Foundations of Physics manuscript No.
(will be inserted by the editor)

Why holography?

Chris Fields · Donald D. Hoffman · Antonino Marcianò · Chetan Prakash · Robert Prentner

Received: date / Accepted: date

Abstract The holographic principle (HP) has become a mainstay of quantum cosmology, but its relation to the rest of physics remains poorly understood. Here we show that the HP is a geometric special case of a more general principle that restricts the classical information transferred between any two subsystems of a physical system to the information that can be encoded on the boundary between them. This more general principle has its origins in the 18th century and has been formulated in a number of disciplinary contexts. We formulate this generalized holographic principle precisely, derive the HP a special case, and examine three consequences for physics: 1) that gauge invariance is a consequence of restricting classical information transfer between subsystems; 2) that environmental decoherence is holographic encoding; and 3) that compliance with the generalized holographic principle suggests a resolution of the black-hole information paradox.

C. Fields
23 Rue de Lavandières, 11160 Caunes Minervois, France
Tel.: +33-(0)6-44-20-68-69
E-mail: fieldsres@gmail.com

D. D. Hoffman
Department of Cognitive Science, University of California, Irvine, CA 92697 USA E-mail: ddhoff@uci.edu

A. Marcianò
Center for Field Theory and Particle Physics & Department of Physics, Fudan University, 200433 Shanghai, China E-mail: marciano@fudan.edu.cn

C. Prakash
Department of Mathematics, California State University, San Bernardino, CA 92407 USA
E-mail: cprakash@csusb.edu

R. Prenter
Department of Philosophy, ETH Zurich, Switzerland E-mail: robert.prentner@phil.gess.ethz.ch

Keywords Black box · Black hole information paradox · Decoherence · Decomposition · Interface theory of perception · Markov blanket · Mereological information

1 Introduction

The Holographic Principle (HP) for physical systems was first stated by 't Hooft [1]: "given any closed surface, we can represent all that happens inside it by degrees of freedom on this surface itself." The HP extends to a general, qualitative principle the area law (in Planck units) $S = A/4$, where S is the entropy and A the horizon area, derived by Bekenstein [2] for black holes. Observers external to any closed system, the HP tells us, are in a position formally analogous to that of observers outside a black hole: they can know no more about what is happening inside the system than what they can observe happening on its exterior surface.

The HP is at first sight quite counterintuitive: whatever is going on inside some bounded volume of space can, the HP tells us, be completely described given what is happening on its boundary. Suppose the bounded volume of space is a memory encoding some quantity of information: the HP tells us that the information it encodes can, as far external observers are concerned, be fully specified on its surface. This restriction of accessible information to surface degrees of freedom hugely simplifies the description of physical systems by making their purely "interior" degrees of freedom, however numerous or complex they may be, irrelevant. As 't Hooft pointed out ([1] p. 289, italics in original):

> "The fact that the total volume inside is irrelevant may be seen as a blessing, since it implies that we do not have to worry about the *metric* inside. The inside metric could be so much curved that an entire universe could be squeezed inside our closed surface, regardless how small it is. Now we see that this possibility will not add to the number of allowed states at all."

If not even the *number* of interior degrees of freedom matters to the physical description of a system, their existence is irrelevant; as 't Hooft put it, "we suspect that there simply *are* not more degrees of freedom to talk about than the ones one can draw on a surface" (p. 289, italics in original). Even if an entire universe with all of its degrees of freedom is confined within the closed system, these confined degrees of freedom are confined in principle, and can be safely ignored by external observers.

It is important to emphasize that the HP holds *only from the perspective of an external observer*. Bousso ([3], Sect. IV) provides several examples of closed volumes that in fact contain more information than can be encoded on their boundaries, showing that any observer-independent claim that "the entropy contained in any spatial region will not exceed the area of the region's boundary" fails. Rovelli [4] shows that radiating black holes must, in particular, contain more degrees of freedom than can be encoded on their horizon surfaces, concluding that "what is bound by the area of the boundary of a region is not the number of possible states in the region, but only the number of states distinguishable from observations outside the region." We provide an additional demonstration along these lines below.

Following its subsequent elaboration by Susskind [5] and the discovery by Maldecena [6] of an important special case, the duality between a string quantum gravity on a d-dimensional anti de Sitter (AdS) spacetime and a conformal quantum field theory (CFT) on its $(d-1)$-dimensional boundary, the HP rapidly became a guiding principle of quantum cosmology (for reviews see [3] [7]). The physical motivation for the HP, however, remains its motivation for 't Hooft: the observational impenetrability of a black hole that is formalized by the Bekenstein area law. While subsequent work has suggested that the HP is completely general, it remains unclear *why* it should be completely general. Hence the HP remains counterintuitive, with Bousso, for example, characterizing it as "an apparent law of physics that stands by itself, both uncontradicted and unexplained by existing theories" that "may still prove incorrect or merely accidental, signifying no deeper origin" ([3] p. 826).

Here we broaden the motivation for, and also the evidence supporting, the HP by showing that it is a geometric special case of a more general principle that restricts the classical information transferred between any two subsystems of a physical system to the information that can be encoded on the boundary between them. Informally, this generalized holographic principle (GHP) states that 1) inter-system boundaries are classical information channels and 2) there are no other classical information channels. The GHP predates 't Hooft's formulation of the HP by at least four decades, having been clearly articulated in the classical cybernetics literature of the 1950s, but both its conceptual and its formal origins are in the 18th century. The GHP cannot, therefore, be regarded as a novel or surprising principle, though its implications for physics as well as other sciences have yet to be thoroughly explored. We review three independent formulations of the GHP, from classical cybernetics [8], the theory of Markov processes [9], and perceptual psychology [10] as background before stating it more formally in Hilbert-space language, showing that it arises naturally in any state space with an associative decomposition operator, and deriving the HP as a special case. We then consider three consequences of the GHP for physics: 1) that gauge invariance follows from the GHP; 2) that environmental decoherence of a quantum system is an instance of holographic encoding; and 3) that strict compliance with the GHP resolves the black-hole information paradox (BHIP), an apparent incompatibility between unitary evolution and general covariance (for review see [11]. We conclude by suggesting that holographic encoding via the GHP captures the full meaning of the troublesome term "observation" in physics.

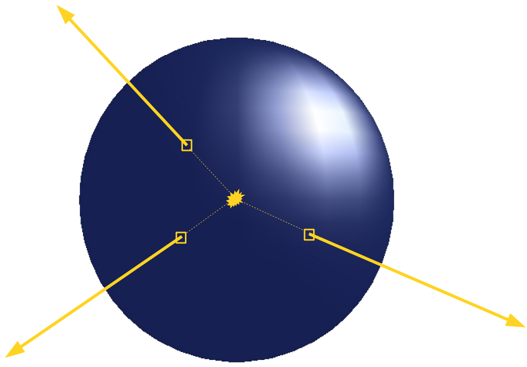

Fig. 1 Photons escape from holes in an opaque spherical shell of radius r. The maximal information obtainable by an external observer at some time t is $4\pi r^2$ in Planck units, achieved if each Planck area contains a hole.

2 Example: Photons escaping through a spherical boundary

As a simple illustration of the HP, consider an opaque, perfectly-insulating spherical shell of radius r containing at its center a point light source with intensity I. Information about photon degrees of freedom inside the shell can be obtained by an external observer only if holes are made in the shell that allow some of the photons to escape. No more information about the state at time t of the internal degrees of freedom can be obtained by an external observer than can be encoded by the photons escaping through the holes at t (Fig. 1). The dimensions of these holes are bounded by the wavelength of the escaping photons, with the Planck area l_P^2 as the lower limit. At this limit, only the presence or absence of a photon conveys information [2]. The maximal information obtainable by an exterior observer is then given by the maximum possible number of Planck-area holes, i.e. by the area $4\pi r^2$ of the sphere in Planck units.

Fixing r and hence the number $N_r \leq 4\pi r^2$ of photons emitted from the sphere, consider an observer stationed at some $R > r$. Here the available information is spread over the larger surface area $4\pi R^2$. The number of photons emitted per unit area of the observer's radius-R sphere (i.e. the intensity of the Poynting vector) falls as $1/R^2$, the same inverse-square law discovered by Newton for gravity and Coulomb for electrostatics. That the flux of any infinite-range, point-source field through a convex boundary in three dimensions satisfies this "area law" was recognized as a general principle by Gauss

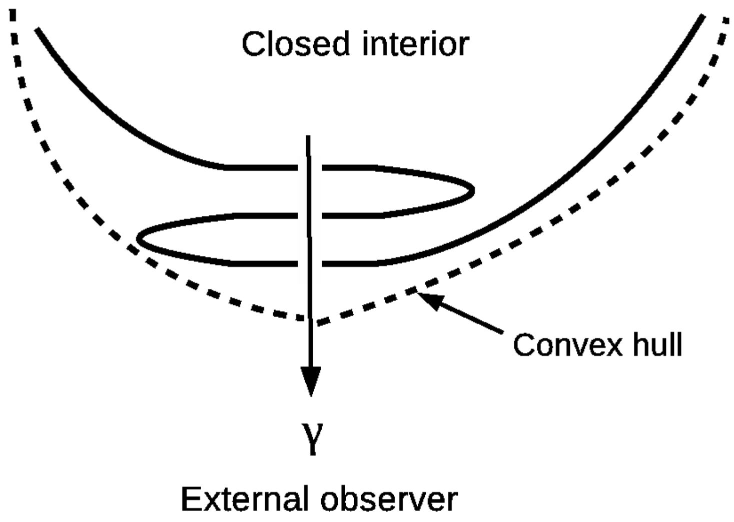

Fig. 2 Non-convex boundaries potentially lead to double-counting of information; hence HP area laws refer to convex hulls and suitably-distant observers.

