The novelty here is that we get the black hole entropy with the correct factor of 1/4 even though the normalization (1) did not include such a factor. Furthermore, (1) was a *definition* of the number of states, whereas (11) calculates something physical. The normalization in (3) was chosen so as to give Newton's law, and as we see from the derivation Newton's law 'knows' about the normalisation of the black hole entropy formula. The factor of 1/4 in Bekenstein's formula, then, follows in Verlinde's scheme from the same principles as Newton's law. This consistency check adds to the robustness of Verlinde's derivation of gravity.

3.3 Comparison

A duality relation such as AdS/CFT with emergent gravity can be construed as in Figure 2. On the left-hand side is the boundary theory, on the right-hand side is the bulk. In Dieks et al. (2014) it was observed that the RG flow can be regarded as an approximation suitable for describing certain situations (e.g. low energies, see section 1.4); and so it is akin to taking a thermodynamic limit. Now in such a scheme, the holographic relation relating the left- and right-hand sides may well be a bijective map acting at each level of coarse graining. If so, and regarding these theories as descriptions of the whole world so that one is justified (or takes oneself to be justified!) in adopting the internal view (ii) of section 2.4, there is no reason why one side should be more fundamental than the other. Rather, it is the *thermodynamic limit* that introduces the possibility for gravity to emerge on the right-hand side.

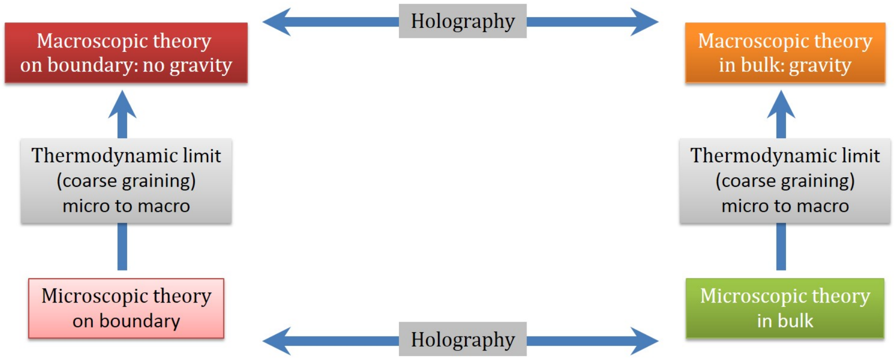

Figure 2: Analysis of holographic duals vs. emergent gravity via suitable approximation scheme.

To sum up: remember that duality alone cannot lead to emergence (Dieks et al. 2014). Figure 2 explicates how gravity can emerge in AdS/CFT, where the RG flow is interpreted as akin to taking the thermodynamic limit.

On the other hand, Verlinde's scheme does not require a microscopic duality, therefore Verlinde has Figure 3 in mind. There is no microscopic theory in the bulk, the only the microscopic theory is on the sphere but the details of the latter are irrelevant; what matters is its universality class. In the thermodynamic limit, this theory flows towards a macroscopic theory on the sphere characterised by E, N, T, n, which is dual to the gravity theory in the bulk. It is here that gravity can emerge from the theory on the sphere in an uncontroversial way. In conclusion, the crucial ingredient in this kind of emergence is not holography, but the existence of a thermodynamic limit on the boundary theory (here, on the surface of the sphere).

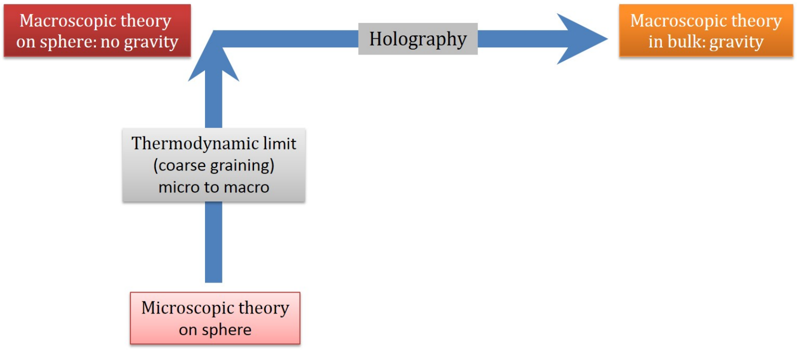

Figure 3: Analysis of Verlinde's scheme.

