Emergence in Holographic Scenarios for Gravity

Dennis Dieks[1], Jeroen van Dongen[1,2] and Sebastian de Haro[2,3]

[1]*Institute for History and Foundations of Science, Utrecht University,*
Utrecht, the Netherlands
[2]*Institute for Theoretical Physics, University of Amsterdam,*
Amsterdam, the Netherlands
[3]*Amsterdam University College, University of Amsterdam,*
Amsterdam, the Netherlands

Abstract

'Holographic' relations between theories have become a main theme in quantum gravity research. These relations entail that a theory without gravity is equivalent to a gravitational theory with an extra spatial dimension. The idea of holography was first proposed in 1993 by Gerard 't Hooft on the basis of his studies of evaporating black holes. Soon afterwards the holographic 'AdS/CFT' duality was introduced, which since has been heavily studied in the string theory community and beyond. Recently, Erik Verlinde has proposed that even Newton's law of gravitation can be related holographically to the thermodynamics of information on screens. We discuss inter-theoretical relations in these scenarios: what is the status of the holographic relation in them and in what sense is gravity, or spacetime, emergent?

1 Introduction

During the last twenty years the concept of holography from quantum gravity research has grown into one of the key successful innovations in theoretical physics. By now it is studied in many diverse subfields and the literature on the subject is enormous. One of the most successful papers on holography, the article that announced the celebrated 'AdS/CFT' correspondence, has been cited more than ten thousand times.[1] Even fields that at first sight would seem far removed from the expected high energies of quantum gravity are now engaging with holography. For example, central issues in condensed matter physics are addressed using holographic ideas.[2]

Philosophers of physics should be aware of this important development: there are several ideas here that not only relate to the foundations of physics but also to more general themes in philosophy of science. In short, the core idea of holography is that a lower dimensional quantum theory without gravitation (for instance, defined on a two-dimensional sphere) is capable of describing physical phenomena, including manifestations

---

[1](Maldacena, 1997); on 10 December 2013, a Google scholar citation search resulted in 11716 hits.
[2]See for example (Heartnoll et al., 2008); (McGreevy, 2010); (Cubrović et al. 2009).

of the gravitational force, that take place in a higher dimensional spacetime (such as the interior of the sphere).[3]

Discussion of the holographic relation clearly must touch on philosophical questions of emergence and reduction.[4] Also in the physics literature these themes have come up, as reflected in some of the titles of articles on the subject: these announce "Emergent spacetime", "Emergent gauge fields" or, e.g., promise a discussion of "Aspects of emergent geometry in the AdS/CFT context."[5] Indeed, one of the publications that we specifically focus on in this article is called "On the origin of gravity and the laws of Newton."[6]

We will discuss a number of holographic scenarios and place them in the context of existing ideas about emergence and reduction. It is not our aim to focus on the general analysis of the concept of emergence itself. Globally speaking, we sympathize with the characterization of emergence as novel and robust behaviour relative to some appropriate comparison class,[7] and we will use the term 'emergence' accordingly. What we wish to investigate here is whether, and if so how, recent holographic scenarios can be interpreted as representing such emergence, and whether one theory in a holographic pair can justifiably be called more fundamental than the other. We will discuss three proposals in particular: Gerard 't Hooft's original formulation of the holographic hypothesis, the AdS/CFT duality from string theory, and Erik Verlinde's recent ideas. Although these proposals are strongly interrelated, we will argue that only Verlinde's account realizes emergence in a straightforward and uncontroversial way: gravity and spacetime here arise as thermodynamic phenomena in a coarse grained description. Verlinde's proposal builds upon the AdS/CFT duality and 't Hooft's ideas. However, it will turn out to be far less clear how the concept of *emergence* of three-dimensional gravity from two-dimensional non-gravitational processes could apply to these two latter cases, even though they are often discussed as scenarios of emergence.

That gravity perhaps emerges from some deeper layer of reality and is different from other forces may intuitively be a rather plausible idea, even if it is an intuition that is alien to the string theory or quantum gravity program. Gravity distinguishes itself because it is universal: it applies to all forms of matter and energy, and relates to the general framework of space and time itself—this may remind one of the universal character of thermodynamic descriptions. Moreover, gravity is notoriously and essentially more difficult to quantize than other forces. This suggests a difference of principle from the ordinary physical forces represented in the standard model. As already mentioned, studies of black hole physics have led to the hypothesis that quantum gravity theories within a volume correspond to theories *without gravitation* on the boundary of this volume. This seems only a small step from the notion that gravity *emerges* from processes described by a theory without gravity; it is this idea that we will critically analyze here.

---

[3]For a systematic statement of the holographic principle that makes reference to appropriate choices of surface and interior, see (Bousso, 2002). For an early but comprehensive overview of AdS/CFT, see (Aharony et al., 2000).

[4]See Rickles (2012) and Teh (2012).

[5]Respectively: (Seiberg, 2006), (Domènech et al., 2010), (Berenstein and Cotta, 2006).

[6](Verlinde, 2011).

[7](Butterfield (2011a, 2011b)).

2 The holographic hypothesis

The central ideas of holography go back to the debates about the black hole information paradox that raged in the early 1990s. Important participants in these discussions were Gerard 't Hooft and Stephen Hawking; the latter famously claimed that black holes destroy information, which was opposed by the former.[8] In 1993, almost twenty years after the first results on the evaporation of black holes had been announced by Hawking, 't Hooft put on the Los Alamos preprint server a short contribution to a future Festschrift honoring particle physicist Abdus Salam. It contained the first formulation of what would soon become known as the *holographic principle* of quantum gravity.[9]

In his article, 't Hooft made a programmatic start with the formulation of a unitary quantum theory of gravity, taking his cue from processes that he hypothesized to take place near black hole horizons. While leaving open what the exact degrees of freedom would be, 't Hooft argued via thermodynamical arguments that the entropy of a black hole system is proportional to its horizon's area A. In natural units, and with the black hole's Schwarzschild radius given by $2M$:

$$S = 4\pi M^2 = A/4. \tag{1}$$

This gives us a handle on *how many* degrees of freedom there are in the black hole system, but it is also suggestive of the *kind* of theory that should be able to describe these fundamental degrees of freedom. 't Hooft concluded that: "The total number of [...] degrees of freedom, n, in a region of space-time surrounding a black hole is:"[10]

$$n = \frac{S}{\log 2} = \frac{A}{4\log 2}. \tag{2}$$

Accordingly, there is a finite number of degrees of freedom in a black hole system.

't Hooft carried the argument one step further by pointing out that if a spherical volume V is bounded by a surface A, the total number of possible states and the entropy inside A are maximized if the volume contains a black hole. Therefore, the number of degrees of freedom contained in any spatial volume is bounded by the size of its boundary surface area, and not by the size of the volume itself. In other words, there are much less degrees of freedom in the volume than one would expect on the basis of traditional calculations. So, "we can represent all that happens inside [the volume] by degrees of freedom on this surface [...]. This suggests that quantum gravity should be described entirely by a topological quantum field theory, in which all physical degrees of freedom can be projected onto the boundary. One Boolean variable per Planckian surface element should suffice."

This observation contains the essence of the *holographic hypothesis*. Again 't Hooft: "We suspect that there simply *are* not more degrees of freedom to talk about than the ones one can draw on a surface [...]. The situation can be compared with a hologram of a three dimensional image on a two dimensional surface."[11] Stretching the surface to infinity,

---

[8](Hawking, 1976), ('t Hooft, 1985).
[9]('t Hooft, 1993).
[10]('t Hooft, 1993), p. 4.
[11]('t Hooft, 1993), p. 6.

the conclusion is that a two-dimensional surface can contain all physical information of three-space.

What does 't Hooft's account imply for the relation between the three-dimensional description and the surface description? The original 1993 text already suggests some possible answers. 't Hooft's 1993 abstract states, interestingly, that at the Planck scale "our world is not 3+1 dimensional." This appears to give precedence to the holographic description: the theory on the surface is more fundamental than the theory in the bulk. However, 't Hooft's paper is not unambiguous on this point: in the same abstract, he says that the observables in our world "can best be described *as if*"[12] they were Boolean variables on a lattice, which suggests that the description on the surface only serves as one possible *representation*. Nevertheless, 't Hooft's account more often assumes that the fundamental ontology is the one of the degrees of freedom that scale with the spacetime's boundary. In fact, 't Hooft argued that quantum gravity theories that are formulated in a four dimensional spacetime, and that one would normally expect to have a number of degrees of freedom that scales with the volume, must be "infinitely correlated" at the Planck scale. The argument is that the real number of degrees of freedom is given by a theory on the surface, and because this number is much smaller that the number of independent degrees of freedom one could fit in the enclosed volume, the volume degrees of freedom cannot be independent. 't Hooft even expressed the hope that this overdetermination might hold the key to an explanation of the notorious EPR correlations.[13] The explanatory arrow here clearly goes from surface to bulk, with the plausible implication that the surface theory should be taken as more basic than the theory of the enclosed volume. One is tempted to express this by saying that the space-time theory of the enclosed volume *emerges* from the description on the surface. On the other hand, the precise correspondence between boundary and bulk degrees of freedom does not immediately suggest the occurrence of new types of behaviour, which speaks against emergence in a more technical sense.

't Hooft proposed no concrete candidate for a theory on the surface. But given the just-mentioned reading of his account, this surface theory—whatever it would be—would be the best pick for a scientific realist who wishes to identify the fundamental objects in the quantum gravity world. So according to this reading there is no ontological 'democracy' between surface and bulk.[14] Yet, there are also elements in 't Hooft's proposal that indicate a more equal status for the bulk and boundary theories. Firstly, 't Hooft attributes the bulk theory a primary role when he points out that it is its black holes that are responsible for the "most direct and obvious *physical* cut-off"[15] of the degrees of freedom, which explains the finiteness of the number of degrees of freedom. Secondly, in the debate on the information paradox 't Hooft proposed that operators of observers moving inwards in the black hole spacetime (in the bulk), and operators of observers that

---

[12]('t Hooft, 1993), p.1, our emphasis.

[13]See e.g. also ('t Hooft, 1999).

