Abstract

We present a mathematical model formalizing the practice of science in nature. Axiomatic science is a significant improvement over the informal practice of science and it has numerous desirable properties. Axiomatic science is a model of science and of physics, and as such, it contains a 'science' part and a 'physics' part. Axiomatic science can derive the 'physics' part, including the laws of physics, using the 'science' part as the starting point. Axiomatic science thus explains the origins of the laws of physics as a theorem of the formal practice of science. After we present the model, we then begin the long program to derive all known physics using axiomatic science as the complete, free of physical baggage, mathematical foundation of physics.
Axiomatic science is a mathematical theory formalizing the practice of science in nature. Axiomatic science contains a 'science' part that describes the world brutally without models, patterns or laws, and a 'physics' part which derives the broadest patterns applicable to the brute description. Axiomatic science introduces (and requires the use of) 'natural models’, in which the laws of physics are derived from the description of nature, as distinct from an 'artificial model’, in which the description of nature (solutions) is/are obtained from (postulated) laws of physics. Axiomatic science is at least as general as informal science and it introduces very strong constraints on what it means for a theory to be scientific in the formal mathematical sense. For instance, all physical theories that are the product of informal science ought to be formalizable as theorems of axiomatic science, lest they would be provably unscientific.
Unlike a usual physical theory containing only a 'physics' part, axiomatic science, as it also contains a 'science' part, is unavoidably a more fundamental representation of reality, than any physical theory resulting from science. Consistent with its scope, axiomatic science proposes solutions to long and enduring problems regarding the foundation of physics. For instance; the problem of time and entropy, the origin of the appearance of a quantum collapse, identifying a preferred interpretation of quantum mechanics, as well as philosophical problems such as "why these laws of physics, and not others?", and even "why are there laws of physics at all?". Axiomatic science explains why it can answer these questions, and also explains why physics is unable to do the same: quite simply, the solutions to these problems are found in the 'science' part which precedes the 'physics' part of the axiomatic framework.

Axiomatic science is constructed using the formalism of theoretic computer science including that of Turing machines and that of algorithmic information theory. This construction negates most, and quite probably all, objections to falsificationism from the philosophy of science. By design, it is constructed to be as close to a 'necessary truth' as possible. Specifically, the domain of axiomatic science is constructed precisely as the set of all formal statements that are necessarily true for all possible state of affairs of the World. Consequently, it is necessarily the case that no formal argument can successfully invalidate elements of its domain. Furthermore, as axiomatic science is sufficiently descriptive to account for all possible state of affairs, it is also necessarily the case that there exists no fact verifiable in the world which is outside its domain. Axiomatic science is universal in the computer theoretic sense and, intuitively, in the 'physical/experimental sense'.

Axiomatic science is philosophically extremely robust, and it exceeds the robustness of informal science. It is because of both its robustness and its universality that axiomatic science, a purely mathematical construction with no physical baggage, can transpose its theorems to the domain of physics. Consequently, it will be by practicing science within the setup of axiomatic science that we will derive the laws of physics in this framework, just as we identify them when we practice science in the wild. However, in the present case, the laws of physics are derived not by experimentation but by formal proof and are derived without physical baggage, without conceptual ambiguity and in their full generally. For these reasons and because it is a formalization of the practice of science, axiomatic science is a candidate model to serve as the fundamental description of nature.

Let us start with a teaser problem to build up the intuition, then we will produce the axiomatic basis of the model.

**Which of the two logically implies the other: The egg, or the molecular theory of organic chemistry?**

As the first step towards understanding axiomatic science, we seek to understand the relationship between the 'science' and the 'physics' part, the role
played by the logical implication, by initial conditions, and by axioms. A peculiar
demand of axiomatic science is to banish what we will call 'artificial models' in
favor exclusively of what we will call 'natural models'. Let us first understand
what we mean using examples, and then we will generalize the idea. Within the
methodology of axiomatic science, the logical implication is used in the direction
that the observations imply the theory. For instance,

1. The discovery of astronomical redshift implies (or at least gives credibility
to) models accounting for a metric expansion of space.

2. The discovery of the cosmic microwave background (CMB) implies (or at
least gives credibility to) Big Bang models.

3. The measured homogeneity of the temperature of the CMB implies (or at
least gives credibility to) inflationary models.

4. The discovery of DNA implies (or at least gives credibility to) natural
selection models regarding the evolution of life on Earth.

5. The observation of objects falling from trees implies (or at least give
credibility to) the theory of gravitation.

In this paradigm, the observations form the basis of the logical argument.
From now on, we will qualify such arguments as natural; in the sense that the
conclusion logically follows from the observations. For natural models, the set of
observations takes the role of the axioms (the premise); they are the brute facts
from which the model is logically implied.

Shockingly, with perhaps axiomatic science as the only exception, we find
that no theory in physics is mathematically constructed as a natural model. Let
us first investigate how a mathematical model of nature is typically constructed,
and then explain why we believe it to be a fallacy — we will refer to it as the
artificial model fallacy.