4 Conclusion

Let me summarize the main points as follows. The concept of duality hinges on the existence of a bijection relating the states and the quantities of the two theories. In AdS/CFT, such a bijection is provided by the fundamental conjecture [`AdSCFT`]. The existence of AdS/CFT as a duality thus depends on the relation [`AdSCFT`] being: (i) complete, (ii) exact. There are no good reasons to believe that (i) fails, but the validity of (ii) is still open (despite a large amount of evidence that, at least in the most symmetric cases, it is satisfied). Failure of (ii) in AdS/CFT, or its generalizations such as dS/CFT and cases with less symmetry, actually opens a new door to the emergence of gravity, as I will review below.

Concerning the important question of background-independence, I gave a minimalist conception of the latter as meaning: general covariance of all the observables and lack of reference to a priori fixed fields in the theory. I argued that AdS/CFT is background-independent on this conception, with the observables falling into two classes: (i) the scalar operator [`AdSCFT`] (the partition function/generating functional) is *invariant* under diffeomorphisms; (ii) tensor operators (obtained by taking functional derivatives of [`AdSCFT`]) are *co-variant* in even bulk dimensions. Thus, in even bulk dimensions, AdS/CFT satisfies the requirements of background-independence. Certainly, more work needs to be done to further refine and generalize the notion of background-independence (see e.g. Belot (2011), Giulini (2007)), but, for the reasons explained, such refinement will not change this important conclusion if the notion keeps Einstein's theory as the paragon of a background-independent theory. In odd bulk dimensions, the story is more subtle because of the presence of the *holographic Weyl anomaly*, which breaks those bulk diffeomorphisms that give rise to non-trivial conformal transformations on the boundary. In particular, the holographic stress-energy tensor is not a tensor but picks up an anomalous term. This anomaly precisely matches the *conformal anomaly* of the CFT for curved backgrounds. It was argued that this anomaly is harmless—it does not imply any kind of inconsistency of the theory (as *is* the case of local anomalies in quantum field theories), but, like global anomalies in quantum field theory, merely expresses the fact that some accidental symmetries of the action are not (need not) be respected upon renormalizing the theory. Thus, for odd bulk dimensions, background-independence is broken by an anomaly that is both unavoidable and harmless—this seems really the best one can do for a theory with negative cosmological constant and with Einstein gravity as its classical limit.

Interpreting dualities, two possibilities were found in Dieks et al. (2014): an external and an internal viewpoint. In this paper I traced the difference between these two viewpoints back to a distinction between theories of physics regarded as: (i) complete descriptions of the world (on the internal viewpoint), or (ii) partial descriptions of empirical reality (on the external viewpoint). On the latter, a *context* must be provided that fixes the interpretation. On the former, the interpretation is internally fixed. The difference between the two viewpoints is *not* simply that existing between structuralist and strong realist metaphysical positions, as other factors other than strong realism can also prompt the external viewpoint. I explained my preference for the internal viewpoint under certain additional conditions.

Finally, regarding the question of emergence, I argued that in the models of space-time considered here, approximations are the key features that can make space-time emerge. In particular, there are two mutually opposed approaches to quantum gravity and I analyzed in which way gravity can emerge in each of them: (i) If the duality breaks down, gravity can emerge from a more fundamental boundary theory. This is the viewpoint in which there is no fundamental theory of quantum gravity, but only a fundamental quantum field theory from which gravity emerges. (ii) If a fundamental theory of quantum gravity does exist, then an approximation scheme (such as the Wilsonian coarse graining approach) can make Einstein gravity to emerge from a more fundamental theory. Physics, not philosophy, is to decide which of the two variants is correct and applicable to our world. The task of philosophical analysis, as viewed here, is to map the possible alternatives and explicate how it is that gravity can emerge. The current analysis of emergence applies independently of the existence of dualities, hence it explicitly shows the consequences of having (or not having) a fundamental theory of quantum gravity for the possibility of gravity to emerge.

As a corollary, I provided a novel derivation of the Bekenstein-Hawking black hole entropy formula in Verlinde's scheme, including the correct numerical factor. This closes a circle of ideas in Verlinde's argument, which started from Bekenstein's thought experiment but left us in the dark as to how to recover the entropy of black holes.