[14]The notion of ontological 'democracy' in this context was proposed by Elena Castellani, lecture at Seven Pines Symposium XVI, "Analogy and Duality in Physics", 16-20 May 2012, Stillwater, MN, USA; compare also Rickles (2011).

[15]('t Hooft, 1993), p. 2, emphasis as in original.

remain at a distance, on the boundary, do not commute.[16] This appears to point in the direction of a kind of complementarity between the two observers' descriptions of the quantum black hole state.[17] 'Complementarity' seems to imply that the two perspectives can claim equal rights in describing the physics of the black hole. So, 't Hooft's holographic proposal wavers between boundary and bulk as fundamental ontologies. There is an interpretative tension here, that will resurface later in this article.

't Hooft's paper was programmatic and did not elaborate much on concrete possibilities for the bulk and boundary theories and their precise mutual relation. But the massive amount of later work on the so-called 'AdS/CFT' duality has changed the situation. Here we have a concrete example of a holographic relation between two theories as suggested by 't Hooft (and subsequently by others, in particular by Leonard Susskind in an article that attracted considerable attention in the string theory community[18]). We will discuss this concrete holographic proposal in the following section. Let us end here by noting that soon after 't Hooft's paper, holography took on the role of a guiding principle in much quantum gravity work, not just in efforts based in string or field theory.[19]

3   The AdS/CFT duality and its interpretation

We will first explain the AdS/CFT correspondence and then briefly discuss its interpretation, in particular with respect to issues of emergence and fundamentality. We also introduce the renormalization group, which is an important ingredient in AdS/CFT and also in Verlinde's scenario (to be discussed in section 4).

3.1   What is AdS/CFT?

The idea of a holographic correspondence between bulk theories and theories defined on the boundaries of their spacetimes has found an explicit illustration in gravitational string theories in Anti-de Sitter (AdS) spacetime. These string theories are believed to correspond exactly to Conformal Field Theories (CFT), without gravity, on the *boundary* of AdS. This 'AdS/CFT duality' was first conjectured by Juan Maldacena in 1997.[20]

The AdS/CFT duality relates string theory in (asymptotically) $d + 1$-dimensional Anti-de Sitter spacetime (AdS_{d+1}) to a conformal field theory on a d-dimensional space isomorphic to the boundary of AdS. In this way it is a concrete realization of holographic ideas, even if the term 'holography' was absent in Maldacena's original paper and initial excitement focused on the duality symmetry itself rather than its holographic aspects.[21]

---

[16]See ('t Hooft, 1996), pp. 30-34; 65-66.

[17]The term 'black hole complementarity' appears to have been introduced by (Susskind et al., 1993); for philosophical discussion, see (Belot et al., 1999), (van Dongen and de Haro, 2004).

[18](Susskind, 1995).

[19]See e.g. its discussion in the book by Lee Smolin (2007, pp. 317-319), that is quite critical of string theory, and advocates other approaches to quantum theories of gravity.

[20](Maldacena, 1997); an important early review article of the subject is (Aharony et al., 1999).

[21]The holographic nature of AdS/CFT was particularly highlighted in influential articles by Leonard Susskind and Edward Witten (Witten, 1998a; Susskind and Witten, 1998).

Let us first look at some of the concepts that underlie the correspondence. Anti-de Sitter spacetime is the maximally symmetric solution of the Einstein equations with a negative cosmological constant. In a suitably chosen local coordinate patch, its metric has the form:

$$ds^2 = \frac{\ell^2}{r^2} \left(dr^2 - dt^2 + d\vec{x}^2 \right), \tag{3}$$

where $\vec{x}$ parametrizes $d-1$ spatial coordinates. This 'bulk' metric has a horizon at $r = \infty$ and spacetime can be extended beyond the horizon by a coordinate transformation. In these coordinates, $r = 0$ is the boundary of the space-time, which has been represented at a finite coordinate distance. The singularity at $r = 0$ is therefore a large-distance singularity of the type one expects for spaces with infinite volumes that are finitely parametrized. Up to a conformal factor, the bulk metric induces a flat d-dimensional Minkowski metric on the boundary at $r = 0$, which is given by $-dt^2 + d\vec{x}^2$. This is the metric of the fixed spacetime background of the conformal field theory.

A conformal field theory is a quantum field theory that is invariant under conformal transformations, that is, angle-preserving coordinate transformations under which the metric is multiplied by a scalar function (a 'conformal factor'). The just-mentioned fact that the bulk metric does not induce a Minkowski metric on the boundary of AdS, but only a class of metrics conformally equivalent to it, is thus insignificant: because the boundary theory is a conformal field theory, it is insensitive to the conformal factor of the metric.[22] In the standard example of AdS_5/CFT_4, in which $d = 4$, the conformal field theory on the boundary is supersymmetric Yang-Mills theory in 3+1 dimensions. The dual theory in this case is a type IIB string theory in $AdS_5 \times S^5$ (S^5 is an additional internal manifold), which in the classical limit reduces to supergravity in AdS_5.

The idea of a duality came to Maldacena when he was struck by the fact that there appeared to be two equivalent ways of describing, in the low-energy limit, the states of a stack of N D3-branes in the type IIB string theory (a 'D-brane' is a generalized type of particle solution, spatially extended in D dimensions). On the one hand, one can use the field theory living on the world volume mapped out by the D-branes. The branes are stacked close together, and their excitations can be described by a Yang-Mills theory with gauge group $SU(N)$ at low values of the coupling constant, in which the number of branes determines the rank N of the gauge group. On the other hand, excitations in the bulk geometry surrounding the D-branes can be described by using IIB supergravity in an $AdS_5 \times S^5$ spacetime, evaluated in the regime of strong coupling. Maldacena took this correspondence between two equivalent descriptions as a hint that there existed a general exact relation between gauge and bulk theories at all values of the coupling.[23]

What this correspondence could look like was further investigated by, particularly, Steven Gubser, Igor Klebanov, Alexander Polyakov and Edward Witten.[24] These authors proposed an exact equality between the partition function of the CFT (deformed by the insertion of an operator coupled to an external source) and the partition function Z_{string} of the quantum gravity theory in the AdS bulk. The partition function fully determines the expectation values of observables; so the claim that such an equality exists is a hypothesis

---

[22] Anomalies however may ruin this in quantum regime, see e.g. (Henningson and Skenderis, 1998).

[23] (Maldacena, 1997).

[24] (Gubser et al., 1998), (Witten, 1998a).

with strong implications. The precise form of the correspondence is given by:

$$\left\langle e^{\int \mathrm{d}^d x\, \phi_{(0)}(x)\, \mathcal{O}(x)} \right\rangle_{\mathrm{CFT}} = Z_{\mathrm{string}} \left(r^{\Delta - d} \phi(x, r) \Big|_{r=0} = \phi_{(0)}(x) \right) . \tag{4}$$

On the left hand side, $\mathcal{O}$ is an operator inserted via a space-dependent coupling parameter $\phi_{(0)}(x)$. This coupling is not a quantum field, but can be thought of as representing a classical external source that probes the system. On the right hand side, the string partition function of the scalar field ϕ is computed with a prescribed boundary condition at $r = 0$, given by $\phi_{(0)}(x)$; Δ is a constant that depends on the dimension of the bulk spacetime and the field's mass.

The essential message of Eq. (4) is that there is a one-to-one correspondence between observables of the bulk theory (represented by fields) and observables of the CFT (operators). Given a boundary coupling parameter $\phi_{(0)}$, associated with an operator $\mathcal{O}$ that couples to it, Eq. (4) enables us to calculate the bulk partition function for all bulk fields ϕ with $\phi_{(0)}(x)$ as their boundary condition[25]. Full knowledge of the bulk theory determines, according to (4), the partition function of the CFT and therefore the expectation values of all observables of the CFT, since these can be computed from the partition function. Conversely, knowledge of the boundary CFT partition function leads to full knowledge of the partition function of the quantum gravity theory in the bulk. Although Eq.(4) only states the AdS/CFT correspondence for scalar operators, vector and tensor operators can be handled in a similar way. Equation (4) and its generalizations thus establish a one-to-one mapping between expectation values of observables of the two theories. This is what we mean when we say that the AdS/CFT correspondence is a 'duality'.

In discussions of duality, especially in the context of AdS/CFT, it is sometimes stated that one is dealing with two theories that are the "same".[26] However, the classical actions of the theories are not the same, so that one would expect different behaviour upon quantization. It is quite remarkable that theories that "look very different",[27] nevertheless yield the same numbers for amplitudes. For this reason, some of the original protagonists of AdS/CFT found it comforting that the theories give these numbers in different ranges of expansion parameters: when calculations in one theory are made at strong coupling, the other should be considered at weak coupling, and vice versa.[28]

In his textbook on string theory, Barton Zwiebach described the situation thus: "the term 'duality' is generally used by physicists to refer to the relationship between two systems that have very different descriptions but identical physics."[29] What such characterizations apparently aim at is the existence of a one-to-one correspondence between

---

[25] See for instance (de Haro et al., 2001).

[26] For example, (Aharony et al., 1999), p. 57, write: "we are led to the conjecture that $\mathcal{N}=4$ SU(N) super-Yang-Mills theory in 3+1 dimensions is the same as (or dual to) type IIB superstring theory on AdS$_5 \times S^5$".

[27] (Aharony et al., 1999), p. 60.

[28] (Aharony et al., 1999), p. 60: "In this fashion we avoid any obvious contradiction due to the fact that the two theories look very different. This is the reason that this correspondence is called a 'duality'. The two theories are conjectured to be exactly the same, but when one side is weakly coupled the other is strongly coupled and vice versa. This makes the correspondence both hard to prove and useful, as we can solve a strongly coupled gauge theory via classical supergravity."

[29] (Zwiebach, 2004), p. 376.

physical quantities (observables) and their expectation values on the respective sides of the duality ('observable' is here used in its technical quantum mechanical sense and so refers to physical quantities that in principle could be measured; there is no direct relation to observability by the unaided human senses).