[12]. Gauss did not have the idea of a finitely-pixelated boundary, so obtained non-physical infinite fluxes as $r \rightarrow 0$.

It is important to emphasize that, as in Bekenstein's formulation [2], the "area law" in this case concerns the information that can be obtained by an observer external to the enclosed volume. It does not refer to an observer-independent "god's eye" [13] measure of the number of photon degrees of freedom (and hence entropy) within the volume, though this distinction between observer-dependent and "objective" measures is not always explicitly made (cf. [3] [4]). In the example above, the spherical shell contains more photons than can be observed escaping it whenever $I > N_r$.

The above example also makes it clear that the relevant area must be convex – or alternatively, the considered observer must be located outside of a convex shell containing the system – to avoid double-counting the escaping photons and hence the encoded bits (Fig. 2). In the case of a non-convex boundary, it is the area of the convex hull of the boundary on which information transferred from inside to a suitably external observer is effectively encoded.

3 The GHP as a limit on inter-system information transfer

Principled limitations on information transfer between systems have been characterized from a number of distinct disciplinary perspectives. Before formulating the GHP precisely as a general physical principle, we examine three such characterizations, each illustrating how inter-system information transfer is restricted to information encodable on an inter-system boundary. “Information” here is *classical* information – what Bartlett, Rudolph and Spekkens [14] call *fungible* information – information that can be written down as a finite bit string and shared with another observer.

3.1 Black box formulation

The classical cybernetics literature of the 1950s introduced the engineer’s concept of a *black box* (BB) as a system for which observers have no interior access. In the canonical picture, a BB is a literal “box” on which knobs and dials, input and output registers, or some other means of bidirectional interaction are mounted. The box is sealed to prevent access to its interior, or is of such nature – e.g. is a bomb – that opening the box may be catastrophic. The interior of a BB can, therefore, be regarded as comprising whatever degrees of freedom the engineer is prevented from or chooses not to measure directly. As Ashby and others recognized, this way of defining the interior renders any physical system a BB for an engineer limited to finite observational means: “The theory of the Black Box is merely the theory of real objects or systems, when close attention is given to the question, relating object and observer, about what information comes from the object, and how it is obtained” ([8] p. 110).

The BB model formally restricts observers to finite resources and hence the exchange of finite bit strings with the BB at each unit time t. Under this restriction, Moore [15] proved that no finite number of observations of any BB is sufficient to determine the machine table, i.e. the rule generating outputs from inputs, executed by the BB. Moore’s proof is disarmingly simple, merely pointing out that any finite set of observations is consistent with an arbitrarily large number of distinct machine tables (for review see [16]). The information specifying the machine table is thus contained within the BB, but inaccessible in principle to external observers. Observers can only obtain, at each t, the finite number of bits displayed by the knobs and dials, input and output registers, or other encodings of input/output data on the “exterior” of the BB. This exterior is the complement of the interior: it comprises the degrees of freedom that the engineer *does* directly measure.

The notion of a BB as a system with behavior that cannot be fully characterized by finite measurements translates into practical terms the realization of Church [17], Turing [18] and others that formal systems exist that can execute any finitely-specified computation, and hence display any finite sequence of behaviors. Should a physical implementation of such a formal system be enclosed within a box, it would *ipso facto* be a BB. Present-day computers, with their layers of virtual machines specifically designed to render lower-lying degrees of freedom inaccessible, are precisely such physical implementations (e.g. [19] [20] [21]). No finite sample of its input-output behavior is, in particular, sufficient to determine what algorithm a general-purpose computer is

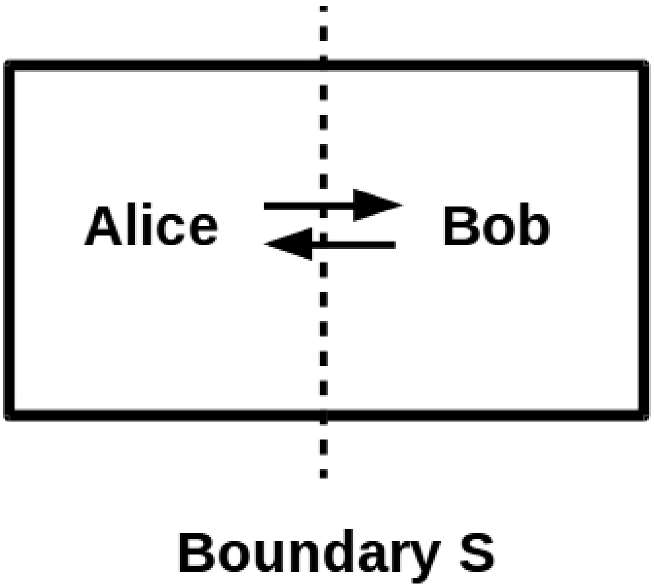

Fig. 3 An observer (Alice) interacts with a BB (Bob) by exchanging information across the surface S that separates them. The exchanged information is encoded by the degrees of freedom of the surface, e.g. the pointer positions of knobs and dials or the digits displayed by input and output registers mounted on the surface. Alice and Bob together comprise a closed system: there are no other information channels between them.

running, not just at the hardware level but at any virtual machine level. An important special case is the unsolvability of the halting problem: no finite sample of input-output behavior is sufficient to determine that a halting state has or will be reached [22].

The exchange of information between an external observer and a BB clearly satisfies the first clause of the GHP informally defined above. The only information about the BB available to the observer is the information encoded by its "exterior" degrees of freedom: the values indicated by the knobs and dials, input and output registers, or other encodings of input/output data. These exterior degrees of freedom constitute, by definition, the information channel connecting the BB to the observer; their possible values constitute the information that can be communicated through this channel. Consistent with the picture of a BB as a literal box, these channel degrees of freedom can be thought of as being displayed on – just as the knobs and dials, digital displays, etc. are physically mounted on – the surface S of the BB. The only information about the observer available to the BB is likewise encoded on this surface (Fig. 3).

The information exchange between observer and BB also satisfies the second clause of the GHP: there is no alternative or "back" information channel between them. This follows from the definition of a BB. Let ξ be any arbitrary degree of freedom. The observer either measures ξ directly or does not.

If the former, ξ is by definition part of the "exterior" of the BB and hence has values encoded on S. If the latter, ξ is, again by definition, part of the "interior" of the BB and hence inaccessible. Moore's result prevents any determination by finite observation that any such ξ, whether observed or not, is irrelevant to the dynamics and hence the possible future behavior of the BB [15]. Any informally-characterized "surrounding environment" is thus observationally indistinguishable from the BB that it surrounds; to the extent that this "environment" is not directly measured, it is part of the interior of the BB. Mere *stipulation* that some assumed "environment" is a distinct entity with its own degrees of freedom cannot physically constrain such degrees of freedom from being or interacting with the interior degrees of freedom of the BB, the identities of which are unknown by definition. This point is discussed further below in the context of decoherence.

The interaction between an observer and a BB need not, and in the classic presentations of [8] [15] does not, involve any spatial degrees of freedom. While the "surface" S of the BB is traditionally regarded as a panel on which knobs and dials are mounted and hence as itself spatial, the locations on S of the knobs and dials are irrelevant to the input or output data values that they indicate. Nothing is changed, in particular, if the input or output data values are regarded written down as a data table or encoded into any other classical data structure; their physical implementation is entirely irrelevant to their communicative function. The spatial degrees of freedom associated with S are, in other words, not part of the information channel connecting the observer to the BB; they are merely a conceptual tool, a pictorial device, for separating and hence distinguishing the input and output data values. This use of the spatial degrees of freedom of a surface to distinguish encoded data values has been carried over into present-day computer interfaces, in which the collapsibility and movability of display windows makes their spatial location entirely arbitrary. The same can be said for traditional print media. Here the temporal order of characters encodes the transferred information; the characters are printed in different locations, not on top of each other in a character-sized palimpsest, only as an aid to readability. Compliance with the GHP does not, therefore, require the spatiality assumed by the HP; any spatial degrees of freedom associated with S are a practical convenience, not a theoretical necessity.

As with Gauss' anticipation of the Bekenstein-Hawking area law, the intellectual roots of the BB model can be traced to the 18th century. The epistemic position it assigns to the observer is that described by the Empiricist philosophical tradition, the tradition claiming that knowledge of the world is obtained via the senses, with no possibility of going "beyond" the senses to learn more than they reveal [23]. Ashby and Moore further operationalized Kant's [24] insight that space and time are not characteristics of the world, but rather structures imposed on sensory data by human perceptual systems (they are "forms of intuition" for Kant). From this perspective, the surface of a black box becomes an interface that humans automatically employ to organize and interpret the perceived behavior of the world.