To produce an axiomatic physical theory, one first "adjust" how one describes
nature by doing what amounts to an 'axiomatic re-organization/compression'
of the data. Specifically, one finds a new set of axioms, different than brute
observations, but believed to account for at least some good portion of the
observations, and then uses these axioms as the new starting point. This re-
organization is in the ideal case logically equivalent to the brute facts. It is
usually justified on the grounds that a more elegant or aesthetically pleasing
model will be produced which would be preferred to the brute facts. For instance,
at CERN, the LHC collision data produces about 25 petabytes of data annually
(it is algorithmically quite inelegant), but the standard model reasonably fits in
a few textbooks (comparatively, it is quite elegant). If one cares about elegance,
this is quite an improvement! As another example, consider that about 100 tons
of cosmic dust fall on earth every day, and that about 10-20 trillion drops of
water fall on Earth in the same period, etc. That is a lot of events to log as data.
But we can compress a good chunk of it by postulating that this simple formula
\[ F = Gm_1m_2/r^2 \] is a law of nature. We can compress an even bigger chunk

4
of this data by adding a few more laws; such as aerodynamics laws, weather patterns, etc.

However, with this new admittedly more aesthetically pleasing axiomatic basis as a starting point, the logical argument has a new but artificial starting point and points in a new but artificial direction. Mathematically, it is now the model that implies the observations. For instance, in physics, it is common to re-organize the presentation of the previously enumerated statements as follows:

1. The metric expansion of the universe implies (predicts) the astronomical redshift.

2. The Big Bang implies (predicts) the CMB.

3. The inflationary period immediately after the Big Bang implies (predicts) the homogeneity of the CMB temperature.

4. Natural selection implies (predicts) the existence of an information-bearing physical structure such that offsprings acquire the phenotypes of their parents (e.g. DNA).

5. Gravity implies (predicts) that objects will fall from trees, should their attachment fail.

With this re-organization, one elevates the axioms of the 'artificial model' above observations. Contrary to the direction of the natural argument, the artificial model, as it is axiomatic, now outranks observations, and consequently, the model becomes vulnerable to falsification. As far as the artificial model is concerned, its theorems are true statements, but as practitioners of science, we know that the model is overselling its salad. We are quite aware that the theorems of the model are mere predictions, not necessarily true in the World, and we do welcome and even expect the discovery of confirmatory or refuting evidence of the model. Indeed, if the axiomatic re-organization of the raw data is not equivalent in the information-theoretic sense to the raw data (in practice it seldom is), then the model will eventually make incorrect predictions and will be falsified.

Axiomatic science exposes this axiomatic re-organization as a fallacy. Axiomatic science, as a framework, connects 'raw data' to 'laws of physics' without requiring a preliminary axiomatic re-organization of the raw data. Since axiomatic science is a formal theory, employing any kind of artificial model becomes strongly prohibited within the framework. For instance, if one holds an egg, then drops it on the floor, then whatever model of reality one holds, it is now constrained to account for a broken egg on the floor. The artificial argument (the model implies the broken egg) is a false implication: there exists no such implication as in all cases the model is simply falsified should it fail to account for the broken egg.

The central tenet of axiomatic science is to construct a framework consistent with the assumption that it is not the model that constrains the World; rather, it is the World that constrains the model. In other words, the data implies the
model, but the model never implies the data. As a result, axiomatic science places the initial conditions, not at the Big Bang, but at the present because it is the present that holds the set of all constraining raw data. Even though causality can, in principle, be used as an artificial model for a subset of all observations, axiomatic science shuns its introduction as a postulate. Within the framework of axiomatic science, even something as common as assuming that the present is caused by the past cannot be done, as it is an artificial argument. Such an assumption, if true, must be formally proven from the framework as a theorem (within the 'physics' part) before it can be adopted. Consequently, it is thus more fundamental within axiomatic science to state that the past is logically implied by the present and that the system’s history may be recoverable by forensic investigation and as a model of the raw data, than it is to say that the present is caused by the past; the latter being a special case abstraction of the former.

1.1 Notation

The parentheses (example) are used to denote the order of operations. For instance \(2(1 + 2) = 6 \neq 4\). To avoid confusing maps with 'order of operations' we will elect to use the square bracket to define valued maps. For instance a map \(f : X \rightarrow \mathbb{R}\) will be written as \(f[x]\). \(S\) will denote the entropy, and \(S\) the action. Sets, unless a prior convention assigns it another symbol, will be written using the latex mathbb typography (ex: \(L, W, Q\), etc.). Matrices will have a hat (ex: \(A\)), vectors will be in bold (ex: \(a, A\)) and most other constructions (ex. scalars) will have normal typography (ex. \(a, A\)). Matrices which are diagonal will have the grave symbol, instead of the hat symbol (ex. \(\hat{a}, \hat{A}\)). The identity matrix is \(\hat{1}\), the null matrix is \(\hat{0}\), the identity vector \(1\) and the null vector \(0\).

2 The axioms of science

Definition 1 (Language). A language \(L\), with alphabet \(\Sigma\), is the set of all sentences \((s_1, s_2, \ldots)\) that can be constructed from the elements of \(\Sigma\) and it includes the empty sentence \(\emptyset\):

\[
L := \{\emptyset, s_1, s_2, \ldots\}
\]

(1)

For instance, the sentences of the binary language are:

\[
L_b := \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots\}
\]

(2)

and its alphabet is:

\[
\Sigma_b = \{0, 1\}
\]

(3)
The fundamental object of study of axiomatic science is not the electron, the quark or even the microscopic super-strings, but the experiment. An experiment represents an atom of verifiable knowledge.