On the other hand, in spite of this 'sameness' the full theoretical structures of the dual theories may appear different. For example, the line element of AdS, $\mathrm{d}s^2 = \frac{\ell^2}{r^2}(\mathrm{d}r^2 - \mathrm{d}t^2 + \mathrm{d}\vec{x}^2)$, does not occur in CFT on Minkowski spacetime, so that a full isomorphism between the mathematical structures of the theories seems absent. But equation (4) nevertheless ensures that numerically correct accounts of any conceivable experiment or problem phrased with one theory's objects and concepts can be given using the concepts and objects of the dual theory even if its mathematical structure looks different.

The above leads to the question of *how* different the dual AdS/CFT theories are; whether they share any structural properties apart from the one-to-one mapping between their observables and expectation values. Obviously, they should share all symmetries between observables. These correspondences between symmetries are indeed found. For example, the space-time symmetry group of the CFT in d dimensions ($\mathrm{SO}(2,d)$) equals the isometry group of $(d+1)$-dimensional AdS. The theories also have a matching number of supersymmetries, and the internal manifold multiplying the AdS factor in the case of $d = 4$, S^5, corresponds to the $\mathrm{SO}(6)$ symmetry of the six scalar fields of $N = 4$ super Yang-Mills. Moreover, both type IIB string theory and super Yang-Mills share a non-perturbative $\mathrm{SL}(2, \mathbb{Z})$ symmetry. These matching symmetries are generally taken as an indication that the AdS/CFT correspondence is *exact*, and not only valid in a perturbative approximation.[30]

If the 'field-operator' correspondence is indeed fully correct, then this could imply an isomorphism between the Hilbert space of string theory in the bulk and the Hilbert space of the CFT. In other words, there could be a one-to-one unitary mapping between the Hilbert space of the bulk and the Hilbert space of the boundary.[31] An isomorphism between Hilbert spaces would obviously imply that much structure is preserved when going from one theory to another, in spite of the absence of a full isomorphism between the mathematical structures of the theories themselves. However, it could be the case that some of the symmetries break down and that as a result the duality itself may prove to be *inexact*—e.g., the match may begin to trail off at some order in the expansions. This distinction between exact and inexact dualities is of importance for interpretational issues, as will become clear later on.

Regardless of whether an exact version of AdS/CFT applies or not, it is clear that AdS/CFT relates bulk degrees of freedom, with gravity, to boundary degrees of freedom of a gravitation-less quantum field theory; it is thus an example of 't Hooft's holography.

The Renormalization Group and AdS/CFT

The gravity side of the AdS/CFT duality suffers from large distance divergences; these are mirrored in the CFT theory by high energy divergences. This is an example of a general feature: large distances on the bulk side are related to low energies on the CFT side

---

[30] As suggested by e.g. (Green, 1999), (Bianchi, 2001), (Drukker et al., 2011).

[31] See (Aharony, 1999) pp. 90-98 and (Giddings, 2011) pp. 15-16 in preprint.

(the infrared or 'IR' part of the spectrum), so that there is an 'UV/IR correspondence'. The divergences we mentioned can be studied with the technique of the 'renormalization group flow' ('RG flow'), which makes the effects of shifts in cutoff parameters explicit. This renormalization technique also plays an important role in Verlinde's scenario.

The RG approach to renormalization, introduced by Ken Wilson in 1974,[32] handles divergences and cutoffs differently from traditional renormalization procedures in quantum field theory. Such procedures typically introduced a cutoff in the integration range of a divergent integral, then performed a calculation (for instance of a path integral) and finally let the cutoff go to infinity. The novelty of Wilson's approach is the insight that there is no absolute need to take cutoffs to infinity: interesting results can be obtained with finite values of the cutoff parameters. This approach disregards higher order quantum processes at energy scales that are above the cutoff value, but this is justified for processes that take place at low energies. Moreover, it is conceivable that completely new theories will be required to deal with processes at very high energies, and that these yet unknown theories will solve the problem of the divergences—this motivates leaving the cutoff at a finite value.

The renormalization group approach begins with limiting the integration range of momenta by introducing a cutoff Λ in the partition function of the theory:[33]

$$Z = \int [\mathcal{D}\phi]_{0 \leq |k| \leq \Lambda} \, e^{-S[\phi]}. \tag{5}$$

Wilson's method essentially consists of repeatedly decreasing the momentum integration range by introducing a novel cutoff $b\Lambda$, while integrating out contributions to the path integral for $b\Lambda \leq |k| \leq \Lambda$. One repeats this process for smaller and smaller b, so $b\Lambda/\Lambda \to 0$. To perform the path integral one splits the field ϕ into Fourier modes $\phi(k)$ with $0 \leq |k| \leq b\Lambda$ and modes $\psi(k)$ with $b\Lambda \leq |k| \leq \Lambda$. As it turns out, the result of integrating out the modes ψ can be represented by an adjustment of the parameters of, and the introduction of additional terms in the original action:

$$Z = \int [\mathcal{D}\phi]_{0 \leq |k| \leq b\Lambda} [\mathcal{D}\psi]_{b\Lambda \leq |k| \leq \Lambda} \, e^{-S[\phi;\psi]} = \int [\mathcal{D}\phi]_{0 \leq |k| \leq b\Lambda} \, e^{-S_{\text{eff}}[\phi]} \tag{6}$$

Rescaling the momenta and coordinates, $k' = k/b$ and $x' = xb$, now leads back to the original range $0 \leq |k'| \leq \Lambda$ in (5)—however, with a new, "effective", action that has "renormalized" couplings. The renormalized action may contain additional terms that were absent from the original form of the action. These additional terms represent the quantum effects of the high-energy modes that were integrated out in the renormalization step.

Successive renormalization steps can be thought of as shifts in a space of theories, with values of the coupling constants on its axes. These shifts, and the corresponding

---

[32]We have relied on (Fisher, 1998) in our presentation of renormalization and renormalization group ideas; philosophical discussions we have found particularly useful include (Hartmann, 2001), (Batterman, 2010).

[33]As Λ is a momentum scale, not the integration range of the field, this step is indicated in subscript under the integration measure.

9

integration over successive shells plus rescaling, defines the 'renormalization group flow'. If a theory is 'renormalizable', the action does not acquire any new terms under renormalization steps. At the endpoint of a flow is a fixed point where the theory becomes scale invariant.

Renormalization group transformations have been interpreted as introducing statistical averaging: when flowing towards a fixed point, some information is thrown away, as processes that become irrelevant in the low energy regime are hidden from view under the RG flow. This procedure can be seen as reducing the fine-grained information available in the quantum description, and RG flow transformations in theory space have accordingly been conceived as steps toward higher entropies.[34] This point of view will be relevant when we discuss Verlinde's ideas.

Fixed points reveal universality classes of theories: different theories may flow towards the same fixed point, and show similar 'statistical' properties. A given Lagrangian may have several fixed points. It may have IR fixed points, as just discussed, but it may also have UV fixed points, i.e. flow toward a fixed point at high energies. A UV fixed point will arise when the corrections are small in the UV, which makes the high energy regime a natural expansion point for the action. According to AdS/CFT, low energy quantum gravity in AdS spacetime is related by the duality relation to a UV fixed point of the corresponding CFT theory. A cutoff at long radial distances in AdS ($1/r = 1/\epsilon$ with ϵ small, i.e. at the boundary of AdS) corresponds to an UV cutoff Λ in the CFT; the inverse of the radial coordinate ($1/r$) in AdS corresponds to the scale $b\Lambda$ up to which the dual field theory is defined.

3.2 AdS/CFT Interpreted

Holography and duality raise interesting interpretational questions.[35] Can one consider one of two dual theories as more fundamental than the other, so that it may become plausible to say that the description given by the less fundamental theory "emerges"? Are we facing situations of empirical underdetermination if there is no difference in fundamentality? After a preliminary look at possible reasons for favouring one theory over another in the context of AdS/CFT, we will attempt a more general appraisal of these questions.

Is one side of the AdS/CFT duality more fundamental?

One option is to consider the non-gravitational theory as more fundamental, and the higher-dimensional space-time and its gravitational degrees of freedom as derived. We saw in section 2 that some of 't Hooft's intuitions went in this direction, when he introduced holography. This viewpoint has the exciting consequence that spacetime (or at least some of its dimensions) would become non-fundamental: apparently (part of) the spacetime description 'comes from' a more fundamental description in non-spatiotemporal terms. For AdS/CFT this point of view has been advocated by, for example, Nathan Seiberg (2007) and Gary Horowitz (2005).

---

[34]For concrete examples, see (Gaite and O'Connor, 1995); (Swingle 2013).
[35]Useful discussions of duality include (Castellani, 2010); (Rickles, 2011).

Seiberg (2007) has argued that spacetime cannot be probed at distances smaller than a certain fundamental length scale, which according to him shows that spacetime cannot be part of a fundamental description: at very high energies the notion of distance loses its meaning. However, apart from a general criticism one might have concerning the step from verifiability to meaning, one should note that in the case of an *exact* duality and correspondence between observables a breakdown of empirical significance in the gravitational theory ('GT') might well be mirrored by a similar defect in the non-gravitational theory ('NGT'), although at a different place in the theoretical structure. So it is not clear at all that this 'loss of verifiability' argument leads to a decision concerning fundamentality.

Horowitz (2005, p. 5) has proposed to consider the gravitational theory as *defined* through the NGT: "since the gauge theory is defined nonperturbatively [in AdS/CFT], one can view this as a nonperturbative and (mostly) background independent definition of string theory." Indeed, due to the UV/IR connection, the NGT is in fact the only available instrument to actually do calculations within the regime of strongly coupled quantum gravity. One should note, however, that assigning precedence to the NGT based on this instrumental aspect has a pragmatic character, at least if one accepts that the duality between NGT and GT is exact. After all, an exact duality implies a one-to-one relation between the values of physical quantities, so that in this case it seems impossible to claim a *descriptive* superiority of NGT over GT: as far as observables are concerned, the NGT and GT describe the physical world equally well or equally badly, even if one theory is more tractable than the other in a certain regime.

In the literature, one finds a near unanimous consensus that the AdS/CFT duality should be taken to be exact, even if there is not yet a proof of this exactness; still, many positive tests underwrite this consensus.[36] If the duality is *not* exact, the question of the relative status of NGT and GT is relatively simple: in this case the correspondence between observables can only be approximate, so that the straight-forward question arises which one of the two theories is better confirmed by experiment. The question of fundamentality in this case reduces to a question of empirical adequacy, even if an actual empirical verification of the differences between the two theories could presently be out of experimental reach.