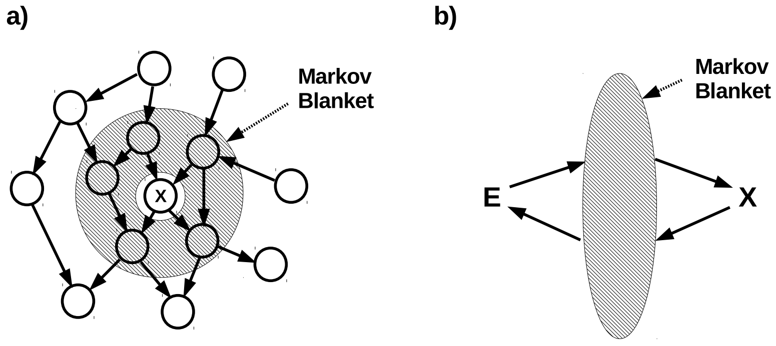

Fig. 4 a) the Markov blanket around a node X in an ergodic network comprises the parents of X, the children of X, and any parents other than X of X's children. All nodes outside the blanket constitute the "environment" E. b) The interaction between X and E is mediated by the Markov blanket between them, which renders their states conditionally independent. Adapted from [27] Fig. 2.

3.2 Markov blanket formulation

The fundamental insight of the BB model was reformulated in the language of ergodic systems by Pearl [9]. Let G be a directed graph in which nodes represent subspaces of states and arrows represent Markov processes. Pearl defines the *Markov blanket* around a node X as the set of nodes containing the parent nodes of X (nodes with arrows to X), the child nodes of X (nodes with arrows from X), and any parents other than X of X's children. All causal connections between X and the nodes "outside" the blanket must flow through the Markov blanket; hence the blanket renders the states of X and those of the "environment" E comprising these external nodes conditionally independent. Markov blankets provide natural models of processing layers in brains [25] and signal transduction in cells and multicellular systems [26], among other phenomena.

Information transfer between any nodes X and E of a graph G that are separated by a Markov blanket satisfies both clauses of the GHP: 1) any information transferred between X and E must be encoded, at the time of its transfer, by the state of the blanket, and 2) there are, by definition, no nodes or arrows outside of the blanket via which information could be transferred between X and E. As in the BB model, the "information" being transferred across the blanket is classical information. A Markov blanket thus serves the same function as the surface of a BB: it limits classical information flow between the subsystems that it separates to the classical information that it can encode. As with the surface of a BB, there is no requirement that a Markov blanket have or represent spatial degrees of freedom.

The Markov blanket formulation makes explicit two features of inter-system boundaries that the BB model leaves implicit. First, such boundaries can themselves be sophisticated information processors. This is clearly true in the case of virtual-machine layers in computers; it is also true of processing layers in neural networks [25] or cellular signal transduction pathways [26] [27] [28]. It is also clearly true of BBs in practice. Nothing says that the knobs and dials on an uncharacterized device cannot contain microchips that arbitrarily transform their inputs into their outputs; if isolated such components are, as Ashby pointed out, themselves black boxes.

Second, and more importantly for what follows, the Markov blanket formulation makes it explicit that system, environment and blanket must be regarded as separable in the technical sense of having separately-definable states if they are to be regarded as exchanging *classical* information. If, in particular, X, E and the blanket B between them are treated as quantum systems, then their joint state must be the separable $|\mathrm{XEB}\rangle = |\mathrm{X}\rangle\ |\mathrm{E}\rangle\ |\mathrm{B}\rangle$. Entanglement, therefore, cannot be a "back channel" for classical information in this formalism. Entanglement and hence separability are, however, themselves observer-relative concepts [29] [30] [31]. The only "observer" that can see all of three of X, E, and B and hence define their separate states is the "god's eye" observer, i.e. the stipulating theorist. Each of the component systems can only "see" the system(s) with which it interacts. Hence X (respectively E, B) cannot demand that E and B (respectively X and B, X and E) not be entangled, just as the observer of a BB cannot demand that its interior not be entangled with the input-output degrees of freedom on its exterior. The requirement of separability can, therefore, always be relativized to the component system regarded as "the observer." As discussed further below, the locus at which information is regarded, from some relevant perspective, as classically encoded is the only locus at which separability can be enforced.

3.3 Perceptual psychology formulation

Perceptual psychologists generally assume that evolution has tuned higher mammals, particularly humans, to perceive the world veridically under normal circumstances. Trivers, for example, argues that "our sensory systems are organized to give us a detailed and accurate view of reality, exactly as we would expect if truth about the outside world helps us to navigate it more effectively" ([32] p. xxvi). Pizlo, Sawada and Steinman similarly claim that "veridicality is an essential characteristic of perception and cognition. It is absolutely essential. *Perception and cognition without veridicality would be like physics without the conservation laws*" ([33] p. 227; emphasis in original). Hoffman, Singh and Prakash review both empirical evidence and theoretical arguments demonstrating that veridical perception is a highly-improbable outcome of any evolutionary process [10]. The "interface theory of perception" (ITP) that they develop as an alternative postulates instead that perceptual experience functions as an interface – much like the user interface of a computer – that encodes the fitness consequences of actions while hiding the states and dynamics that produce those consequences (see also [34] [35]).

The function of the perceptual interface, as it is defined within ITP, is to encode information transferred from the perceiver's world to the perceiver; it thus satisfies the first clause of the GHP. As in the case of a BB or a Markov blanket, the interface is by definition the only information channel from the world to the perceiver; hence it also satisfies the second clause. The only foundational assumption made about this interface is that it is a measurable space and hence supports probability distributions over its possible states. Computations can be built into the interface to support short- and long-term memory and the recognition of distinct "objects" as correlated sets of fitness consequences [36]; however, the existence of such correlations imposes no structural requirements on the perceiver's world beyond a lower limit on its number of degrees of freedom [10]. As in both the BB and Markov blanket formulations, no requirement that the interface has or encodes spatial degrees of freedom is imposed; we have suggested elsewhere that the role of spatial degrees of freedom in ITP is to provide an error correction mechanism for the encoded fitness consequences [37].

4 Formal statement of the GHP

4.1 Hilbert space formulation

Let $\mathbf{A}$ and $\mathbf{B}$ be distinct, non-overlapping physical systems, the possible states of which compose Hilbert spaces $\mathcal{H}_{\mathbf{A}}$ and $\mathcal{H}_{\mathbf{B}}$, $\mathcal{H}_{\mathbf{A}} \cap \mathcal{H}_{\mathbf{B}} = \emptyset$, and assume the joint system $\mathbf{S} = \mathbf{A} \oplus \mathbf{B}$ with states in $\mathcal{H}_{\mathbf{S}} = \mathcal{H}_{\mathbf{A}} \otimes \mathcal{H}_{\mathbf{B}}$ is closed. Let $H_{\mathbf{A}}$ and $H_{\mathbf{B}}$ be Hamiltonian operators acting separately on $\mathbf{A}$ and $\mathbf{B}$ respectively; in this case the Hamiltonian $H_{\mathbf{S}} = H_{\mathbf{A}} \otimes I_{\mathbf{B}} + I_{\mathbf{A}} \otimes H_{\mathbf{B}} + H_{\mathbf{AB}}$, where $I_{\mathbf{A}}$ and $I_{\mathbf{B}}$ are the identity operators on $\mathbf{A}$ and $\mathbf{B}$ respectively and $H_{\mathbf{AB}}$ represents the $\mathbf{A} - \mathbf{B}$ interaction. Consistently with $\mathbf{S}$ being closed, we assume that the time-evolution operator $e^{-(i/\hbar)H_{\mathbf{S}}t}$ is unitary, where t is the time dimension associated with $\mathbf{S}$. The *generalized holographic principle* is then:

> **GHP**: No more classical information can be transferred between two distinct, jointly-closed physical systems $\mathbf{A}$ and $\mathbf{B}$ than can be encoded by the eigenvalues of their interaction Hamiltonian $H_{\mathbf{AB}}$.

The GHP is not a statement about the "objective" classical correlation between $\mathbf{A}$ and $\mathbf{B}$ as observed by some hypothetical "god's eye" observer external to $\mathbf{S}$, nor does it concern the entanglement between $\mathbf{A}$ and $\mathbf{B}$ produced by the unitary evolution of $\mathbf{S}$, which is similarly only observable by an observer external to $\mathbf{S}$. It is rather a statement about classical information transfer within the closed system $\mathbf{S}$, i.e. about the "observational outcomes" $\mathbf{A}$ can obtain from **B** or vice versa. If the eigenvalues of $H_{\mathbf{AB}}$ are continuous, the GHP places no limit on classical information transfer between **A** and **B**; the GHP becomes interesting only if $H_{\mathbf{AB}}$ has a finite number of eigenvalues, each of which is defined with finite precision. In this case, the eigenvalues can be specified, assuming some complete basis and an ordering, by a finite sequence of bits. The GHP is, therefore, an *intrinsically quantum* principle: it is significant only when the $\mathbf{A} - \mathbf{B}$ interaction is both finite and discrete. As all physically-realizable detectors are limited to finite resolution, this seemingly-limited case is the only one of interest in practice; it is analogous to the restriction to finite observational resources found in classical cybernetics. It will, therefore, be assumed in what follows.