**Definition 2 (Experiment).** An experiment $p$ is a tuple comprising two sentences of $L$. The first sentence, $h$, is called the hypothesis. The second sentence, $TM$, is called the protocol. Let $UTM : L \times L \rightarrow L$ be a universal Turing machine, then we say that the experiment holds if $UTM[TM, h]$ halts, and fails otherwise:

$$UTM[TM, h] \begin{cases} = r & \text{halts} \implies p \text{ holds} \\ \neq & \text{halts} \implies p \text{ fails} \end{cases}$$

If $p$ holds, we say that the protocol verifies the hypothesis. Finally, $r$, also a sentence of $L$, is the result. Of course, in the general case, there exists no computable function which can decide if an experiment holds or doesn’t.

An experiment, so defined, is formally reproducible. Indeed, for the protocol $TM$ to be a Turing machine, the protocol must specify all steps of the experiment including the complete inner workings of any instrumentation used for the experiment. The protocol must be described as an effective method equivalent to an abstract computer program. Should the protocol fail to verify the hypothesis, the entire experiment; that is, the group comprising the hypothesis, the protocol and including its complete description of all instrumentation, is falsified.

The set of all experiments are the programs that halt. The set includes all provable mathematical statements and it is universal in the computer theoretic sense.

**Definition 3 (Domain).** Let $D$ be the domain (Dom) of axiomatic science. We can define $D$ in reference to a universal Turing machine $UTM$ as:

$$D := \text{Dom}[UTM]$$

Thus, for all sentences $s$ in $L$, if $UTM[s]$ halts, then $s \in D$.

**Definition 4 (Manifest).** A manifest $M$ is a subset of $D$:

$$M \subset D$$

**Definition 5 (Set of all manifests).** Let $P[A]$ denote the power set of $A$. Then the set of all manifests $\mathcal{W}$ is:

$$\mathcal{W} := P[D]$$

Thus, $M \in \mathcal{W}$.  

7
Assumption 1 (The fundamental assumption of science). The state of affairs of the World is describable as a set of reproducible experiments. Therefore, the state of affairs is describable as a manifest. Furthermore, to each state of affairs corresponds a manifest, and finally, the manifest is a complete description of the state of affairs.

Axiom 1 (Existence of the reference manifest). As the World is in a given state of affairs, then there exists, as a brute fact, one and only one manifest $\mathbb{M}$ which corresponds to its state.

\[ \exists! \mathbb{M} \]

- $\mathbb{M}$ is called the ‘reference manifest’.
- The symbol $\mathbb{M}$ will denote any manifest in $\mathbb{W}$, whereas $\mathbb{M}$ specifically denotes the reference manifest corresponding to the present state of affairs.
- Finally, we consider the symbol $\exists$, ontological existence, as distinct from the symbol $\exists$, mathematical existence. For instance, in set theory, all manifests $\mathbb{M}$ exist ($\exists$), but in axiomatic science only $\mathbb{M}$ exists ontologically ($\exists$).

We now solidify the intuitive notions of ‘the world’ and ‘state of affairs’ used in Assumption 1 and Axiom 1.

Definition 6 (World). Mathematically, we define ‘The World’ as a tuple: \{(\mathbb{W}) \times \mathcal{P}[\mathbb{D}]\}. Specifically, it comprises the set $\mathbb{W}$ along with the reference manifest $\mathbb{M}$. Thus,

\[ \hat{\mathbb{W}} = (\mathbb{W}, \mathbb{M}) \]

- $\mathbb{M}$ is the state of affairs of $\hat{\mathbb{W}}$.
- As $\hat{\mathbb{W}}$ contains $\mathbb{M}$ as a member of the tuple, $\hat{\mathbb{W}}$ inherits the ontological qualifier previously attributed to $\mathbb{M}$ by Axiom 1.
- One may refer to a possible world (other than the one referenced in Axiom 1) by removing the over-circle notation: $\mathbb{W} = (\mathbb{W}, \mathbb{M})$.

Intuition: The reference manifest is how the world presents itself to us in the most direct, unmodelled, uninterpreted and in an uncompressed manner. Brutely knowing the manifest is how one perceives the world without understanding any patterns and without knowing any laws of physics.

As infinitely many manifests $\mathbb{M}$ can be constructed from the elements of $\mathbb{D}$, one may wonder why it is the reference manifest $\mathbb{M}$ that is actual, and not any other. This brings us to the next assumption.
Assumption 2 (The fundamental assumption of 'nature'). The reference manifest is randomly selected from the set of all possible manifests \( \mathcal{W} \) according to a probability measure \( \rho[\mathcal{M}] \).

With this assumption, we abandon all hope, as difficult to cope with as it may be, of there being a model which tells us why \( \mathcal{M} \) and not \( \mathcal{M} \) is actual. This assumption is most directly responsible for necessitating that any physical model is derived as a natural model. Essentially, it is the mathematical formulation of the intuitive notion that the state of affairs is not implied by the model.

However, as dreadful as this assumption might be, it is the key to recover the corpus of physics. The first step is to associate knowledge of \( \mathcal{M} \) to information, and it is precisely because it is randomly selected from a larger set that this is possible. We briefly recall the mathematical theory of information of Claude Shannon: Specifically, \( \mathcal{M} \) will be interpreted as a message randomly selected from the set \( \mathcal{W} \). Therefore, we can quantify the amount of information in the message \( \mathcal{M} \) as follows:

Definition 7 (Natural Information). We define natural information as the information one gains by knowing which manifest is randomly selected from \( \mathcal{W} \), according to the probability distribution \( \rho[\mathcal{M}] \). Let

\[
P := \left\{ \rho : \mathcal{W} \rightarrow [0, 1] \mid \sum_{\mathcal{M} \in \mathcal{W}} \rho[\mathcal{M}] = 1 \right\}
\]