If one of the AdS/CFT theories would thus turn out to be more fundamental than the other on empirical grounds, it could of course be that the gravitational side is found to be the more fundamental one. In this case there would clearly be no motivation for the notion that spacetime and gravity emerge from the boundary description. For instance, it may be that exact duality fails in strong quantum gravity regimes, far from the semi-classical limit, where not many tests have yet been performed; it could be that strong quantum gravity phenomena cannot be captured by a CFT. The gravity side of the duality would be superior in this case, in the uncontroversial sense of better fitting nature, even if it would be convenient to employ the NGT as an instrument in calculations. The NGT would then be a calculational tool of limited validity.

However, one instead more often encounters in the literature the notion that spacetime and gravity are derivative and emergent in some way.[37] The difficulties that non-

---

[36]See for instance (Aharony et al. 2000).

[37]Even Albert Einstein expressed that if one desired a *quantum* theory of gravity, one would have to

perturbative formulations of quantum gravity encounter, in combination with the universal character of gravity that distinguishes it from other forces, may play a role in this expectation and in the relative unpopularity of the point of view that gravitation is basic.[38]

Duality, fundamentality and emergence

The distinction between exact and approximate dualities is important. In the latter case empirical equivalence between theories in a dual pair is not to be expected, so that uncontroversial criteria for theory comparison can be deployed. In this situation there is clear scope for the notion of 'emergence'. The duality now boils down to an inter-theoretic relation that could resemble the one between thermodynamics and atomic theory: one description approximates the other. In this analogy the atomistic description is uncontroversially the more fundamental one, even if in most practical situations it is impossibly inconvenient to take recourse to calculations in terms of atoms—this latter circumstance is relevant to the pragmatics of the situation, not to considerations about fundamentality. Continuing the analogy: on the thermodynamic level concepts like 'temperature' and 'pressure' become applicable—concepts that capture objective aspects of physical reality, even though they cannot be applied to the more fundamental atomistic level of description. 'Temperature' and 'pressure' can be said to emerge from the atomistic description, in a clear and uncontroversial sense of emergence: these concepts figure in the description of novel and robust behaviour that is insensitive to the underlying atomistic and molecular details. This emergence involves an asymmetry between the theories that are involved: thermodynamics emerges from the atomistic description, but not the other way around. A relation of approximate duality might well be similar in relevant aspects and give rise to an effective description that emerges from the more fundamental theory in the dual pair; as we will see in section 4, the situation in Verlinde's proposal is a case in point.

However, most discussions about duality and its philosophical consequences take place against the background of the assumption that the duality is exact, and here it is less clear how we should judge the relative status of the theories that are involved. By definition there is in this case a precise one-to-one mapping between the observables and their values in the two theories. This suggests that the theories are empirically equivalent: for each physically significant number in one theory there is an exact counterpart in the other.

One might object that this mapping by itself does not imply empirical equivalence, on the grounds that these equal numbers may pertain to different physical quantities, and different regimes of coupling strengths. But this objection presupposes that the

___

get rid of the spacetime continuum and thus arrive at a "purely algebraic physics"; see his letter to Paul Langevin, 3 October 1935, cited in (Stachel 1993), p. 285. Nevertheless, he strongly preferred to stick with his own attempts at a continuum unification theory, despite these insights; see (van Dongen 2010), pp. 174-183.

[38]We have not found, in the AdS/CFT literature, any explicit statements that gravity should be considered as more fundamental than gauge theories. However, there are numerous articles in which the gauge theory side of the duality is used as a *tool* for predicting bulk physics, while the latter seems implicitly assumed to be more fundamental: as boundary calculations are here treated merely instrumentally, these applications do suggest a presumed precedence of the bulk over the boundary. See for instance the discussion of the Big Bang scenario by Hertog and Horowitz (2005).

physical meaning of the quantities in each of the theories has been fixed independently; that we already know what terms such as 'energy' and 'distance' mean before we start contemplating how such terms figure in the theories. This will be the case if there exists an external point of view, from outside the two theories, that determines the reference relation between the theories and physical reality. If this is the situation that is being considered, duality between theories expresses a *symmetry* in the physical world: exactly the same relations that obtain between, say, the energies of particles also obtain between certain distances.

An example of this kind is provided by the source-free Maxwell equations, which exhibit perfect symmetry between the E and B fields. When we consider the application of these equations to a source-free region of space, the form invariance under an exchange of E and B reflects a physical property of this region, namely that it is source-free. However, if other regions do contain charges, this determines objectively what the E and B refer to. Exchanging E and B in the source-free region does not change anything in the form of the equations, but does imply a drastic change in the physical situation that is described: electric fields are replaced by magnetic fields. So in this case the duality connects *different* aspects of the world that possess an isomorphic internal structure. This makes it possible that one theory can serve as a *model* for the other.

Clearly, in this situation the notion that the duality is connected with emergence does not even suggest itself. With regard to the example: the symmetry between E and B does not entail anything about a possible origin of electricity in magnetism or the other way around. Similarly, reflection symmetry has no implication for a possible emergence of 'left' from 'right' or *vice versa*.

The situation becomes more interesting, and more in the spirit of discussions about duality in the context of present-day fundamental physics, if the two dual theories are candidate descriptions of the whole world, without an external point of view that fixes the meanings of terms. In this case we are dealing with two structures of observables and their values that have exactly the same internal relations to each other in the two respective theories. Without an external viewpoint, the only thing to go on with regard to the meaning of observables are their positions in the two networks of relations. But this means that we arrive at the conclusion that isomorphism between the structures of observables can be cashed out in terms of *equality* rather than *symmetry*. The symbols used in the two theories may be very different, but in view of the role quantities play in relation to other quantities, and the values they assume, identifications can be established: A in one theory will denote exactly the same physical quantity as B denotes in the other if these quantities occupy structurally identical nodes in their respective webs of observables and assume the same values.[39]

We are thus dealing with a strong case of empirical equivalence: even though the complete mathematical structures of the two theories may well differ, their substructures of observables will coincide. In the recent philosophical literature about empirical equivalence and underdetermination (for the greater part responding to the seminal paper by

---

[39]What is assumed here is not a doctrine about what the world is like, but rather a view about how a mathematically formulated 'theory of everything' can correspond to the world if no *a priori* 'rules of correspondence' between the theory and world are given. In this case it is only the internal structure of the theory that can decide how it can be applied to the world. So the structuralism here is epistemic.

Laudan and Leplin) the possibility of such a thoroughgoing empirical equivalence is often doubted.[40] But the case of exact duality resists most of the usual arguments. For instance, in examples of (potential) duality discussed in the literature it certainly is not true that one of the two theories fails to meet standards of theoreticity, or is an artificial parasite on the other. Such standard objections against empirical equivalence have the purpose of removing the threat of theoretical underdetermination, the dilemma that arises when it is impossible to reach an empirically justified choice between theories. But exact duality seems to revive this threat, by surviving standard objections against empirical equivalence.

In the relevant philosophical literature it is usually argued that it would testify to a superficial instrumentalism to *identify* two empirically equivalent theories—to consider them as just variant formulations of one and the same theory. The thought behind this objection is that differences in theoretical structure between theories may well refer to differences in physical reality, even if these differences are not (yet) observable: we should not assume that the descriptive physical content of theories is exhausted by the theory's observable consequences. But in our cases of exact duality the situation is different from what is usually assumed in philosophical discussions of empirical equivalence. As we have pointed out before, the 'observables' that are in one-to-one correspondence with each other in cases of exact duality are not defined via a notion of observability like in the debate about empiricism and scientific realism. Rather, they stand for what is physically real and meaningful according to the theories under discussion (i.e. expectation values of all physical quantities), even if there are no possibilities of direct observation. So what we are facing is not the standard situation of empirical equivalence in which two different physical theories coincide 'on the surface of observable phenomena': we are dealing with theories that coincide exactly in everything they deem physically real.

We therefore conclude that in the case of an exact duality between theories without external rules of correspondence a very strong form of empirical equivalence arises that does *not* lead to theoretical underdetermination. Because it is inherent in the notion of exact duality in this case that the two theories completely agree on what is physically meaningful, the two sides of the duality can only be taken to be different representations of one and the same physical theory. The two theories collapse into one; and there is no emergence of one side of the duality from the other. Emergence therefore seems a potentially applicable notion only when we are dealing with approximate duality.

Renormalization and emergence in AdS/CFT

However, even when we are dealing with essentially one theory, *within* this theory there could still be emergence. For example, there could be an effective macroscopic description if the system possesses very many degrees of freedom and this could justify speaking about an emergent macro-behaviour. Could Einstein gravity be emergent in this way in AdS/CFT? This is not how one has usually imagined emergence in the context of AdS/CFT, but it may nevertheless be a possible scenario.

---

[40]See (Laudan and Leplin, 1991); for a recent critical discussion of their proposal, see (Acuña and Dieks, 2014).

We have already noted that in the vast majority of actual examples of AdS/CFT one relates a *semi-classical* bulk theory to a CFT that is considered at large N, i.e. for high values of the rank of the gauge group $\mathrm{SU}(N)$. This represents a particular kind of (semi-)classical limit,[41] which leaves open whether gravity as we know it from classical theories is fundamental. In fact, it has been suggested that the metric field (the central quantity in general relativity) will not be one of the fundamental fields in a fundamental microscopic bulk theory but will somehow appear in a limit, from a fundamental microscopic theory that has very different space-time properties. A concrete example that supports this idea is the result in string theory that the metric is reproduced in the regime of very small string length.[42] One might thus speculate that ordinary space-time concepts in the bulk only make sense after taking some limit.

There may be indications that that limit is thermodynamic. It turns out that progressively neglecting quantum corrections to the Einstein equations in the bulk corresponds, via the AdS/CFT duality, to renormalization transformations in the CFT that throw out higher order terms in the action.[43] If we interpret this sequence of coarse graining/renormalization steps as the transition to a thermodynamic limit, we get a hint of how a thermodynamic limit on the boundary may be associated with the emergence of classical Einsteinian gravity in the bulk. In such a scenario gravity is not emergent due to duality but rather because of coarse graining and the existence of a huge number of degrees of freedom. As we will see, this way of creating a place for emergence in AdS/CFT is closely related to Verlinde's proposal.