The GHP effectively defines the interaction Hamiltonian $H_{\mathbf{AB}}$ as the exclusive classical information channel between the distinct systems **A** and **B**. This channel can be represented as an N-dimensional discrete space, where N is the number of bits in a bit-sequence representation of the eigenvalues of $H_{\mathbf{AB}}$. As the $\mathbf{A} - \mathbf{B}$ information channel is the locus of the $\mathbf{A} - \mathbf{B}$ interaction, it can also be considered the "boundary" between **A** and **B**. It is then naturally interpreted as the subspace of $\mathcal{H}_{\mathbf{S}}$ on which $H_{\mathbf{AB}}$ is non-zero, which has dimensionality smaller than that of $\mathcal{H}_{\mathbf{S}}$ provided $H_{\mathbf{AB}} \neq H_{\mathbf{S}}$. The encoding of information on this boundary is, in this case, "holographic" in a natural sense. Thinking of the boundary as a discrete array of bits makes the "encoding" function of the boundary explicit. Provided **A** and **B** are distinct as required, it is clear from the definition of $H_{\mathbf{AB}}$ that the $\mathbf{A} - \mathbf{B}$ boundary is the only classical information channel between **A** and **B**. Hence the statement above formalizes both clauses of the informal GHP exemplified by the BB model, the Markov blanket formalism, and ITP.

4.2 The GHP forbids inter-system transfer of mereological information

The interaction $H_{\mathbf{AB}}$ is well-defined only if it is invariant under further decompositions of either **A** or **B** into subsystems, an invariance that is guaranteed, for Hilbert spaces, by the associativity of the tensor product operator $\otimes$. The eigenvalues of $H_{\mathbf{AB}}$ must similarly be invariant under decompositions of either **A** or **B** into subsystems. These eigenvalues cannot, therefore, encode information that depends on a specific decomposition of either **A** or **B** into subsystems; they cannot, in other words, encode mereological information, i.e. information specifying part-whole relationships, about either **A** or **B**. The GHP therefore forbids any transfer of mereological information, in particular any transfer of information specifying decomposition into subsystems or interactions between subsystems of either **A** or **B** across the $\mathbf{A} - \mathbf{B}$ boundary. The GHP thus embodies the linearity requirement, for any observable O, decomposition $\mathbf{A}_1 \otimes \mathbf{A}_2 = \mathbf{A}$, and basis transformation $[c_{ij}^{i'j'}] : \mathbf{A}_1 \otimes \mathbf{A}_2 \rightarrow \mathbf{A'}_1 \otimes \mathbf{A'}_2 = \mathbf{A}$, that:

$$\begin{aligned} < O >_{\mathbf{A}} &= \mathrm{Tr}[\rho_{\mathbf{A}} O] = \rho^{ijkjl} \langle \mathbf{A}_{1i}\mathbf{A}_{2j}|O|\mathbf{A}_{1k}\mathbf{A}_{2l}\rangle \\ &= \rho^{ijkjl}[c_{ij}^{*i'j'}][c_{kl}^{k'l'}]\langle \mathbf{A'}_{1i'}\mathbf{A'}_{2j'}|O|\mathbf{A'}_{1k'}\mathbf{A'}_{2l'}\rangle \,. \end{aligned} \tag{1}$$

Similarly, for any basis transformation $[d_{ij}^{i'j'}] : \mathbf{B}_1 \otimes \mathbf{B}_2 \rightarrow \mathbf{B'}_1 \otimes \mathbf{B'}_2 = \mathbf{B}$, it must hold that

$$\begin{aligned} < O >_{\mathbf{B}} &= \mathrm{Tr}[\rho_{\mathbf{B}}\, O] = \rho^{ijkjl} \langle \mathbf{B}_{1i}\mathbf{B}_{2j}|O|\mathbf{B}_{1k}\mathbf{B}_{2l}\rangle \\ &= \rho^{ijkjl}[d_{ij}^{*i'j'}][d_{kl}^{k'l'}]\langle \mathbf{B'}_{1i'}\mathbf{B'}_{2j'}|O|\mathbf{B'}_{1k'}\mathbf{B'}_{2l'}\rangle \,. \end{aligned} \tag{2}$$

The irrelevance of mereological or internal-interaction information enforced by the GHP formalizes the inaccessibility of internal "parts" or "mechanisms" familiar from the classical BB model; it similarly formalizes the "shielding" of information by Markov blankets and inaccessibility of structural or dynamical information about the "world" found in ITP. The GHP thus imposes a powerful symmetry on any physical description based solely on the information available on an inter-system boundary: no such description can depend on subsystem separability, and hence subsystem identity, on either side of the boundary [16]. If the interaction between **A** and **B** satisfies the GHP, no description of the interaction can require either **A** or **B** to occupy a separable state. If **A** is regarded as "observing" **B**, in particular, **A**'s description of **B** cannot require that **B**'s state is separable. Here again the intrinsically quantum nature of the GHP becomes clear: by generating entangled states, unitary evolution effectively erases subsystem identity and hence mereological information within both **A** and **B**. Such information cannot, therefore, be considered observer-independent in any quantum-theoretic description.

4.3 The HP is a special case of the GHP

As in its exemplars reviewed above, no assumption of spatial degrees of freedom has been made in stating the GHP or in characterizing the inter-system boundary that it implicitly defines. There are, in particular, no spatial degrees of freedom (or any other degrees of freedom) "between" **A** and **B** on which their interaction can depend; **A** and **B** are "separated" only by the notional decompositional boundary at which $H_{\mathbf{AB}}$ is defined. This interaction is, moreover, confined by definition to this decompositional boundary; it cannot depend on any spatial degrees of freedom purely within the "bulk" of either **A** or **B**. We can, therefore, assume without loss of generality that $\mathcal{H}_{\mathbf{S}}$ includes no spatial degrees of freedom. If, however, we add d ancillary spatial dimensions on which $H_{\mathbf{S}}$ and therefore $H_{\mathbf{AB}}$ do not depend to both $\mathcal{H}_{\mathbf{A}}$ and $\mathcal{H}_{\mathbf{B}}$, and hence to $\mathcal{H}_{\mathbf{S}}$, an ancillary, compact $(d-1)$-dimensional spatial structure is induced on the boundary. Adding such ancillary spatial degrees of freedom is, to continue the analogy to the classical BB, attaching the input and output devices of the BB to an instrument panel that provides them with a spatial

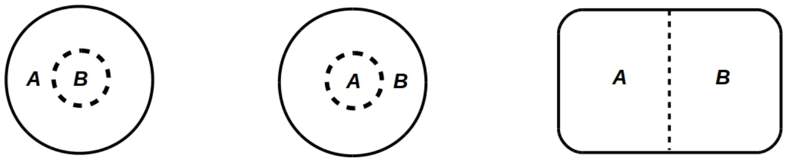

Fig. 5 Three ways of adding spatial dimensionality to the **A** – **B** boundary (dashed line). As $\mathbf{S} = \mathbf{A} \oplus \mathbf{B}$ is by assumption closed, these are equivalent.

layout. As **S** is by assumption closed, the spatial boundary of **S**, including the boundary between **A** and **B** formed by these ancillary spatial dimensions, must also be closed (Fig. 5). As the **A** – **B** boundary must still function as a classical information channel, we can envision it as "encoding" the transmitted information or as has having "punctures" through which the transmitted bits can flow.

The HP as originally formulated by 't Hooft follows immediately in this setting. Consider **B** to be the "system of interest" and **A** to be the observer. From **A**'s perspective, **B** has a closed spatial boundary. All information about **B** accessible to **A** is encoded on this boundary by $H_{\mathbf{AB}}$; the "degrees of freedom of **B**" to which **A** has access are those represented by the eigenvalues of $H_{\mathbf{AB}}$. These eigenvalues tell **A** nothing about **B**'s internal "bulk" structure or about the "bulk" interaction $H_{\mathbf{B}}$; from **A**'s external perspective, the internal degrees of freedom of **B** may as well not exist. In particular, **A** can determine nothing about the metric structure inside **B**, or even whether such a metric structure exists.

The physical meaning of the Bekenstein-Hawking area law also becomes clear in this setting. The physics represented by $H_{\mathbf{AB}}$ is by assumption independent of the ancillary spatial dimensions added to the **A** – **B** boundary; therefore the spatial locations on the boundary at which eigenvalues are encoded are arbitrary. The amount of information encoded on the boundary remains invariant under any shuffling of eigenvalue-encoding locations. No spatial property of the boundary except its overall area can, therefore, be correlated with its information content. Its overall area, on the other hand, must be correlated with its information content if the ancillary spatial dimensions are independent of the encoded eigenvalues, i.e. independent of $H_{\mathbf{AB}}$. The area law for a sphere follows, as Bekenstein originally derived it, from Gauss' theorem together with the assumption of a minimal area l_P^2 to encode each nat (i.e. 1/ln2 bit) of information.

5 Consequences of the GHP for physics

5.1 The GHP implies gauge invariance

In §4.2, we employed coordinate transformations (Eq. (1), (2)) to represent decompositional changes. Elaborating more on this, we may extend the invariance under choice of bases to develop the concept of invariance of extended systems (i.e. fields) under gauge transformations. As an exploratory example, we consider two physical systems, **A** representing bosonic particles, in our case photons, and **B** representing fermionic particles, which can exchange information only through the interaction Hamiltonian.