Then, the entropy of natural information is the functional:

\[
S : \{\mathcal{W}\} \times P \rightarrow [0, \infty[ \quad (\mathcal{W}, \rho) \mapsto -\sum_{\mathcal{M} \in \mathcal{W}} \rho[\mathcal{M}] \ln \rho[\mathcal{M}]
\]

We recall that to construct an artificial model, in the informal case, one would re-organize/compress the raw data into a shorter more aesthetically pleasing and, hopefully, logically equivalent set of axioms, then call the set of axioms a model of the physical system. Intuitively, we understand that one attempted to maximize 'something' but precisely what (aesthetics? elegance?...?) was not quite clear; in the sense that the process was done heuristically and that no specific functions were maximized. This brings us to our next assumption:

Assumption 3 (The fundamental assumption of physics). The fundamental relations that result from maximizing the entropy of natural information in the world are the laws of physics.

Axiomatic science reveals that the quantity which one attempted to maximize as one informally constructed an artificial model of the data, is, in actuality, the entropy of natural information. The problem of finding the laws of physics is thus reduced to what amounts to maximizing the entropy of natural information using \( \mathcal{M} \) as the message and \( \mathcal{W} \) as the set of possible messages. With these tools,
we can solve for the laws of physics without first having to re-organize the raw data as axioms, and thus without inadvertently producing an artificial model. Indeed, as the starting point is Axiom 1, the laws of physics will necessarily be derived by a natural argument hence they will be immunized against the pathologies present in artificial models. Absent these problems, and because the laws of physics will be derived as theorems of axiomatic science, we will then attack the fundamental questions raised in the introduction. In short, the thesis of this manuscript is to prove that the world (Definition 6) with the assumptions (Assumptions 1, 2, 3 and 4) implies the laws of physics.

3 Technical introduction

To understand the relationship between natural information, entropy, statistical physics and how this implies the laws of physics, we will introduce geometric (or generalized/non-commutative) thermodynamics, but first, we will provide a recap of statistical physics, and then of algorithmic thermodynamics.

3.1 Recap: Statistical physics

Generally speaking, in statistical physics, we are interested in the distribution that maximizes the Boltzmann entropy. Let $Q$ be a set of micro-states and let:

$$
P := \left\{ \rho : Q \rightarrow [0, 1] \ \left| \sum_{q \in Q} \rho[q] = 1 \right. \right\} \quad (12)$$

Then, the Boltzmann entropy is the functional:

$$
S : \{Q\} \times P \rightarrow [0, \infty[ \quad (Q, \rho) \mapsto -k_B \sum_{q \in Q} \rho[q] \ln \rho[q] \quad (13)
$$

In statistical physics, the entropy is subject to fixed macroscopic quantities known as the statistical priors. The probability measure which maximizes the entropy under these constraints is the Gibbs ensemble. Typical thermodynamic quantities are shown in Table 1. These quantities, in general, are functions defined as:

$$
A_i : Q \rightarrow \mathbb{R} \quad q \mapsto A_i[q] \quad (14)
$$

The partition function is:

$$
Z : \{Q\} \times \mathbb{R}^n \rightarrow \mathbb{R} \quad (Q, \alpha_1, \ldots, \alpha_n) \mapsto \sum_{q \in Q} \exp \left( -\alpha_1 A_1[q] - \cdots - \alpha_n A_n[q] \right) \quad (15)
$$
Table 1: Typical thermodynamic quantities

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[q]$</td>
<td>energy</td>
<td>Joule</td>
<td>extensive</td>
</tr>
<tr>
<td>$1/T = k_B\beta$</td>
<td>temperature</td>
<td>1/Kelvin</td>
<td>intensive</td>
</tr>
<tr>
<td>$\mathcal{E}$</td>
<td>average energy</td>
<td>Joule</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$V[q]$</td>
<td>volume</td>
<td>meter$^3$</td>
<td>extensive</td>
</tr>
<tr>
<td>$p/T = k_B\gamma$</td>
<td>pressure</td>
<td>Joule/(Kelvin-meter$^3$)</td>
<td>intensive</td>
</tr>
<tr>
<td>$\mathcal{V}$</td>
<td>average volume</td>
<td>meter$^3$</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$N[q]$</td>
<td>number of particles</td>
<td>kg</td>
<td>extensive</td>
</tr>
<tr>
<td>$-\mu/T = k_B\delta$</td>
<td>chemical potential</td>
<td>Joule/(Kelvin-kg)</td>
<td>intensive</td>
</tr>
<tr>
<td>$\mathcal{N}$</td>
<td>average number of particles</td>
<td>kg</td>
<td>macroscopic</td>
</tr>
</tbody>
</table>

where $\alpha_1, \ldots, \alpha_n$ are Lagrange multipliers.