4 Gravity as an entropic force

A further holographic scenario that we want to analyze in some detail is the recent rather spectacular theory of gravity proposed by Erik Verlinde. We first list its key assumptions, and add details in section 4.2. We will try to disentangle the logical structure of Verlinde's argument and assess some of its conceptual and interpretative consequences in 4.3.

4.1 Holography and the Newton force law

Imagine a closed two-dimensional space, e.g. the surface of a sphere, on which a quantum theory is defined. We are going to remain quite unspecific about this quantum theory (this is one of the salient points in Verlinde's proposal). It is sufficient to assume that the theory describes physical processes on the surface by means of bits and that these processes can therefore be characterized in an information-theoretical way, as 'changes

---

[41]Technically, for the case of a four-dimensional CFT, one takes N to be large but keeps the product $g^2 N$ fixed, where g is the coupling constant. The quantity that is held fixed is called the 't Hooft coupling. In this limit, only a limited class of Feynman diagrams (called 'planar' because they can be written on the plane) contribute to the observables, and these diagrams are generally reproduced by the saddle point of a classical theory.

[42]See e.g. (Green et al., 1987), p. 115, where an excitation of the string is found that corresponds to a nearly flat ambient metric. This also requires small string coupling. Increasing the string coupling allows for more highly curved metrics, see pp. 166-183.

[43]See (de Boer et al., 1999).

in information'. Accordingly, the surface area of the sphere is imagined to be divided in small cells, each of which can contain one bit of information. The holographic principle now suggests that such a surface theory can also fully represent physical processes that go on *inside* the sphere. In particular, the bits on the surface may encode where matter is located in the interior.

The number of bits on the sphere is assumed to be very large, so that an effective thermodynamical description can be used instead of the original quantum theory defined on the micro-level of the cells and bits. The physical processes that take place on the surface (changes in the 0-s and 1-s in the cells) can thermodynamically be characterized as processes that maximize entropy: the distribution of 0-s and 1-s tends to an equilibrium distribution. The core of Verlinde's proposal is that this entropic process on the surface corresponds, via the holographic principle, to *gravitational* processes in the bulk (the interior of the sphere). In other words, the changes in 0-s and 1-s on the surface, described in the thermodynamical limit, yields a gravitational description of matter in the interior (e.g. for matter that is falling inward). This idea is made plausible through a simple deduction of Newton's law by a holographic translation of thermodynamical equations on the surface—we will reproduce this derivation here.

First, the number of active bits N on the sphere is assumed to be proportional to the sphere's total surface area A (constants are introduced and handpicked for later convenience):

$$N = \frac{Ac^3}{G\hbar} \ , \tag{7}$$

in which the area of the sphere is

$$A = 4\pi R^2. \tag{8}$$

From the internal point of view of the surface theory there is no outside world and no third dimension, so that from this perspective R can be considered as a quantity that is *defined* by Eqs.(8) and (7). As already announced, it will be a central assumption in the derivation of Newton's law that the thermodynamic limit can be taken on the surface. A *temperature T* will therefore be definable, and a thermodynamic energy E that obeys the law of equipartition:

$$E = \frac{1}{2} N k_{\mathrm{B}} T \ . \tag{9}$$

We can now *define* the quantity M by:

$$E = Mc^2. \tag{10}$$

On the screen, M is just an alternative expression for the thermodynamic energy; but via the holographic correspondence it will soon acquire the interpretation of the total mass that is present in the interior.

As we have seen in the discussion of the Renormalization Group (see section 3), the quantum theory on the surface can be subjected to coarse-graining renormalization steps.
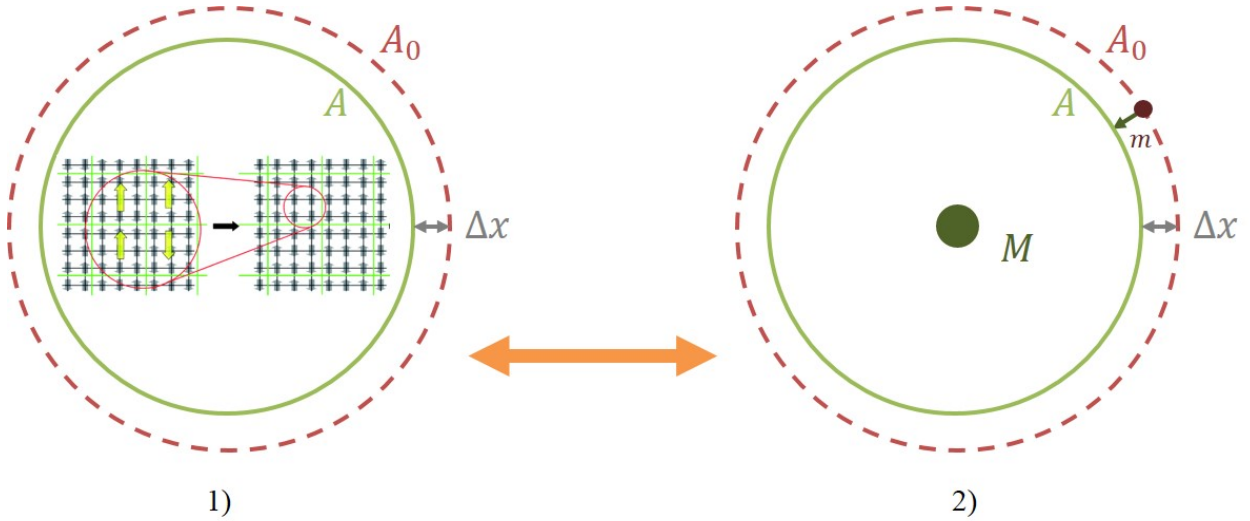
Figure 1: A test mass in three-dimensional space, close to the position of one of the screens, feels a force. This force, due to the holographic hypothesis, can be seen as an expression of the tendency towards increasing entropy on the screen. The growth in entropy allows for increasingly coarse grained descriptions realized by renormalization transformations in the surface system, as depicted on the left; renormalization group transformations are implied when the screens 'follow the particle inwards'. On the sphere on the left, an example is sketched of 'integrating out' degrees of freedom via an RG-like step, i.e., a block spin transformation: groups of microspins are identified with an overall spin value, after which the lattice is reduced in size, along with the reduction in surface area of the sphere. Such a transformation is associated with an increase in entropy, as 'micro-information is thrown out', i.e. in the coarse grained description single spin values in fact represent more spins on a 'finer grained' level. This microscopic structure becomes irrelevant when the system relaxes.

A concrete example is renormalization of a theory that describes bits as quantum spins on a lattice. The renormalization steps will produce a change in lattice cell sizes,[44] and thus effectively change the area of the surface (by changing the size of the cells and subsequent rescaling). Going to a more coarse grained description, and therefore discarding part of the fine grained information (in the high energy modes), is thus equivalent to considering a quantum theory on a smaller surface than we had before (see figure 1; we will return to this). Using the relation between surface and radial distance, (8), we are thus led to a representation in three dimensions by means of a nested set of spheres, all with the same center but with different radii. In this way it becomes possible to talk about both the inside and the outside of any given sphere, although all quantities were originally defined within two-dimensional theories.

Now imagine that in this external bulk description (in three-dimensional space) a particle with mass m comes from outside a screen and changes its distance with respect to it by Δx. Via holography, the bits on the screen encode everything that is going on within the screen's interior. After the approach of the particle, and its subsequent fall into

---

[44]On these transformations, see e.g. (Fisher 1998) pp. 666-669.

the interior, there are more particles inside and therefore more has to be encoded on the surface. This must correspond to an increase in the number of active bits on the screen and an increase in the entropy. Generalizing Bekenstein's ideas about entropy changes when masses fall into a black hole, Verlinde takes this change in entropy to be:[45]

$$\Delta S = 2\pi k_{\mathrm{B}} \, \frac{mc}{\hbar} \, \Delta x \; . \tag{11}$$

At the thermodynamic level the process of falling inward can be described as an approach towards equilibrium, which can be characterized phenomenologically as the result of the operation of an effective 'entropic force' F that relates to changes in entropy:

$$F \, \Delta x = T \Delta S \; . \tag{12}$$

The peculiarity of an entropic force is that it does not derive from an interaction on the microscopic level, but only arises in an effective thermodynamic description. A typical example is the force that can be used to describe the behavior of a polymer, stretched in the direction Δx. On the fundamental level, viz. of the atoms that make up the polymer, there need not be any force: the polymer may consist of short chains of atoms that are connected but can rotate freely with respect to each other. However, as a result of random thermal motion, the polymer will with overwhelming probability end up in a macroscopic state that corresponds to a large phase space volume; this will be a state in which the polymer is coiled up (there are many more coiled up microstates than states in which the polymer is stretched). So from the macroscopic point of view a definite directedness in the behaviour of the polymer manifests itself: it tends to coil up, in spite of the microscopic randomness. This tendency (associated with an increase in entropy) can be phenomenologically described as caused by an elastic force obeying Eq. (12).

Going back to our case of a particle that approaches a surface while increasing its entropy, we can now determine the magnitude of the effective force. Simply combining the above relations (7)–(12) yields the suggestive result

$$F = G \, \frac{Mm}{R^2} \; . \tag{13}$$

The conclusion therefore suggests itself that gravity is an *entropic force* whose "corresponding potential has no microscopic meaning", as Erik Verlinde puts it.[46] In his paper Verlinde shows that it is possible to give a similar derivation of the Einstein equations.[47]

---

[45] Verlinde fixes the numerical factor relating ΔS and Δx by consideration of a thought experiment worked out by Jakob Bekenstein in the 1970s in the context of black holes: Bekenstein had argued that when a particle is added to a black hole, the latter's area increases with, minimally, $8\pi\hbar$, which can be added when the particle is at one Compton wavelength from the horizon (Bekenstein 1973, p. 2338). Arguably, it can then no longer be distinguished from the black hole. In the same way, if the particle is at the distance of its Compton wavelength from the screen, the entropy on the screen is raised by one bit (with a factor of 2π put in by hand): $\Delta S = 2\pi k_{\mathrm{B}}$ when $\Delta x = \frac{\hbar}{mc}$. Generalizing for arbitrary distances leads to relation (11).