If we disregard the gravitational interaction, mathematically the system **A** can be described by a real vector field $A_\mu(x)$ in the four dimensional Minkowski space-time $\mathcal{M}_4$ that satisfies the covariant Maxwell equations[1]. The system **B** is instead described by a Dirac field ψ, which is an element of $\mathbb{C}^4$ that carries internal spinorial indices and transforms under the fundamental representation of $SU_L(2) \times SU_R(2)$. The invariance with respect to the choice of basis can be generalized to the local gauge freedom for the choice of the field A_μ. This is defined for the abelian vector field under scrutiny by

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) - \partial_\mu \lambda(x)\,, \tag{3}$$

where $\lambda(x)$ denotes a scalar function that is C^1, i.e. continuous with its first derivatives. Eq. (3) introduces a redundancy in the description of the system that does not affect the physics. In absence of sources, the Maxwell theory specifies the meaning of the latter statement: picking Cartesian coordinates and denoting $A_\mu = \{\phi, \mathbf{A}\}$, the electric field[2] reads $\mathbf{E} = \dot{\mathbf{A}} - \nabla\phi$ while the magnetic field is $\mathbf{B} = \nabla \wedge \mathbf{A}$. These are the observable fields of the system **A**, which are gauge invariant under Eq. (3).

A gauge fixing for the vector field A_μ of the system **A** can be thought as a choice of observers, in analogy to the choice of basis for the system considered in §4.2. In the language of field theory, the independence of the system's description under the choice of observers corresponds to an invariance under gauge transformations of the system of photons and fermions. Therefore we argue that in addition to forbidding the transfer of mereological information, the GHP also leads naturally to the principle of gauge invariance for the two systems considered. Even more, we can show that the application of the GHP to this system of fields dictates the form of their interaction. This is the greatest success of the principle of gauge symmetry.

---

[1] The gauge $A_\mu(x)$ is endowed with a U(1) symmetry. In general, we can think at the gauge group as being provided by a hypercharge sector. Alternatively, we may directly focus on the specific example of QED, and imagine that the hypercharge U(1) is mixed with the third component of the isospin symmetry group SU(2). In the latter case, the U(1) we are dealing with represents the electromagnetic sector. As customary, greek letters μ, ν label space-time indices, while with $x := x_\mu$ we denote a chart of coordinates on $\mathcal{M}_4$.

[2] With dots we denote time derivatives.

We start by writing the partition function for the system **A**, namely the U(1) gauge sector.

$$\mathcal{Z}_{\mathbf{A}}[A] = \int \mathcal{D}A_\mu \, e^{\imath \mathcal{S}(A)} \,, \tag{4}$$

in which $\mathcal{D}A_\mu$ denotes the path integral measure over the copies of the gauge field and $\mathcal{S}(A)$ stands for the classical action. The GHP now implies that for different observers, related by different choices of the gauge fixing, the expectation value in the path integral of any functional observable $\mathcal{O}[A]$ is invariant under gauge transformation, and by choices of gauge fixing. The first statement is easily recovered, if we factor out the redundancy due to the gauge transformations, by fixing the gauge functional G and then imposing gauge invariance. Mathematically the statement reads

$$1 = \int \mathcal{D}\lambda \delta(G(A^\lambda)) \left| \det \frac{\delta G(A^\lambda)}{\delta \lambda(x)} \right| \,, \tag{5}$$

in which we have denoted $A^\lambda_\mu(x) = A_\mu(x) + \partial_\mu \lambda(x)$. The Lorentz functional $G(A) = \partial_\mu A^\mu$ provides the simplest choice of gauge functional we can resort to in order to implement gauge invariance of the path integral. Under gauge transformations

$$G(A) = \partial_\mu A^\mu \quad \to \quad G(A^\lambda) = \partial_\mu A^\mu + \Box \lambda \,. \tag{6}$$

The invariance of the path-integral is then recovered

$$\begin{aligned} \mathcal{Z}_{\mathbf{A}}[A] &= \mathcal{N} \int \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A)} \int \mathcal{D}\lambda \delta(G(A^\lambda)) \left| \det \frac{\delta G(A^\lambda)}{\delta \lambda(x)} \right| \\ &= \mathcal{N} \left| \det \Box \right| \int \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A)} \int \mathcal{D}\lambda \delta(G(A^\lambda)) \\ &= \mathcal{N}' \int \mathcal{D}\lambda \, \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A)} \, \delta(G(A)) \\ &= \mathcal{N}'' \int \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A)} \, \delta(G(A)) \,, \end{aligned} \tag{7}$$

in which $\mathcal{N}$, $\mathcal{N}'$ and $\mathcal{N}''$ are normalization functions that can be neglected.

We may parametrize the gauge condition employing an arbitrary function f, so that

$$G_f(A) = \partial_\mu A^\mu - f \tag{8}$$

represents a family of gauge-fixing terms to be considered. Since no physical observables must depend on the gauge fixing, we can further average the different gauge fixing terms by means of the factor $\exp(-\frac{\imath}{2\xi} \int d^4x f^2(x))$, with ξ a positive parameter, and an integration over f. The expression for the path integral in Eq. (7) then becomes

$$\begin{aligned} \mathcal{Z}_{\mathbf{A}}[A] &= \mathcal{N} \int \mathcal{D}f \, \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A) - \frac{\imath}{2\xi} \int d^4x \, f^2(x)} \\ &= \mathcal{N} \int \mathcal{D}A_\mu \, e^{\imath \mathcal{S}_{\mathrm{A}}(A) - \frac{\imath}{2\xi} \int d^4x (\partial_\mu A^\mu(x))^2} \,. \end{aligned} \tag{9}$$

Invariance under different choices of the gauge fixing condition, i.e. different choices of f, extends the result found in Eq. (1), as it can now be recast for the expectation value of the observables $\mathcal{O}$,

$$\begin{aligned}\langle\mathcal{O}[A]\rangle_f &= \mathcal{N}\int \mathcal{D}f\,\mathcal{D}A_\mu\, e^{\imath\mathcal{S}_{\rm A}(A)-\frac{\imath}{2\xi}\int d^4x\, f^2(x)}\,\mathcal{O}[A]\\ &= \mathcal{N}\int \mathcal{D}A_\mu\, e^{\imath\mathcal{S}_{\rm A}(A)-\frac{\imath}{2\xi}\int d^4x(\partial_\mu A^\mu(x))^2}\,\mathcal{O}[A]\\ &= \mathcal{N}\int \mathcal{D}g\,\mathcal{D}A_\mu\, e^{\imath\mathcal{S}_{\rm A}(A)-\frac{\imath}{2\xi}\int d^4x\, g^2(x)}\,\mathcal{O}[A]\\ &= \langle\mathcal{O}[A]\rangle_g\,, \end{aligned}\tag{10}$$

having denoted with f and g two different gauge-fixings.

We now consider the system **B**, which is composed by Dirac fields. For simplicity, we restrict our focus to only one fermionic species ψ, and write the path integral formulation of the system, i.e.

$$\mathcal{Z}_{\mathbf{B}}[\psi,\bar\psi] = \int \mathcal{D}\psi\,\mathcal{D}\bar\psi\, e^{\imath\mathcal{S}_{\rm D}(\psi,\bar\psi)}\,, \tag{11}$$

in which

$$\mathcal{S}_{\rm D}(\psi,\bar\psi) = \int d^4x\,\bar\psi\,(\imath\gamma^\mu\partial_\mu - m)\,\psi\,. \tag{12}$$

Observable quantities, which we denote with $\mathcal{O}[\psi,\bar\psi]$, are bilinear in the fermionic fields ψ and $\bar\psi$,

$$\mathcal{O}[\psi,\bar\psi] = \bar\psi\, O(\Gamma_I)\psi\,,$$

in which the matrix O, for which spinorial indices are suppressed, depends on the elements of the Clifford algebra Γ_I, with $I = 1\dots16$. A local gauge transformation that does not change the values of $\mathcal{O}$ can be introduced:

$$\psi(x)\to\psi'(x) = e^{\imath q\lambda(x)}\psi(x)\,,\qquad \bar\psi(x)\to\bar\psi'(x) = e^{-\imath q\lambda(x)}\bar\psi(x)\,, \tag{13}$$

with q a charge parameter. The generator of the transformation, which is U(1), clearly commutes with the matrix O, and invariance of any observable $\mathcal{O}$ under Eq. (13) is immediately recovered. Selecting an infinitesimal global transformation, for instance by $\lambda(x) = \alpha\in\mathbb{R}$, the conserved charge

$$Q = \int_\Sigma d^3x\,\psi^\dagger\psi\,,$$

with Σ a spatial hypersurface, can be introduced. This generates U(1) transformations $U = e^{\imath\alpha Q}$ acting on the Hilbert space of the theory.