Replacing the generalized quantities by the typical thermodynamic quantities, in Table 1:

\[
\begin{align*}
\alpha_1 & := \beta \\
\alpha_2 & := \gamma \\
\alpha_3 & := \delta \\
A_1[q] & := E[q] \\
A_2[q] & := V[q] \\
A_3[q] & := N[q]
\end{align*}
\]

the partition function would be:

\[
Z[Q, \beta, \gamma, \delta] = \sum_{q \in Q} \exp \left( -\beta E[q] + \gamma V[q] + \delta N[q] \right)
\]

The probability of occupation of a micro-state (Gibbs measure) derived by maximizing the entropy (using the method of the Lagrange multipliers) is:

\[
\rho : \{Q\} \times Q \times \mathbb{R}^n \rightarrow [0, 1] \\
(Q, q, \alpha_1, \ldots, \alpha_n) \rightarrow Z^{-1} \exp \left( -\alpha_1 A_1[q] - \cdots - \alpha_n A_n[q] \right)
\]

and a typical example is:

\[
\rho(Q, q, \beta, \gamma, \delta) = \frac{1}{Z} \exp \left( -\beta E[q] - \gamma V[q] - \delta N[q] \right)
\]

The thermodynamic bulk state is defined by a set of $n$ constraints:
\[ \overline{\mathcal{A}}_i = Z^{-1} \sum_{q \in \mathcal{Q}} A_i[q] \exp (-\alpha_1 A_1[q] - \cdots - \alpha_n A_n[q]) \quad (25) \]

and typical examples are:

\[ \overline{E} = \frac{1}{Z} \sum_{q \in \mathcal{Q}} E[q] \exp (-\beta E[q] - \gamma V[q] - \delta N[q]) \quad (26) \]

\[ \overline{V} = \frac{1}{Z} \sum_{q \in \mathcal{Q}} V[q] \exp (-\beta E[q] - \gamma V[q] - \delta N[q]) \quad (27) \]

\[ \overline{N} = \frac{1}{Z} \sum_{q \in \mathcal{Q}} N[q] \exp (-\beta E[q] - \gamma V[q] - \delta N[q]) \quad (28) \]

The thermodynamic bulk quantities are also given by the following relation:

\[ \frac{\partial \ln Z[\mathcal{Q}, \alpha_1, \ldots, \alpha_n]}{\partial \alpha_i} = \overline{\mathcal{A}}_i \quad (29) \]

And the variance by the following relations:

\[ \frac{\partial^2 \ln Z[\mathcal{Q}, \alpha_1, \ldots, \alpha_n]}{\partial \alpha_i^2} = (\Delta \overline{\mathcal{A}}_i)^2 \quad (30) \]

The entropy for this ensemble is:

\[ S[\mathcal{Q}, \alpha_1, \ldots, \alpha_n] = k_B (\ln Z + \alpha_1 \overline{A}_1 + \cdots + \alpha_n \overline{A}_n) \quad (31) \]

Taking the total derivative of the entropy, we obtain:

\[ dS[\mathcal{Q}, \alpha_1, \ldots, \alpha_n] = k_B (\alpha_1 d\overline{A}_1 + \cdots + \alpha_n d\overline{A}_n) \quad (32) \]

This is the equation of the state of the system.

Typical examples are:

\[ S[\mathcal{Q}, \beta, \gamma, \delta] = k_B (\ln Z + \beta \overline{E} + \gamma \overline{V} + \delta \overline{N}) \quad (33) \]

and

\[ dS[\mathcal{Q}, \beta, \gamma, \delta] = k_B (\beta d\overline{E} + \gamma d\overline{V} + \delta d\overline{N}) \quad (34) \]

respectively.
The Fischer information metric is defined by taking the partial derivatives of \( \ln Z \) with respect to \( \alpha_i \) then to \( \alpha_j \), as follows:

\[
g_{ij} = \frac{\partial^2 \ln Z}{\partial \alpha_i \partial \alpha_j} \tag{35}
\]

For example:

\[
(ds)^2 = (d\beta \ d\gamma \ d\delta) \left( \begin{array}{ccc} \frac{\partial^2 \ln Z}{\partial \beta \partial \beta} & \frac{\partial^2 \ln Z}{\partial \beta \partial \gamma} & \frac{\partial^2 \ln Z}{\partial \beta \partial \delta} \\ \frac{\partial^2 \ln Z}{\partial \gamma \partial \beta} & \frac{\partial^2 \ln Z}{\partial \gamma \partial \gamma} & \frac{\partial^2 \ln Z}{\partial \gamma \partial \delta} \\ \frac{\partial^2 \ln Z}{\partial \delta \partial \beta} & \frac{\partial^2 \ln Z}{\partial \delta \partial \gamma} & \frac{\partial^2 \ln Z}{\partial \delta \partial \delta} \end{array} \right) (d\beta \ d\gamma \ d\delta) \tag{36}
\]

Thermodynamics is derived from statistical physics. It is concerned primarily by the fundamental relation (32). Thermodynamic changes (and cycles) can be realized by changing the quantities \( \{\alpha_1, \ldots, \alpha_n\} \) and/or by modifications of \( Q \). Under modification of \( Q \), usually by cross product: \( Q \times Q_1 = Q_2 \), or by set complement \( Q \setminus Q_3 = Q_4 \), quantities which are invariant \( \{\alpha_1, \ldots, \alpha_n\} \) are called intensive, and quantities which are variant \( \{A_1, \ldots, A_n\} \) are called extensive.