[46] (Verlinde, 2011) p. 4.

[47] A related derivation of the Einstein equations has been given earlier by T. Jacobson (1995).

So space and gravity may be explainable from a non-spatial and non-gravitational background: in the above account a third spatial dimension was introduced from considerations on a two-dimensional surface (the third dimension appeared as a coarse-graining parameter in the surface theories) and gravitation came in as the three-dimensional translation of a thermodynamic description of the surface. Indeed, Verlinde states that he has "reversed the arguments" that have yielded holography and black hole thermodynamics, so that from holography and thermodynamics we can understand gravity, and that this has "shed new light on the origin of gravity".[48] But there are questions about the status of the various assumptions that have been made. And, most important for our purposes here, is it justified to say that in Verlinde's scenario the surface theory is *more fundamental* than the bulk theory, so that the surface theory can be called the *origin* of gravity?

Two essential ingredients in Verlinde's proposal are responsible for the derivability of gravity from the gravitation-less surface theory, namely 1) the holographic correspondence between surface and bulk descriptions, and 2) the transition from the microscopic to the thermodynamic mode of description, which grounds the characterization of gravity as an entropic phenomenon. We will now discuss these two core assumptions in more detail.

4.2 The correspondence between information loss and gravity

The first correspondence we consider, **a**, is the holographic identification of degrees of freedom between two microscopic systems *a1* and *a2* (see Figure 2). The system *a1* is defined on the surface and is described by a theory without gravity. The microscopic dynamical details of this system are not relevant, but it should be possible to speak about the thermodynamic limit and the number of degrees of freedom of the theory for the argument to take hold. The system *a2* is a system of masses inside the sphere. In contrast to what is usually assumed in discussions of the holographic principle, in Verlinde's scenario we do not need to assume that *a2* is described by a microscopic theory including gravity in any traditional sense (a quantum gravity theory, or Einsteinian gravity). The theory dealing with *a2* is just a three-dimensional translation of the microscopic surface theory and could still be without gravity in a recognizable form. In Verlinde's proposal gravity as we know it is taken to arise from thermodynamics.[49]

Holography is the leading idea in this correspondence between two-dimensional and three-dimensional descriptions. Verlinde does not explicitly state which precise form of holographic relation he has in mind, but his various remarks indicate that, initially at least, he is thinking of a *bijective map* between a microscopic quantum theory in the bulk and a microscopic quantum theory on the screen, so that we have an exact duality. We will return to this later.

The correspondence **a** leads naturally to the consideration of a second correspondence

---

[48](Verlinde, 2011) p. 9.

[49]On this point Verlinde's discussion could in principle be falsified: if the theory *a2* would turn out to be a theory of gravity, then gravity would not *emerge* as stipulated in the Verlinde scenario. In that case, however, one must also determine when microscopic phenomena count as 'gravity' (for instance, a microscopic force mediated by a spin two excitation connected with diffeomorphism invariance may already qualify as gravity). For the sake of our analysis of Verlinde's work, we will use the term 'gravity' in a restricted sense, namely either as Newtonian or Einsteinian gravity.
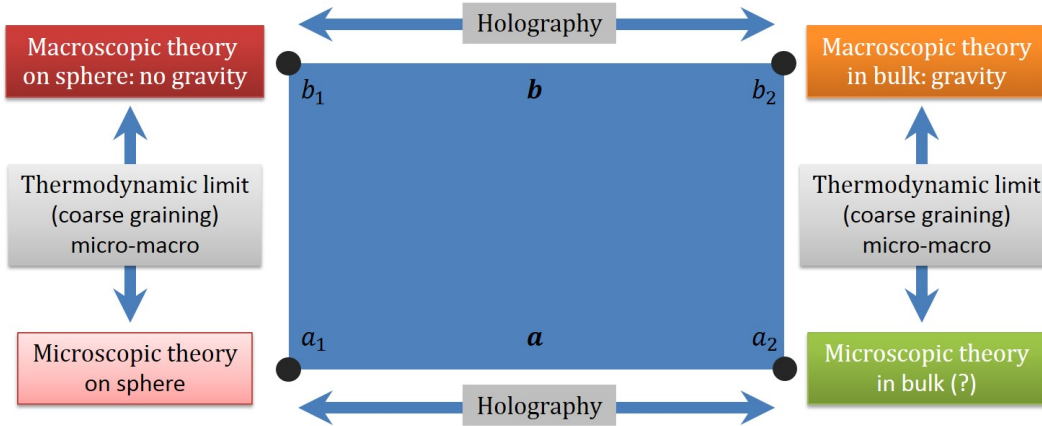
Figure 2: First interpretation of Verlinde's scenario for holography and emergence: at the microscopic scale, one generally assumes that a holographic and exact duality **a** holds between screen quantum theory *a1* and bulk quantum theory *a2*. For large numbers of degrees of freedom, coarse grained descriptions suffice for capturing the relevant physics; the microscopic duality and the thermodynamic limit give a new holographic relation **b** between a thermodynamical description of the system on the screen, *b1*, and gravity in the bulk, *b2*.

relation, **b**, between two systems *b1* and *b2*. These two systems are the macroscopic or thermodynamic versions of the systems mentioned under **a**. The system *b1* is the system described by the surface theory again, but now considered in the limit in which the number of degrees of freedom is very large and an effective macroscopic description can be employed (see figure 2). This transition to a macroscopic description consists in the 'throwing away' of irrelevant degrees of freedom, which can formally be represented by RG transformations, block spin transformations, or similar coarse graining steps. In the surface language: when less degrees of freedom are taken into account, a smaller surface suffices for the description (a surface with less cells, and therefore with less information carrying bits). As macrostates that can be realized by more microstates are more probable, the system will move towards them, as in the analogy of the released polymer. While the system evolves to states that are closer to equilibrium, it becomes equally well describable, from the macroscopic point of view, by less microscopic variables; less 'information' is relevant to the macroscopic description of the system, and in this way entropy grows.

As we saw before, a concrete example of how this can work is given by a screen theory that describes a large number of quantum spins on a lattice. While the spin system diffuses naturally, it can be characterized by increasingly coarse grained theories: successive block spin transformations lead to more efficient theories, with less degrees of freedom but equally well suited to describe the system. In block spin transformations averaging over spins in a number of cells takes place, after which several individual spins are replaced with one bigger cell in which there is one new spin value, namely the average of the original cell values. This in turn is followed by a rescaling of the cell size, so that we finally arrive at a smaller copy of the original lattice, i.e. a lattice with less spin positions.

The renormalization steps thus produce a smaller copy of the screen with less microscopic information. In a three-dimensional picture the new screen can be imagined as placed inside the original one; the striving for thermodynamic equilibrium on the screen then corresponds to the adequacy of using a succession of increasingly coarse grained theories, defined on smaller and smaller spheres with shrinking interiors.[50]

The system $b2$ is the system described by the holographic counterpart of the theory describing $a2$, again in the thermodynamic limit. As we have seen, Verlinde's central claim is that this macroscopic bulk theory describes the interior of the spheres in terms of masses and gravitational forces between them: it is a gravitational theory. Via the holographic correspondence **b** (conceived as a bijective mapping) $b1$ and $b2$ become two alternative ways of describing the same thermodynamical system. For example, the parameter x (Eq. 12) is defined as a cutoff parameter in the surface theory, which keeps track of coarse-graining steps on the surface, but it becomes an added spatial dimension in the bulk description. However, in both cases it figures in the same formulas so that structurally the descriptions are the same. The process of entropy growth on the screens is thus represented as a gravitational process that needs less and less space for its description—because masses fall inwards during their approach to gravitational equilibrium—and in which all gravitational bulk properties correspond one-to-one to thermodynamical quantities defined on the screens. So, for example, the gradient of the gravitational potential in the bulk turns out to track the level of coarse graining of the surface theories; the force felt by a test particle in the bulk in this way translates the entropy gains on the surfaces.

One argument in favour of these ideas is that in the thermodynamic limit details of the underlying microscopic theory become unimportant: there exists a striking universality in thermodynamic behaviour. This universality now appears as connected to the universality of gravitational attraction: all systems, whatever their non-gravitational interactions, display the same gravitational behaviour. If gravitation is indeed the manifestation of thermodynamic behaviour of a system that at the microscopic level is gravitation-less, then the universality of gravitation has the prospect of being explained in the same way as the universality of thermodynamics.

The correspondence **b** in our scheme can be seen as resulting from the combination of a surface-bulk correspondence at the micro-level and the thermodynamic limit (see Figure 2). One interesting feature arising from this combination of ideas from holography and thermodynamics is that the correspondence **a** on the microscopic level could link two *non*-gravitational theories, one in two dimensions and one in three. On the other hand, the correspondence **b** is between a non-gravitational thermodynamic surface theory and a theory of gravity in the bulk, as usual in holography (cf. the introduction of holography by 't Hooft, reported in section 2). In Verlinde's scenario the standard holographic relation can therefore be interpreted as arising from a more elementary non-gravitational holographic mapping **a**, plus a thermodynamic limit. The presentation of Verlinde's sce-

---

[50]Note that the *maximum* possible amount of entropy *decreases* as the surface decreases. In the gravitational correspondence that we are discussing, the screen capacity cannot be further reduced when we reach the horizon of a black hole. This is the final equilibrium situation in which a further growth of entropy is impossible. Until that point has been reached, however, 'reducing screen size by a renormalization group step' or considering successively smaller screens can go hand in hand with increasing physical entropy.

nario in Figure 2 differs from the AdS/CFT case discussed earlier in that the microscopic theory of the former does not need to contain gravity, whereas in usual interpretations of AdS/CFT it is assumed that *a2* is a microscopic theory of gravity.

Actually, for Verlinde's argument an appeal to a microscopic theory on the bulk side (dealing with *a2*) is not necessary at all: such a theory plays no active role in Verlinde's argument for the emergence of gravity. The holographic reinterpretation of the thermodynamics on the screen suffices for the introduction of gravity; so it is sufficient to look at the bulk counterpart of the surface theory in which the thermodynamic limit has already been taken. This raises the question of whether we have to assume a mapping between microscopic theories, **a** in the above, at all. In other words, we should consider the possibility that there might be a holographic mapping between bulk and surface theories only *after* the thermodynamic limit has been taken, on the macro-level. Anticipating a more extensive analysis, we might say that holography itself could emerge in the thermodynamical limit.