Notice however that the dynamics of the system **B** as specified by the Dirac Lagrangian in Eq. (12) is not yet gauge invariant. The introduction of a covariant derivative

$$\partial_\mu \to D_\mu = \partial_\mu - \imath q A_\mu$$

is required, which in turn introduces the interaction Lagrangian density

$$\mathcal{L}_{\mathbf{AB}} := \mathcal{L}_{\text{int}} = e\, A_\mu\, \bar{\psi}\gamma^\mu\psi\,. \tag{14}$$

At the level of the path integral formulation, the theory for the joint system $\mathbf{S} = \mathbf{A} \oplus \mathbf{B}$ is defined by

$$\mathcal{Z}_{\mathbf{S}}[A, \psi, \bar{\psi}] = \int \mathcal{D}A_\mu\, \mathcal{D}\psi\, \mathcal{D}\bar{\psi}\, e^{\imath \mathcal{S}_{\mathrm{A}}(A) + \mathcal{S}_{\mathrm{D}}(\psi,\bar{\psi}) + \mathcal{S}_{\mathrm{int}}(A,\psi,\bar{\psi})}\,. \tag{15}$$

Averaging over the spatial hypersurface, we obtain the interaction Hamiltonian

$$H_{\mathbf{AB}} = -\,e \int_{\Sigma} d^3x\, A_\mu\, \bar{\psi}\gamma^\mu\psi, \tag{16}$$

through which the two systems **A** and **B** exchange quanta and vary their internal particle numbers. Once Eq. (16) is taken into account, the Hamiltonian $H_{\mathbf{S}} = H_{\mathbf{A}} \otimes I_{\mathbf{B}} + I_{\mathbf{A}} \otimes H_{\mathbf{B}} + H_{\mathbf{AB}}$ becomes gauge invariant. Naturally, the only possible $\mathbf{A} - \mathbf{B}$ interaction is through the boundary defined by Eqs. (14)-(16). But the source of this interaction is just the necessity to hold the two systems **A** and **B** gauge invariant. The eigenvalues of $H_{\mathbf{AB}}$ must be invariant under gauge transformations acting on the fields, whether we are focusing on the system **A** or on the system **B**. Mutatis mutandis, a similar statement as the one enunciated in §4.2 concerning the transfer of mereological information can be put forward: the eigenvalues of $H_{\mathbf{AB}}$ cannot encode information that depends on a specific gauge fixing of either the system **A** or the system **B**. The requirement of gauge invariance can, therefore, be seen as equivalent to the requirement of invariance under changes in subsystem decomposition. A gauge invariant field cannot encode information specifying decompositional boundaries.

5.2 Decoherence = holography

Since its introduction by Zeh [38] [39] [40] and Zurek [41] [42], environmental decoherence has been widely, though not universally, regarded as explaining the "collapse" of quantum states into effectively classical states. The mechanism of decoherence combines entanglement with selective observation: entangling a system of interest **X** with an environment **E** removes quantum coherence from ("decoheres") **X** for an observer **O** that measures the state of **X** but not the state of **E** (Fig. 6a; see [43] [44] for reviews). Environmental decoherence may also be regarded, in the "environment as witness" formulation, as redundantly encoding the eigenvalues of the $\mathbf{X} - \mathbf{E}$ interaction $H_{\mathbf{XE}}$ into the state of **E** for any **O** that measures the state of **E** but not the state of **X** (Fig. 6b; [45] [46]). Such "state broadcasting" has been offered as an explanation of the classical appearance of the macroscopic world [47] [48] [49].

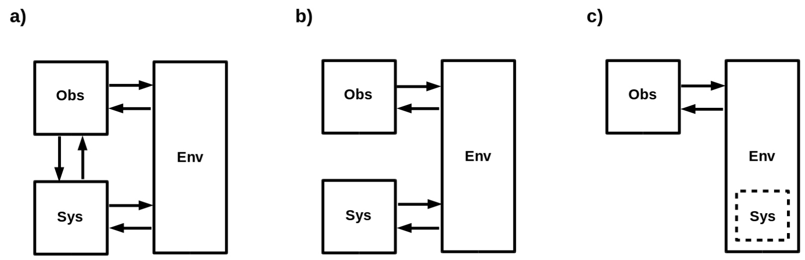

Fig. 6 Environmental decoherence in a) the original formulation of [38], b) the environment as witness formulation of [45], and c) a single-boundary formulation consistent with the GHP. Adapted from [37] Fig. 2.

From a more practical perspective, environmental decoherence poses a major design obstacle to quantum computing[3].

Both standard formulations of environmental decoherence, however, require the transfer of mereological information across inter-system boundaries and hence violate the GHP. In Zeh's original formulation (see [54] for a recent and explicit example), **O** interacts separately with both **X** and **E**, and is able to determine from this interaction (or stipulate a priori) both the location of the decompositional boundary between **X** and **E** and their interaction $H_{\mathbf{XE}}$. In the environment as witness formulation, **O** interacts only with **E**, and this interaction transfers information about both the decompositional boundary between **X** and **E** and their interaction $H_{\mathbf{XE}}$. As seen in §4.2, however, the GHP forbids any physical interaction from transferring such information. The eigenvalues of $H_{\mathbf{OE}}$, in particular, cannot specifically encode information about the eigenvalues of $H_{\mathbf{XE}}$ without violating the fundamental assumption that the relevant Hilbert spaces are associatively decomposable.

Consistency with the GHP requires a formulation of environmental decoherence in which all available information is encoded by the eigenvalues of one cross-boundary interaction and no cross-boundary transmission of mereological information is assumed (Fig. 6c). The most natural such formulation embeds **X** into **E** and considers information transfer across the **O** – **E** boundary; embedding **O** into **E** and considering information transfer across the **X** – **E** boundary is also possible, and corresponds to observation by a "witnessing" environment. Treating decoherence at the **O** – **E** boundary makes explicit the fact that **X** and **E** must be entangled from **O**'s perspective for decoherence to occur. As in the environment as witness formulation, an effective or apparent

[3] Environmental decoherence as a physical mechanism is to be distinguished from more abstract conceptions of decoherence between Everett branches [50] [51] or histories of measurements [52] [53].

state of **X** may be deduced by **O** if, but only if, it is encoded by the eigenvalues of $H_{\mathbf{EO}}$. As mereological information cannot be transmitted by $H_{\mathbf{EO}}$, any such effective state information must be arbitrarily ambiguous about both the location of the boundary of **X** within **E** and the eigenvalues of $H_{\mathbf{XE}}$.

When decoherence is reconceptualized in this way, it becomes clear that the "encoding" that results – what Zurek [41] [42] termed "einselection" – is holographic encoding on the single specified (e.g. **O** – **E**) boundary. Suppose **O** interacts with **E** by deploying n binary-outcome measurements M_i with an average energetic cost per outcome bit recorded of ck_BT, $c > \ln 2$, with k_B Boltzmann's constant and T tempertaure. The **O** – **E** interaction can then be written:

$$(1/ck_BT)H_{OE}(t) = \sum_i^n \alpha_i(t)M_i \tag{17}$$

where the functions $\alpha_i(t), \sum_i^n \alpha_i(t) = 1$ at each t, determine which measurement is performed at each t. Choosing the M_i is, in effect, choosing a gauge or equivalently by the reasoning above, a (stipulated) decomposition of **E** into "observed" or "apparent" systems. An apparent system $\mathcal{X}$ can, in this case, be considered to be a subset of $m < n$ outcome values obtained by acting on **E** with the M_i. For $\mathcal{X}$ to be re-identifiable at multiple observation times, a subset of $m^{(id)} < m$ of these outcome values must be fixed as "identifying criteria" of $\mathcal{X}$; the remaining $m - m^{(id)}$ values can vary with time and hence indicate "degrees of freedom" of $\mathcal{X}$. The apparent "environment" $\bar{\mathcal{X}}$ of $\mathcal{X}$ is everything else observable, i.e. the complementary subset of $n - m$ outcome values. *Decoherence* of $\mathcal{X}$ by $\bar{\mathcal{X}}$ corresponds, in this case, to $\mathcal{X}$ and $\bar{\mathcal{X}}$ being observationally distinguishable, i.e. to the $m^{(id)}$ outcome values that serve to identify $\mathcal{X}$ being encoded by H_{OW}. Coherence between $\mathcal{X}$ and $\bar{\mathcal{X}}$ corresponds to observational indistinguishability of $\mathcal{X}$ and $\bar{\mathcal{X}}$, as is required if their joint state is non-separable.

Reconceptualizing decoherence in this way relocates the "quantum-to-classical transition" from the objective world to the merely-notional decompositional boundary between **O** and **E**. "Systems" as well as their "states" become entirely relative to the choice of measurement operators M_i, where in this case the M_i are themselves only defined at the **O** – **E** decompositional boundary. Strict compliance with the GHP, in other words, reformulates physics in entirely device-independent, operational terms, as has already been suggested by Grinbaum [55].

5.3 Symmetry across the horizon boundary resolves the BHIP

The BHIP was introduced by Hawking [56], who argued that under intuitively-reasonable assumptions evaporating black holes (BH) destroy information and hence violate unitarity. The HP was originally formulated, in part, to address this problem by showing that the "missing" information could be found on the surface of the horizon. Since the work of Susskind and Thorlacius [57], much discussion of the BHIP has focused on the relationship between observations made by asymptotic observers of Hawking radiation from a BH and those made by an observer falling freely into the same BH. The possibility that the same quantum-state information could be obtained by both observers, thus violating the no-cloning theorem and hence unitarity, was sharpened by Almheiri, Marolf, Polchinski and Sully [58], who argued that this possibility exists for a single observer, the infalling one.