3.2 Recap: Algorithmic statistical physics

Many authors\[1, 2, 3, 4, 5, 6, 7, 8, 9\] have discussed the similarity between the Gibbs entropy \( S = -k_B \sum_{q \in Q} \rho[q] \ln \rho[q] \) and the entropy in information theory \( H = -\sum_{q \in Q} \rho[q] \log_2 \rho[q] \). Furthermore, the similarity between the halting probability \( \Omega \) and the Gibbs ensemble of statistical physics has also been studied\[10, 11, 12, 8\]. First let us introduce \( \Omega \). Let \( \mathcal{U} \) be the set of all universal Turing machines, then:

\[
\Omega : \mathcal{U} \rightarrow [0, 1] \\
\text{UTM} \rightarrow \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \tag{37}
\]

Here, \(|p|\) denotes the length of \( p \), a computer program. The domain, \( \text{Dom}[\text{UTM}] \), is the domain of the universal Turing machine (the set of all programs that halt for it). The sum represents the probability that a random program will halt on UTM. The Chaitin’s construction\[2\] (a.k.a. \( \Omega \), halting probability, Chaitin’s constant) is defined for a universal Turing machine as a sum over its domain (the set of programs that halts for it) where the term \( 2^{-|p|} \) acts as a special probability distribution which guarantees that the value of the sum, \( \Omega \), is between zero and one (The Kraft inequality \[13\]). As the sum does not erase halting information, knowing \( \Omega \) is enough to know the programs that halt and those that do not on UTM. Since the halting problem is unsolvable, \( \Omega \) must, therefore, be non-computable. \( \Omega \)’s connection to the halting problem guarantees that it is algorithmically random, normal and incompressible.
It is possible to calculate some small (always finite) quantity of bits of $\Omega$. As such, Calude\cite{14} calculated the first 64 bits of $\Omega_{\text{utm}}$ for some universal Turing machine utm as:

$$\Omega_{\text{utm}} = 0.000000100000010000000110...2$$  \hspace{1cm} (38)

Running the calculation for a handful of bits is certainly possible, however, any finitely axiomatic systems will eventually run out of steam and hit a wall. Calculating the digits of $\pi$, for instance, will not hit this kind of limitation. For $\pi$, the axioms of arithmetic are sufficiently powerful to compute as many bits as we wish to calculate, limited only by the physical resources of the computers at our disposal. To understand why this is not the case for $\Omega$, we have to realize that solving $\Omega$ requires solving problems of arbitrarily higher complexity, the complexity of which always eventually outclasses the power of any finitely axiomatic system.

In 2002, Tadaki\cite{8} suggested augmenting $\Omega$ with a multiplication constant $D$, which acts as an ’algorithmic decompression’ term on $\Omega$.

$$\Omega[\text{UTM}] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \rightarrow \Omega[\text{UTM}, D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|}$$  \hspace{1cm} (39)

With this change, Tadaki argued that the Gibbs ensemble compares to the Tadaki ensemble as follows:

<table>
<thead>
<tr>
<th>Gibbs ensemble</th>
<th>Tadaki ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z[Q, \beta] = \sum_{q \in Q} e^{-\beta E[q]}$</td>
<td>$\Omega[\text{UTM}, D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D</td>
</tr>
</tbody>
</table>

Interpreted as a Gibbs ensemble, the Tadaki construction forms a statistical ensemble where each program corresponds to one of its micro-state. The Tadaki ensemble admits the following quantities; the prefix code of length $|q|$ conjugated with $D$. As a result, it describes the partition function of a system which maximizes the entropy subject to the constraint that the average length of the codes is some quantity $|p|$:  

$$|p| : \mathbb{U} \times \mathbb{R} \rightarrow \mathbb{R}$$

(UTM, D) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} |p|2^{-D|p|}  \hspace{1cm} (41)

The entropy of the Tadaki ensemble is proportional to the average length of prefix-free codes available to encode programs:

$$S[\text{UTM}, D] = \ln \Omega + D|p| \ln 2$$  \hspace{1cm} (42)
The constant \( \ln 2 \) comes from the base 2 of the halting probability function instead of base \( e \) of the Gibbs ensemble.

John C. Baez and Mike Stay\[12\] took the analogy further by suggesting a connection between algorithmic information theory and thermodynamics, where the characteristics of the ensemble of programs are equivalent to thermodynamic observables. A stated goal was to import tools of statistical physics into algorithmic information theory to facilitate its study. In algorithmic thermodynamics, one extends \( \Omega \) with algorithmic quantities to obtain:

Baez-Stay ensemble:

\[
\Omega : \mathbb{U} \times \mathbb{R}^3 \rightarrow \mathbb{R}
\]

\[
(\text{UTM, } \beta, \gamma, \delta) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-\beta E[p]} - \gamma V[p] - \delta N[p]
\]

Noting its similarities to the Gibbs ensemble of statistical physics\[15\], these authors suggest an interpretation where \( E[p] \) is the expected value of the logarithm of the program’s runtime, \( V[p] \) is the expected value of the length of the program, and \( N[p] \) is the expected value of the program’s output. Furthermore, they interpret the conjugate variables as (quoted verbatim from their paper):

1. \( T = 1/\beta \) is the **algorithmic temperature** (analogous to temperature). Roughly speaking, this counts how many times you must double the runtime in order to double the number of programs in the ensemble while holding their mean length and output fixed.

2. \( p = \gamma/\beta \) is the **algorithmic pressure** (analogous to pressure). This measures the trade-off between runtime and length. Roughly speaking, it counts how much you need to decrease the mean length to increase the mean log runtime by a specified amount while holding the number of programs in the ensemble and their mean output fixed.

3. \( \mu = -\delta/\beta \) is the **algorithmic potential** (analogous to chemical potential). Roughly speaking, this counts how much the mean log runtime increases when you increase the mean output while holding the number of programs in the ensemble and their mean length fixed.