4.3   Emergence, holography, and thermodynamics

Verlinde presents his proposal as a scheme in which three-dimensional gravity emerges from a two-dimensional theory without gravity—as already indicated by the title of his paper: "On the *Origin* of Gravity and the Laws of Newton". In this section we consider the question whether use of the term 'emergence' is justified: can it be said that the surface theory is more fundamental than the bulk theory? And, as a preliminary to that question, what exactly does it mean to say that space and gravity emerge?

As explained in the previous sections, *two* relations play an essential role in the correspondence between the surface theory and the bulk: the holographic correspondence and the thermodynamic limit. First we will add some comments on how holography and coarse graining work together in Verlinde's scheme, then we will discuss to what extent gravity can be said to emerge, and finally we will address the question of fundamentality.

As we pointed out in the discussion of Figure 2, invoking *a2* is not needed for arriving at the gravitational system *b2*. One may therefore consider an alternate possibility in which the holographic relation **a** is not there and in which **b** appears as an independent relation that only makes sense on the thermodynamic level, see figure 3. In this case a holographic relation exists only between the screen theory in which the thermodynamic limit has been taken and the gravitational theory in the bulk. This holographic relation will not be analyzable as the combined result of a more fundamental microscopic holographic relation plus the thermodynamic limit: according to this suggestion there simply *is* no holography at the microscopic level—holography is a thermodynamic phenomenon itself.

In the context of the conceptual possibility represented by figure 2 we have to distinguish between the holographic relation between surface and bulk and the thermodynamic limit that gives rise to an effective description. Verlinde's article discusses the emergence of gravity and space without differentiating these two relations too strictly; however, the difference between them is important. As we have argued in section 3.2, in the case of an exact duality between surface and bulk we can differentiate between an internal and an external point of view. Judging from an externally fixed viewpoint, the two theories linked by the duality are structurally similar but still different: indeed, one is about a
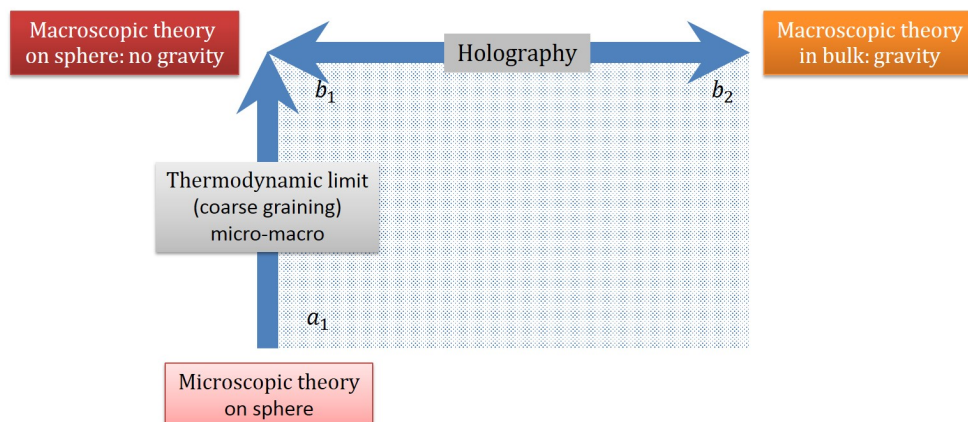
Figure 3: Second interpretation of Verlinde's emergence scenario: thermodynamic emergence of space, gravity and holography without a microscopic bulk theory.

two-dimensional surface and the other about three-dimensional space. However, if an external point of reference is lacking, as in fundamental theories that in principle describe the whole universe, we cannot tell *a priori* which quantities should be called spatial and how many spatial dimensions there are—in this case the meaning of all physical quantities can only be given by the role they play in the theoretical framework. According to this internal viewpoint, which seems the appropriate one in the present context, the fact that all observables and their relations in the two respective theories stand in a bijective correspondence to each other means that we are dealing with two different formulations of one and the same physical theory.

It follows, as already argued in 3.2, that the notion that gravity and space can emerge out of an exact holographic duality is mistaken. In the case of an exact duality no novelty of behaviour can be present on one side of the duality if it is not also there on the other side—there is complete symmetry. On the other hand, that taking the thermodynamic limit may lead to emergence is a more or less standard conclusion. Indeed, the transition to thermodynamics opens up a new level of description that is characterized by new concepts and by patterns of physical behaviour that to a large extent are independent of the microscopic details of the underlying theory. In this sense gravity *can* be said to emerge in Verlinde's proposal: as we have seen, it appears as an entropic force that has no counterpart on the micro-level. Its characteristics are independent of the details of the microscopic interactions and depend only on thermodynamic relations. Thus, the universality of gravitation receives an unexpected explanation: it is due to its emergent thermodynamic character.

Verlinde writes that according to his proposal "space is emergent through a holographic scenario".[51] As we have just seen, holography by itself cannot be responsible for this emergence; if space is emergent in Verlinde's scheme, it must be the thermodynamic limit that does the work. But is space really emergent in Verlinde's proposal? It is true that

---

[51]From the abstract in (Verlinde, 2011).

the spatial coordinate in the bulk theory corresponds to a coarse graining variable in the surface theory (or the number of renormalization steps that were taken), but this variable is just a *parameter* that keeps track of the level of coarse graining (see section 4.1). In other words, it is not a thermodynamic quantity. The parameter x on the surface side is *reinterpreted* as a spatial coordinate on the bulk side via the holographic connection, but as we have discussed, this reinterpretation by itself does not lead to emergence. It therefore follows that although gravity can be said to emerge as a thermodynamic phenomenon, space itself does *not* emerge in this scenario.

The lower-right corner of figure 2 is the place where a microscopic bulk counterpart of the microscopic screen theory should be located according to standard holographic ideas. We have noted several times that we can do without this microscopic bulk counterpart of the surface theory, and may consider a scheme in which system *a2* no longer occurs—see figure 3. This agrees with one of the motivating ideas behind Verlinde's approach, namely that there may be no need for quantizing gravity and for grand unification. "The quest for unification of gravity with these other [quantum] forces of Nature, at a microscopic level, may [...] not be the right approach", Verlinde writes, with reference to the many problems that this quantization approach has produced.[52] This accords with the idea that there is no microscopic bulk theory *of gravity*; but it would also agree with the idea that there is no microscopic three-dimensional bulk at all.

When we accept the analysis of the situation as depicted in Figure 3, the possibilities with respect to emergence change drastically. The holographic correspondence now only arises *after*, and *because*, we have taken the thermodynamic limit, and there is no holography on the micro-level. In other words, holography itself arises as a novel feature on the level of thermodynamic description: it becomes an emergent phenomenon itself. But in this case the whole correspondence between surface theory and its three-dimensional counterpart makes sense only on the level of an effective thermodynamic description, and the existence of three-dimensional space need not be admitted on the micro-level. As a consequence, in this alternative scenario the thesis that "space emerges together with gravity"[53] *can* be justified.

In the above we have used novelty of behaviour and robustness, in the sense of independence of microscopic details, as criteria for emergence. This is in accordance with Verlinde's leading thought: "gravity can emerge from a microscopic description that does not know about its existence. The universality of gravity suggests that its emergence should be understood from general principles that are independent of the specific details of the underlying microscopic theory."[54] The universality of gravitation is an example of thermodynamic universality in Verlinde's approach; gravity is 'robust' in the sense that the form of the emergent force is stable under a wide range of choices for the microscopic surface theory.

---

[52](Verlinde, 2011), on pp. 1-2.
[53](Verlinde, 2011), on p. 2.
[54]As in (Verlinde, 2011), p. 2.

5 Conclusion: emergence and holography

We have reviewed three cases: 't Hooft's original holographic proposal, AdS/CFT, and Verlinde's recent scenario. In 't Hooft's 1993 introduction of the holographic hypothesis we have not found a clear argument for the existence of a relation of emergence, even though 't Hooft's text in places suggests a more fundamental status for the physics on the boundary. The original introduction of holography is programmatic and rather ambiguous as far as interpretation is concerned.

The case of AdS/CFT is more clear-cut, because in it the notion of holography is made more precise due to the duality relation. We have argued that if the duality is exact, as is generally expected and suggested by many tests, there is no reason to consider one of the two holographically related quantum theories as emergent from the other: bulk and boundary theories are two representations of one theory. Although many have expressed the intuition that the spacetime bulk is somehow emergent from the boundary field theory, this does not seem a viable interpretation when exact duality is accepted together with an internalist viewpoint. The only possible place for emergence in this case appears to be the emergence of *Einsteinian* gravity, via RG flow, from an underlying microscopic theory that is not explicitly gravitational. This is, however, *not* the same thing as emergence of gravity from the boundary theory.

Continuing this thought, it is important to note that in studies of AdS/CFT the gravity side has mostly been formulated in a semi-classical regime of one kind or another, which means that a macroscopic limit has implicitly been taken. It could be that only in this limit a spacetime point of view becomes generally applicable. In this case the holographic relation would arise together with the interpretation of certain degrees of freedom as gravitational. In this scenario gravity could be a manifestation of thermodynamic behaviour and a microscopic quantum gravity theory would not be needed.

In this (non-standard) AdS/CFT interpretation, AdS/CFT would fit the same schema as the interpretation of Verlinde's scenario that we have proposed in connection with Figure 3. In the latter case, we have found that understanding gravity as emergent is fairly unproblematic: gravity has a thermodynamical background, and the gravitational force is an entropic force without a corresponding microphysical interaction. One can interpret Verlinde's proposal such that holography, spacetime and gravity emerge together in the thermodynamical limit of a microphysical theory on a screen. Could one permit this kind of interpretation also for AdS/CFT? Verlinde seems to say as much when he states that the "gravitational side" of AdS/CFT should not be seen as "independently defined"; he compares referring to gravity in AdS/CFT to "using stress tensors in a continuous medium half a century before knowing about atoms."[55] This appears to go in the direction of a critical reinterpretation of AdS/CFT in which one abandons the notion of microscopic gravity. Of course, Verlinde's own proposal takes an explicit thermodynamical step, which is not present in standard discussions of AdS/CFT. However, in AdS/CFT moving inward in the bulk is related to coarse-graining renormalization transformations on the boundary which suggests a relation with statistical physics and thermodynamics here as well (this aspect of AdS/CFT in fact has been one of the inspirations for Verlinde's scenario).[56]

---

[55](Verlinde, 2011), p. 21.