We show here that strict compliance with the GHP imposes a symmetry that resolves the BHIP; indeed the BHIP can be seen as a consequence of not taking sufficiently seriously 't Hooft's remark that nothing can be known about the metric inside a BH. Let S be the horizon surface of a sufficiently large BH and consider two asymptotic observers, Alice and Bob, stationed outside and inside the BH respectively. Assume that Alice and Bob are each equipped with quantum reference frames (i.e. local physical systems encoding non-fungible reference information [14]) for mass, charge and angular momentum. Strictly speaking, the GHP requires us to view these reference frames as internal to Alice and Bob, respectively [16] and indeed to view S as the Alice-Bob boundary as discussed above; we will, however, continue to use the conventional language in what follows.

The Bekenstein-Hawking area law fixes the entropy and hence coding capacity of S to $A/(4 \ln 2)$ bits. Alice on the outside interprets $A/(4 \ln 2)$ as the (classical) instantaneous information content of the BH and the specific bit pattern as encoding the instantaneous state of the BH. Similarly, Bob on the inside interprets $A/(4 \ln 2)$ as the (classical) instantaneous information content of the outside (i.e. Alice's) universe and the specific bit pattern as encoding its instantaneous state. Alice and Bob have, in other words, complementary interpretations of the single data structure S: Alice sees S as the horizon surface of a BH, and so does Bob. As 't Hooft emphasized, the metric $g_{\mu\nu}$(inside) inside the BH is unobservable by and cannot matter to Alice; similarly $g_{\mu\nu}$(outside) is unobservable by and cannot matter to Bob. Neither Alice nor Bob can, in particular, require that the geometry on the other side of S be a Schwartzschild singularity, even though an analytic continuation of the geometry observed on their side of the horizon may indicate this. Bob, in particular, is in the vicinity of a singularity only according to Alice's analytic continuation of the behavior of her outside metric; from Bob's internal perspective, it is Alice who is in the vicinity of a singularity. The geometry of the situation, can, therefore, be represented as completely symmetrical (Fig. 7). This symmetric geometry corresponds to the "exterior" components of two Schwartzschild solutions for which the respective horizons have been identified; it differs from standard wormhole geometries by having one horizon, not two. This construction confirms Rovelli's result that BHs contain more information than can be encoded on their horizon surfaces [4]; here the BH observed by Bob contains all of the information in Alice's universe and vice-versa.

In the symmetric geometry of Fig. 7, covariance requires that both asymptotic observers be regarded as making equivalent observations. Alice observes

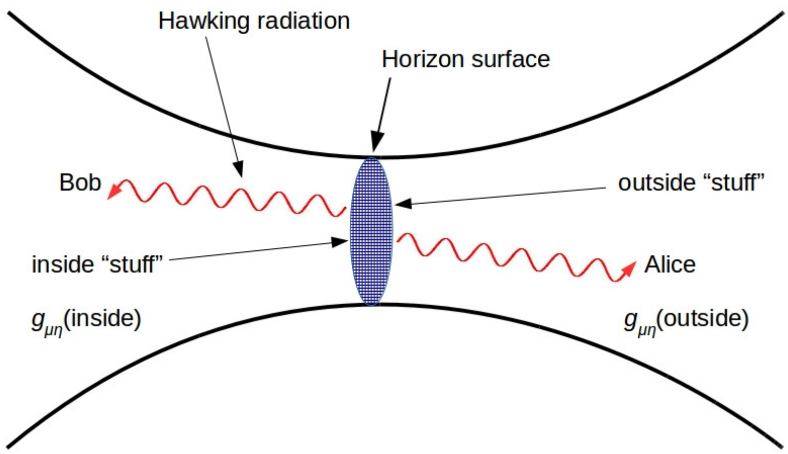

Fig. 7 Symmetric representation of observations of a horizon surface S made by asymptotic observers on either side of S. "Stuff" (matter) falls into the horizon from both sides. The S-matrix from infalling matter to outgoing radiation can be viewed as acting from one side of the horizon to the other. The GHP forbids either observer knowing the S-matrix.

matter falling into the BH and Hawking radiation being emitted. For this Hawking radiation to be detectable by Alice, it must be (mostly) matter. Bob similarly observes matter falling into the BH and Hawking radiation being emitted. For this Hawking radiation to be detectable by Bob, it must be (mostly) matter. In either case, however, the antimatter partners of the emitted matter fall into the BH; Bob's Hawking-radiation matter must, therefore, be Alice's Hawking-radiation antimatter and vice-versa. The effective change in the signature of the metric while passing through the horizon can thus be thought of as being reabsorbed in the energy of the anti-particle produced at the horizon. These are, equivalently, requirements for both observers to have positive time coordinates. The sign of the time coordinate must, therefore, be flipped when passing through the horizon, a situation that is allowed by each observer's ignorance of the other's metric. Alice's and Bob's other reference frames must similarly be CPT symmetric. This CPT symmetry at S is reminiscent of that obtained in a field-theoretic construction (e.g. [11]) but, except for the existence of Hawking radiation, without explicit field theoretic considerations.

Covariance requires that if Alice abandons her asymptotic station and falls freely into the BH, she will observe nothing unusual. This is indeed the case in a symmetric geometry: her mass reference frame shows the local curvature, which is CPT-symmetric inside S, continuing to increase toward a singularity as she would expect on the basis of her (Schwartzschild) analytical continuation. Consistent with Susskind-Thorlacius horizon complementarity, Alice cannot send information about entangled states observed near or beyond the horizon back to an asymptotic colleague in her universe. Nor can she communicate with Bob; their mutually CPT-symmetric reference frames would instead annihilate.

We can now ask whether Alice, using her reference frames, can observe a violation of no-cloning, and hence whether a high-energy firewall is required to prevent Alice's approach to the horizon. Almheiri, Marolf, Polchinski and Sully [58] begin by assuming a unitary S-matrix from infalling matter to outgoing Hawking radiation and further assume, critically, that Alice can know both the initial state and this S-matrix. This latter assumption, however, immediately violates the GHP. Unitarity for the BH plus exterior system requires that the S-matrix can equivalently be represented as transforming, with time reversal, particles incident on the horizon from the inside to particles evaporated from the horizon on the outside and vice-versa (Fig. 7). In this representation, the GHP forbids knowledge of the S-matrix for either observer. Equivalently, knowing the S-matrix would require Alice to access reference frames at past and future null infinity, as only these could measure all soft hair correlations across S [59] [60].

Setting this problem of knowledge aside and assuming that Alice has observed entanglement between subsystems comprising late and early Hawking radiation as Almheiri, Marolf, Polchinski and Sully propose, Alice can observe a no-cloning violation only if she is also capable of observing entanglement between the late radiation and its partner modes across the horizon. Such entanglement is easily inferred: each outgoing Hawking-radiation particle is entangled with its anti-particle that has fallen into the BH. Observing such entanglement, however, requires a horizon-crossing reference frame, which the GHP again forbids. A partner mode is, by definition, a subsystem, and the GHP forbids the transfer of subsystem-dependent information across S.

The symmetric geometry of Fig. 7 provides a simple explanation of why Alice's information (assuming that she can obtain it) about entanglement outside the BH cannot penetrate the horizon to be compared with information about entanglement inside the BH. Were Alice to fall, in her proper time, through S, Bob with his CPT-symmetric reference frames would observe her only as a shower of Hawking radiation. The "firewall" that consumes Alice is, in other words, on the *inside* of the horizon, and it is apparent only to inside observers using their CPT-symmetric reference frames. Alice with her external reference frames observes nothing unusual as noted earlier. A similar firewall exists outside the horizon, apparent only to outside observers, that consumes not infalling matter and information like Alice, but rather matter and information falling out from the inside. The firewall, in other words, is an alternative description of the Hawking evaporation mechanism. Its physical role is the destruction of GHP-forbidden mereological information transiting S, including any information about entanglement between subsystems on the opposite side of S.

6 Conclusion

The HP appears, at first glance, to be both counterintuitive and limited to contexts where gravitation and quantum theory intersect. We have shown here that it is neither. It is, instead, a special case of a more general principle, the GHP, that can be translated roughly as "information is available only through interaction." This GHP has both philosophical and formal roots in the 18th century, and has been formulated independently in multiple disciplines from the mid-20th century onwards. It limits the classical information transmitted across any inter-system boundary to the information that can be encoded by the eigenvalues of the inter-system interaction. As to observe something is to obtain information from it by interacting with it, the GHP places a quantitative upper limit on the bandwidth of observations employing finite resources.

We offer, in summary, the following conclusions:

1. The GHP characterizes all transfers of classical information across decompositional boundaries in associatively-decomposable state spaces. It constitutes a significant limit on information transfer whenever the encodable classical information is finite.
2. The GHP applies not just to $(d-1)$-dimensional boundaries of d-dimensional geometric spaces, but to any collection of degrees of freedom bounding a larger collection of degrees of freedom. The HP as originally formulated is a geometric special case.
3. The spatial degrees of freedom on an information-encoding boundary are strictly ancillary. The interaction that transfers information across the boundary is independent, in principle, of such degrees of freedom.
4. The GHP strictly forbids transfers of mereological information across inter-system boundaries. Subsystem boundaries or inter-subsystem interactions within an overarching system cannot be considered ontological by observers confined to any other (sub)system.
5. "Observation" is holographic encoding. Observational outcomes and hence the "observed world" are both classical and strictly observer-relative by definition.