—John C. Baez and Mike Stay

From equation (43), they derive analogs of Maxwell’s relations and consider thermodynamic cycles, such as the Carnot cycle or Stoddard cycle. For this, they introduce the concepts of **algorithmic heat** and **algorithmic work**. Finally, we note that other authors have suggested other alternative mappings in other but related contexts\[10\] [9].
3.3 Applicability to axiomatic science

Comparing the axioms of science to the very similar computer theoretic setup for algorithmic thermodynamics, it is clear that the framework will play a major role. Axiomatic science defines experiments as protocols verifying a hypothesis, which is analogous to a program halting for an input. With algorithmic thermodynamics, we now have an algorithmic analog to statistical physics, a framework already familiar to physics and capable of producing conservation equations in the form of an equation of state, that can be applied to our axiomatic model of science.

What is left to do is to apply the suitable statistical framework to the axioms of science in such a way that the resulting equation of state is mathematically the same as the laws of physics. Will that be easy to do or will it be hard? Well, let us investigate. For practical reasons, out of the many attempts we have explored we will summarize our efforts as two attempts, then we will give the retained solution. First, let us state what will be easy to do. Using the framework of algorithmic thermodynamics, one can produce an equation of state of computing resources. In this case, one interprets algorithmic thermodynamics as describing a maximally informative computation over a set of programs randomly selected from the space of all programs. Then, augmenting the ensemble with a thermodynamic temperature, the computing resources are connected to the conservation of energy (for these resources), and a series of Landauer-type relations are obtainable. This will be attempt 1.

3.4 Attempt 1: Algorithmic thermodynamics

The Journal of Natural Computing defines the subject as:

"Natural Computing refers to computational processes observed in nature, and human-designed computing inspired by nature."[1]

We are interested in how systems of algorithmic thermodynamics relate to the first part of this definition; how and under what conditions are such systems realized/realizable in nature? A related question is how much of nature can be described as natural computing — is it all of it, or is it only part of it? One (naive) application of algorithmic thermodynamics could be as follows: consider the archetypal ensemble of statistical physics: the classical system of a perfect gas in a box of constant volume. One can surely interpret the changing distribution of the gas molecules within the box as a computation that, over time, maps out the space of solutions for the dynamical equations for the perfect gas. How insightful is that application likely to be? Well, this application amounts to just plastering a computing description on top of an already satisfactory physical description of the system. Why hinder ourselves with the additional overhead? Instead, we will be looking for a much more fundamental description; we want the computing system to stand on its own merits. Specifically, our goal is not to describe the laws of physics as analogous to performing a computation, but

to instead find the proper statistical description under which the equation of
state of the computation gives us the laws of physics (we note that in axiomatic
science the computation is the practice of science and it comes before the laws
of physics).

The first step to connect algorithmic thermodynamics to nature is to not shy
away from the computer-theoretic origins of algorithmic thermodynamics, and
to use quantities consistent with this origin. Therefore, instead of arbitrarily
mapping, say the runtime to the energy, and the program length to the volume
(or permutations of such) we will ground said quantities within the terminology
of computer science.

We will introduce two partition functions. The first is a canonical ensemble
over the domain of a universal Turing machine. The quantities of this partition
function are listed in Table 2. They are $o_k$, the computing repetency
conjugated with $O_x[p]$ the program length, and $o_f$ the computing frequency
conjugated with $O_t[p]$ the program time. The partition function is:

$$Z : U \times \mathbb{R}^2 \rightarrow \mathbb{R},
(UTM, o_k, o_f) \mapsto \sum_{p \in \text{Dom}[UTM]} 2^{-o_k O_x[p] - o_f O_t[p]} \quad (44)$$

The second partition function is a grand canonical ensemble. It is similar to
the previous case but the sum is over the finite elements of the power set of
the domain:

$$\mathcal{W} := \mathcal{P}(\text{Dom}[UTM]) \quad (45)$$

where $\mathcal{P}(\text{Dom}[UTM])$ is the power set of the set of halting programs for UTM.
Executing a manifest $\mathcal{M} \in \mathcal{W}$ of programs on a universal Turing machine refers
to a specific computation involving multiple programs. In this ensemble, we add
the quantity $o_\mu$, the computing overhead conjugated to $O_n[\mathcal{M}]$, the quantity of
programs in the manifest. The quantities of this ensemble are shown in Table 3
and its partition function is:
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_x[M]$</td>
<td>length of programs in the manifest</td>
<td>[bit]</td>
<td>extensive</td>
</tr>
<tr>
<td>$o_k$</td>
<td>computing repetency</td>
<td>[1/bit]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\overline{O}_x$</td>
<td>average tape usage</td>
<td>[bit]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$O_t[M]$</td>
<td>running time of programs in the manifest</td>
<td>[operation]</td>
<td>extensive</td>
</tr>
<tr>
<td>$o_f$</td>
<td>computing frequency</td>
<td>[1/operation]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\overline{O}_t$</td>
<td>average clock time</td>
<td>[operation]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$O_n[M]$</td>
<td>quantity of programs in the manifest</td>
<td>[program]</td>
<td>extensive</td>
</tr>
<tr>
<td>$o_\mu$</td>
<td>computing overhead</td>
<td>[1/program]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\overline{O}_n$</td>
<td>average concurrency</td>
<td>[program]</td>
<td>macroscopic</td>
</tr>
</tbody>
</table>

Table 3: Algorithmic quantities of the grand canonical ensemble of programs

\[
Z : \{\mathcal{W}\} \times \mathbb{R}^3 \longrightarrow \mathbb{R} \\
(W, o_k, o_f, o_\mu) \mapsto \sum_{M \in \mathcal{W}} 2^{-o_k O_x[M]-o_f O_t[M]-o_\mu O_n[M]} \tag{46}
\]