Acknowledgements

It is a pleasure to thank Dennis Dieks and Jeroen van Dongen. Also special thanks to Jeremy Butterfield for joint work on this topic and for careful reading of the manuscript. I would like to thank Elena Castellani and Dean Rickles for organising a wonderful workshop, as well as all of the participants for the helpful discussions.

References

Aharony, O., S.S. Gubser, J.M. Maldacena, H. Ooguri, and Y. Oz. (2000). "Large N field theories, string theory and gravity", *Physics Reports*, 323(3-4), 183-386. [hep-th/9905111].

Anderson, J. (1964). "Relativity principles and the role of coordinates in physics." In: H.Y. Chiu, W. Hoffman (eds.) Gravitation and Relativity, pp. 175194. W.A. Benjamin, Inc., New York.

Anderson, J. (1967). "Principles of Relativity Physics". Academic Press, New York.

Belot, G. (2011). "Background-Independence," Gen. Rel. Grav. **43** (2011) 2865 [arXiv:1106.0920 [gr-qc]].

Bokulich, P. (2015), contribution to this volume.

Butterfield, J.N. (2014), "Duality as a symmetry between theories".

Castellani, E. (2010). "Dualities and intertheoretic relations", pp. 9-19 in: Suarez, M., M. Dorato and M. Redéi (eds.). *EPSA Philosophical Issues in the Sciences.* Dordrecht: Springer.

Chandler, D. (1987), "Introduction to Modern Statistical Mechanics", New York, Oxford: Oxford University Press.

de Haro, S., Skenderis, K., and Solodukhin, S. (2001). "Holographic reconstruction of spacetime and renormalization in the AdS/CFT correspondence", *Communications in Mathematical Physics*, 217(3), 595-622. [hep-th/0002230].

Dieks, D., Dongen, J. van, Haro, S. de, (2014), "Emergence in Holographic Scenarios for Gravity", PhilSci 10606, *Studies in History and Philosophy of Modern Physics,* submitted.

Di Francesco, P. , Mathieu, P. , Sénéchal (1996). "Conformal Field Theory". Springer-Verlag New York.

Fraser, D. (2015), contribution to this volume.

Giulini, D. (2007). “Some remarks on the notions of general covariance and background independence,” Lect. Notes Phys. **721** (2007) 105 [gr-qc/0603087].

Heemskerk, I. and Polchinski, J. (2011). “Holographic and Wilsonian Renormalization Groups,” JHEP **1106** 031 [arXiv:1010.1264 [hep-th]].

Henningson, M. , Skenderis, K. (1998). “The Holographic Weyl anomaly,” JHEP **9807** (1998) 023 [hep-th/9806087].

Huggett, N. (2015), contribution to this volume.

Knox, E. (2015), contribution to this volume.

Maldacena, J. (1998). “The large N limit of superconformal field theories and supergravity”, *Advances in Theoretical and Mathematical Physics* 2, 231-252. [hep-th/9711200].

Polchinski, J. (2015), “Dualities of Fields and Strings,” arXiv:1412.5704 [hep-th]. Contribution to this volume.

Rickles, D. (2011). “A philosopher looks at string dualities”, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 42(1), 54-67.

Rickles, D. (2012). “AdS/CFT duality and the emergence of spacetime”, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3), 312-320.

Teh, N.J. (2013). “Holography and emergence”, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3), 300-311.

’t Hooft, G. (1993). “Dimensional reduction in quantum gravity”, in: Ali, A., J. Ellis and S. Randjbar-Daemi, *Salamfestschrift*. Singapore: World Scientific. [gr-qc/9310026].

’t Hooft, G. (1999). Comments at the Spinoza Meeting on Black Holes (discussion with Juan Maldacena).

’t Hooft, G. (2013). “On the Foundations of Superstring Theory”, *Foundations of Physics*, 43 (1), 46-53

Verlinde, E. (2011). “On the origin of gravity and the laws of Newton”, *Journal of High Energy Physics,* 2011 029. [arXiv:1001.0785 [hep-th]].

Vistarini, T. (2015), contribution to this volume.