While its simplicity, deep history and independent discovery by multiple disciplines indicate that the GHP must be considered to be in some sense "intuitive" or even "obvious," the above indicate that it is also extraordinarily powerful. Strict compliance with the GHP sheds new light on unitarity and entanglement, provides a simple and novel explanation of gauge invariance, shows that "decoherence" and "observation" are alternative terms for the same process, and suggests that the apparent conflict between unitarity and covariance highlighted by the BHIP is due not to unitarity or to covariance but rather to pre-theoretical assumptions about cross-horizon access to mereological information that the GHP explicitly forbids.

Acknowledgements CF, DDH and CP thank Federico Faggin for relevant discussions and the Federico and Elvia Faggin Foundation for financial support.

References

1. t Hooft G., Dimensional reduction in quantum gravity. In: Ali A., Ellis J. & Randjbar-Daemi S. (Eds.) Salamfestschrift. Singapore: World Scientific, Singapore, pp. 284–296 (1993).
2. Bekenstein, J. D., Black holes and entropy. Physical Review D 7(8): 2333–2346 (1973).
3. Bousso R., The holographic principle. Reviews of Modern Physics 74(3): 825–874 (2002).
4. Rovelli, C., Black holes have more states than those giving the Bekenstein-Hawking entropy: A simple argument. Preprint arXiv:1710:00218 [gr-qc] (2017).
5. Susskind, L., The world as a hologram. Journal of Mathematical Physics 36: 6377–6396 (1995).
6. Maldecena, J., The large N limit of superconformal field theories and supergravity. Advances in Theoretical and Mathematical Physics 2: 231–252 (1998).
7. Bekenstein, J. D., Black holes and information theory. Contemporary Physics 45(1): 31–43 (2004).
8. Ashby, W. R., Introduction to Cybernetics. Chapman and Hall, London (1956).
9. Pearl J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo CA (1988).
10. Hoffman D. D., Singh M. and Prakash C., The interface theory of perception. Psychonomic Bulletin & Review 22(6): 1480–1506 (2015).
11. Harlow, D., Jerusalem lectures on black holes and quantum information. Reviews of Modern Physics 88: 015002 (2016).
12. Gauss, C. F., Theorie der Anziehung homogener Ellipsoide. Commentationes societatis regiae scientarum Gottingensis recentiores II: 51–74 (1813).
13. Koenderink, J. J., The all seeing eye? Perception 43: 1–6 (2014).
14. Bartlett, S.D., Rudolph, T. and Spekkens, R.W., Reference frames, superselection rules, and quantum information. Reviews of Modern Physics 79: 555–609 (2007).
15. Moore, E.F., Gedankenexperiments on sequential machines. In: Shannon, C.W., and McCarthy, J. (Eds.) Autonoma Studies. Princeton University Press, Princeton, NJ, pp. 129-155 (1956).
16. Fields, C., Building the observer into the system: Toward a realistic description of human interaction with the world. Systems 4: 32 (2016).
17. Church, A., An unsolvable problem of elementary number theory. American Journal of Mathematics 58(2): 345–363 (1936).
18. Turing, A. R., On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society 442: 230–265 (1936).
19. Goldberg, R. P., A survey of virtual machine research. IEEE Computer 7(6), 34–45 (1974).
20. Tanenbaum, A. S., Structured Computer Organization. Prentice Hall, Upper Saddle River, NJ (1976).
21. Smith, J. E. and Nair, R., The architecture of virtual machines. IEEE Computer 38(5), 32–38 (2005).
22. Hopcroft, J. E. and Ullman, J. D. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, Reading, MA (1979).
23. Hume, D., An Inquiry Concerning Human Understanding. Oxford University Press, Oxford, UK (1777/2008).
24. Kant, E., Critique of Pure Reason. Cambridge University Press, Cambridge, UK (1781/1998).
25. Friston, K., Life as we know it. Journal of the Royal Society: Interface 10: 20130475 (2013).
26. Friston, K., Levin, M., Sengupta, B. and Pezzulo G., Knowing ones place: A free-energy approach to pattern regulation. Journal of the Royal Society: Interface 12: 20141383 (2015).

27. Fields, C. and Levin, M., Multiscale memory and bioelectric error correction in the cytoplasm-cytoskeleton-membrane system. WIREs Systems Biology and Medicine 10: e1410 (2018).
28. Kirchhoff, M., Parr, T., Palacios, E., Friston, K. and Kiverstein, J., The Markov blankets of life: Autonomy, active inference and the free energy principle. Journal of the Royal Society: Interface 15: 20170792 (2018).
29. Zanardi, P., Virtual quantum subsystems. Physical Review Letters 87: 077901 (2001).
30. Zanardi, P., Lidar, D. A. and Lloyd, S., Quantum tensor product structures are observable- induced. Physical Review Letters 92: 060402 (2004).
31. Harshman, N. L. and Ranade, K. S., Observables can be tailored to change the entanglement of any pure state. Physical Review A 84: 012303 (2011).
32. Trivers, R. L., The Folly of Fools. Basic Books, New York (2011).
33. Pizlo, Z., Li, Y., Sawada, T. and Steinman, R.M., Making a Machine that Sees Like Us. Oxford University Press, New York (2014).
34. Hoffman, D. D., The interface theory of perception. Current Directions in Psychological Science 25: 157–161 (2016).
35. Hoffman, D. D., The interface theory of perception. In Serences, J. (Ed.) The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Vol. II: Sensation, Perception & Attention. John Wiley, Hoboken, NJ (in press).
36. Fields, C., Hoffman, D. D., Prakash, C. and Singh, M., Conscious agent networks: Formal analysis and application to cognition. Cognitive Systems Research 47: 186–213 (2018).
37. Fields, C., Hoffman, D. D., Prakash, C. and Prentner, R., Eigenforms, interfaces and holographic encoding : Toward an evolutionary account of objects and spacetime. Constructivist Foundations 12(3): 265–273 (2017).
38. Zeh, D., On the interpretation of measurement in quantum theory. Foundations of Physics 1(1): 69–76 (1970).
39. Zeh, D., Toward a quantum theory of observation. Foundations of Physics 3(1): 109–116 (1973).
40. Joos, E.; Zeh, D. The emergence of classical properties through interaction with the environment. Zeitschrift für Physik B: Condensed Matter 59: 223–243 (1985).
41. Zurek, W.H. Pointer basis of the quantum apparatus: Into what mixture does the wave packet collapse? Physical Review D 24: 1516–1525 (1981).
42. Zurek, W.H. Environment-induced superselection rules. Physical Review D 26: 1862–1880 (1982).
43. Zurek W. H., Decoherence, einselection, and the quantum origins of the classical. Reviews of Modern Physics 75(3): 715–775 (2003).
44. Schlosshauer M., Decoherence and the Quantum to Classical Transition. Springer, Berlin (2007).
45. Ollivier, H., Poulin, D. and Zurek, W. H., Objective properties from subjective quantum states: Environment as a witness. Physical Review Letters 93: 220401 (2004).
46. Ollivier, H., Poulin, D. and Zurek, W. H., Environment as a witness: Selective proliferation of information and emergence of objectivity in a quantum universe. Physical Review A 72: 042113 (2005).
47. Blume-Kohout, R. and Zurek, W. H., Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. Physical Review A 73: 062310 (2006).
48. Chiribella, G. and DAriano, G.M., Quantum information becomes classical when distributed to many users. Physical Review Letters 97: 250503 (2006).
49. Zurek, W. H. Quantum Darwinism. Nature Physics 5: 181–188 (2009).
50. Tegmark, M. Many worlds in context. In: Saunders, S., Barrett, J., Kent, A. and Wallace, D. (Eds) Many Worlds? Everett, Quantum Theory and Reality. Oxford University Press, Oxford, UK, pp 553–581 (2010).
51. Wallace, D. Decoherence and ontology. In: Saunders, S., Barrett, J., Kent, A. and Wallace, D. (Eds) Many Worlds? Everett, Quantum Theory and Reality. Oxford University Press, Oxford, UK, pp 53–72 (2010).
52. Griffiths, R. B. Consistent Quantum Theory. Cambridge University Press, New York (2002).

53. Hartle, J. B. The quasiclassical realms of this quantum universe. Foundations of Physics 41: 982–1006 (2008).
54. Tegmark, M., How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem. Physical Review D 85:123517 (2012).
55. Grinbaum, A. How device-independent approaches change the meaning of physical theory. Studies in History and Philosophy of Modern Physics 58: 22–30 (2017).
56. Hawking, S. W., Breakdown of predictability in gravitational collapse. Physical Review D 14: 2460-2473 (1976).
57. Susskind, L. and Thorlacius, L., Gedanken experiments involving black holes. Physical Review D 49: 966-974 (1994).
58. Almheiri, A., Marolf, D., Polchinski, J. and Sully, J., Black Holes: Complementarity or firewalls? Journal of High Energy Physics 1302: 062 (2013).
59. Hawking, S. W., Perry, M. J. and Strominger, A., Soft hair on black holes. Physical Review Letters 116: 231301 (2016).
60. Strominger, A., Black hole information revisited. Preprint arXiv:1706.07143v1 [hep-th] (2017).