[56]Verlinde (2011, pp. 20-25) sees a number of other reasons that support the idea of abandoning the

We believe that Verlinde has given us an interesting account of how gravitation could be understood as an emergent phenomenon in holographic scenarios,[57] and we think that his proposal contains ideas with potentially far-reaching ramifications. For example, if gravity is an entropic force, then, indeed, there would be no point in looking for a microscopic quantum theory of gravity, or seek gravity's unification with other microscopic forces. Furthermore, if gravity is a thermal phenomenon, one may expect fluctuations around the macroscopic equilibrium; that is, small deviations from the Einstein theory.[58]

Our analysis shows that it is the thermodynamic limit, and not holography, that produces the emergent properties in these new scenarios, as in traditional and familiar examples of emergence in physics. In particular, a novel conceptual possibility that has arisen in this investigation is that the holographic relation itself may emerge, along with gravity, in AdS/CFT-like accounts.

The question has sometimes been raised whether the technical philosophical literature on emergence and reduction is not isolationist and divorced from research problems of the scientists whose work it should relate to.[59] The case of holography in fundamental physics, and the possibilities and questions it raises for the emergence of gravity and spacetime, illustrates, however, that philosophical analysis can indeed be of direct relevance to frontier science.

References

Acuña, P. and D. Dieks, (2014). "Another look at empirical equivalence and underdetermination of theory choice", *European Journal for Philosophy of Science*, in press.

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].

Bardeen, J.M., B. Carter, and S.W. Hawking. (1973). "The four laws of black hole mechanics", *Communications in Mathematical Physics*, 31(2), 161-170.

Batterman, R.W. (2011). "Emergence, singularities and symmetry breaking", *Founda-*

---

notion of microscopic gravity theories. He points to UV/IR relations (of which AdS/CFT is only one example, another being open/closed string duality) as indications that long range gravitational forces seem to know about high energy, short distance physics. This could be a sign that gravity should not be considered as an independently defined quantum force on the micro-level. Another indication could be seen in the set of relations known collectively as 'black hole thermodynamics', which originated in the 1970s from work by Jakob Bekenstein, Stephen Hawking and others (see e.g. [Bekenstein 1973]; [Hawking et al., 1973]). Here we find relations between black hole quantities, originally defined exclusively in terms of Einsteinian gravity, that completely mimic the familiar thermodynamic laws. A final reason for the hypothesis that there is no microscopic gravity may be the general universality of gravity mentioned before.

[57]As said earlier, important early work was done by Jacobson (1995).

[58]Such deviations could of course also arise in a theory that successfully quantizes gravity, so they should not be considered a unique possibility of only the Verlinde scenario.

[59]For instance, this point came up at the Munich 2013 conference on 'Reduction and Emergence in the Sciences'.

tions of Physics, 41(6), 1031-1050.

Bekenstein, J. (1973). "Black holes and entropy", *Physical Review D*, 7(8), 2333-2346.

Belot, G., J. Earman, and L. Ruetsche (1999). "The Hawking information loss paradox: the anatomy of controversy", *The British Journal for the Philosophy of Science*, 50(2), 189-229.

Berenstein, D., and R. Cotta. (2006). "Aspects of emergent geometry in the AdS/CFT context", *Physical Review D*, 74(2), 026006. [hep-th/0605220].

Bianchi, M. (2001). "(Non-)perturbative tests of the AdS/CFT correspondence", *Nuclear Physics B. Proceedings Supplements*, 102-103, 56-64. [hep-th/0103112].

Boer, J. de, E.P. Verlinde, and H.L. Verlinde, (2000). "On the holographic renormalization group", *Journal of High Energy Physics*, 08 003. [hep-th/9912012].

Bousso, R. (2002). "The holographic principle", *Reviews of Modern Physics*, 74(3), 825-874. [hep-th/0203101].

Butterfield, J. (2011a). "Emergence, reduction and supervenience: a varied landscape", *Foundations of Physics*, 41(6), 920-959.

Butterfield, J. (2011b). "Less is different: emergence and reduction reconciled", *Foundations of Physics*, 41(6), 1065-1135.

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.

Csaki, C., H. Ooguri, Y. Oz, and J. Terning. (1999). "Glueball mass spectrum from supergravity", *Journal of High Energy Physics*, 01 017. [hep-th/9806021].

Cubrović, M., J. Zaanen, K. Schalm. (2009). "String theory, quantum phase transitions, and the emergent fermi liquid", *Science*, 325(5939), 439-444. [arXiv:0904.1993 [hep-th]].

Domènech, O., M. Montull, A. Pomarol, A. Salvio, and P.J. Silva. (2010). "Emergent gauge fields in holographic superconductors", *Journal of High Energy Physics*, 2010 033. [arXiv:1005.1776 [hep-th]].

Dongen, J. van. (2010). *Einstein's unification*. Cambridge: Cambridge University Press.

Dongen, J. van, and S. de Haro. (2004). "On black hole complementarity", *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(3), 509-525.

Drukker, N., M. Marino, and P. Putrov. (2011). "Nonperturbative aspects of ABJM theory", *Journal of High Energy Physics*, 2011 141. [arXiv:1103.4844 [hep-th]].

Fisher, M.E. (1998). "Renormalization group theory: Its basis and formulation in statistical physics", *Reviews of Modern Physics*, 70(2), 653-681.

Gaite, J.C. and D. O'Connor. (1996). "Field theory entropy, the H theorem, and the renormalization group", *Physical Review D*, 54, 5163-5173. [hep-th/9511090].

Giddings, S.B. (2011). "Is string theory a theory of quantum gravity?" *Foundations of Physics*, 43(1), 115-139. [arXiv:1105.6359 [hep-th]].

Green, M.B. (1999). "Interconnections between type II superstrings, M theory and $N=4$ supersymmetric Yang-Mills", pp. 22-96 in: Ceresole, A., C. Kounnas, D. Lüst, and S. Theisen (eds.). *Quantum aspects of gauge theories, supersymmetry and unification.* Berlin, Heidelberg: Springer. [hep-th/9903124].

Green, M.B., J.H. Schwarz and E. Witten. (1987). *Superstring theory. Vol. 1: Introduction.* Cambridge: Cambridge University Press.

Gubser, S.S., I.R. Klebanov and A.M. Polyakov. (1998). "Gauge theory correlators from noncritical string theory", *Physics Letters B*, 428(1-2), 105-114. [hep-th/9802109].

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

Hartmann, S. (2001). "Effective field theories, reductionism and scientific explanation", *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 32(2), 267-304.

Hartnoll, S.A., C.P. Herzog, and G.T. Horowitz. (2008). "Building a holographic superconductor", *Physical Review Letters*, 101(3), 031601.

Hawking, S.W. (1976). "Breakdown of predictability in gravitational collapse", *Physical Review D*, 14(10), 2460-2473.

Henningson, M. and K. Skenderis (1998). "The holographic Weyl anomaly", *Journal of High Energy Physics*, 9807 023. [hep-th/9806087].

Hertog, T. and G. T. Horowitz. (2005). "Holographic description of AdS cosmologies", *Journal of High Energy Physics*, 0504 005. [hep-th/0503071].

Hooft, G. 't. (1985). "On the quantum structure of a black hole", *Nuclear Physics*

B, 256, 727-745.

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

Hooft, G. 't. (1996). "The scattering matrix approach for the quantum black hole: an overview", *International Journal of Modern Physics A*, 11(26), 4623-4688. [gr-qc/9607022].

Hooft, G. 't. (1999). "Quantum gravity as a dissipative deterministic system", *Classical and Quantum Gravity*, 16(10), 3263-3279. [gr-qc/9903084].

Horowitz, G.T. (2005). "Spacetime in string theory", *New Journal of Physics*, 7 201. [gr-qc/0410049].

Jacobson, T. (1995). "Thermodynamics of spacetime: The Einstein equation of state", *Physical Review Letters*, 87(7), 1260-1263.

Laudan, L. and L. Leplin. (1991). "Empirical equivalence and underdetermination", *The Journal of Philosophy*, 88(9), 449-472.

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].

McGreevy, J. (2010). "Holographic duality with a view toward many-body physics", *Advances in High Energy Physics*, 723105. [arXiv:0909.0518].

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.

Seiberg, N. (2007). "Emergent spacetime", pp. 163-213 in: Gross, D., M. Henneaux, and A. Sevrin (eds.). *The quantum structure of space and time. Proceedings of the 23rd Solvay conference on physics.* Singapore: World Scientific. [hep-th/0601234].

Smolin, L. (2007). *The Trouble With Physics: The Rise of String Theory, The Fall of a Science, and What Comes Next.* New York: Mariner Books.

Stachel, J. (1993). "The other Einstein: Einstein contra field theory", *Science in Context*, 6(1), 275-290.

Susskind, L., L. Thorlacius, and J. Uglum. (1993). "The stretched horizon and black

hole complementarity", *Physical Review D*, 48(8), 3743-3761. [hep-th/9306069].

Susskind, L. (1995). "The world as a hologram", *Journal of Mathematical Physics*, 36(11), 6377-6396. [hep-th/9409089].

Susskind, L. and E. Witten. (1998). "The holographic bound in anti-de Sitter space", 10 pp. [hep-th/9805114].

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.

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]].

Witten, E. (1998a). "Anti-de Sitter space and holography", *Advances in Theoretical and Mathematical Physics*, 2, 253-291. [hep-th/9802150].

Witten, E. (1998b). "Anti-de Sitter space, thermal phase transition, and confinement in gauge theories," *Advances in Theoretical and Mathematical Physics*, 2, 505-532. [hep-th/9803131].

Zwiebach, B. (2004). *A first course in string theory.* Cambridge: Cambridge University Press.