The corresponding probability measure is:

\[
\rho : \{\mathcal{W}\} \times \mathcal{W} \times \mathbb{R}^3 \longrightarrow \mathbb{R} \\
(W, M, o_k, o_f, o_\mu) \mapsto Z^{-1}2^{-o_k O_x[M]-o_f O_t[M]-o_\mu O_n[M]} \tag{47}
\]

The probability measure maximizes the entropy subject to the following bulk constraints:

\[
\overline{O}_x = \sum_{M \in \mathcal{W}} O_x[M] \exp \left(2^{-\sigma_k O_x[M]-\sigma_f O_t[M]-\sigma_\mu O_n[M]}\right) \tag{48}
\]
\[
\overline{O}_t = \sum_{M \in \mathcal{W}} O_t[M] \exp \left(2^{-\sigma_k O_x[M]-\sigma_f O_t[M]-\sigma_\mu O_n[M]}\right) \tag{49}
\]
\[
\overline{O}_n = \sum_{M \in \mathcal{W}} O_n[M] \exp \left(2^{-\sigma_k O_x[M]-\sigma_f O_t[M]-\sigma_\mu O_n[M]}\right) \tag{50}
\]

The Lagrange multipliers ($o_k$, $o_f$ and $o_\mu$) are interpreted, in the style of Baez and Stay, as:

- The computing repetency $o_k$ counts how many times the average tape usage $\overline{O}_x$ must be doubled to double the entropy of the ensemble while holding the average clock time $\overline{O}_t$ and the average concurrency $\overline{O}_n$ fixed.
- The computing frequency $o_f$ counts how many times the average clock time $\overline{O}_t$ must be doubled to double the entropy of the ensemble while holding the average tape usage $\overline{O}_x$ and the average concurrency $\overline{O}_n$ fixed.

18
The computing overhead $o_\mu$ counts how many times the average concurrency $O_n$ must be doubled to double the entropy of the ensemble while holding the average clock time $O_t$ and the average tape usage $O_r$ fixed.

Various systems of natural computing can be produced using other resources. Let us give a few examples.

1. **Computing time to program frequency formulation:**

   \[
   Z' : \quad \mathbb{U} \times \mathbb{R}^2 \rightarrow \mathbb{R} \\
   (\text{UTM}, o_k, o_t) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-o_t O_x[p] - o_t O_f[p]} \tag{51}
   \]

   To formulate this relation, we introduce the program frequency $O_f[p]$ as the inverse of the program time $O_t[p]$, thus $O_f[p] := 1/O_t[p]$. This formulation fixes an average clock frequency $O_f$ by having the programs executed under a constant computing time $o_t$:

   • The computing time $o_t$ counts how many times the average clock frequency $O_f$ must be doubled to double the entropy of the ensemble while holding the average tape usage $O_r$ and the average concurrency $O_n$ fixed.

2. **Size-cutoff formulation:**

   \[
   Z'' : \quad \mathbb{U} \times \mathbb{R}^2 \rightarrow \mathbb{R} \\
   (\text{UTM}, o_k, x) \mapsto \sum_{p \in \{q \in \text{Dom}[\text{UTM}] | O_x[q] < x\}} 2^{-o_k O_x[p]} \tag{52}
   \]

   The sum $Z''$ only includes programs with length less than or equal to $x$. $\Omega$ is recovered in the limit when $x \rightarrow \infty$ (and when $o_k = 1$). $Z''$ represents the first $n$ bits of $\Omega$ up to a cutoff proportional to $x$.

3. **Time-cutoff formulation:**

   \[
   Z''' : \quad \mathbb{U} \times \mathbb{R}^3 \rightarrow \mathbb{R} \\
   (\text{UTM}, o_k, o_f, t) \mapsto \sum_{p \in \{q \in \text{Dom}[\text{UTM}] | O_t[q] < t\}} 2^{-o_k O_x[p] - o_f O_x[p]} \tag{53}
   \]

   The sum $Z'''$ only includes programs that halt within a time cutoff $t$. Thus, $Z'''$ contains no "non-halting information" and is computable. $\Omega$ is recovered in the limit when $t \rightarrow \infty$ (and when $o_k = 1$).
4. **Computational-complexity formulation:**

\[
Z'''' : U \times \mathbb{R} \times \{ G : f \to h \} \rightarrow \mathbb{R} \\
(\text{UTM}, o_k, g) \mapsto \sum_{p \in \{ q : \text{Dom}(\text{UTM}) \mid O(\text{UTM}[q]) \leq g \}} 2^{-o_k O_x[p]} 
\]  

The sum only includes programs that halt and whose computational complexity is less than or equal to \( g \).

**Interpretation:**

1. **Feasible computing complexity:**

It is not a coincidence that we elected to use the letter O (with indices) to identify the functions that define the quantity of resources required to execute a program to termination (for instance \( O_x[p], O_f[p] \), etc.).

Recall the Big O notation used in computational complexity theory. Unlike in algorithmic thermodynamics, computational complexity theory has no need for physical resource indicators (clock speed, time-cutoffs, etc) to define the computational complexity of programs because said difficulty is defined as the relation between the size of the input and the number of steps required to solve the problem (a definition independent of physical resource availability). For example, in complexity theory a program with input \( n \) which takes \( 10^{9999} n \) steps to halt would likely take longer to run than the age of the universe on any physical computer even for \( n = 1 \), but nonetheless computational complexity theory considers this intractable problem to be an easier problem than a problem which takes \( n + 10^{-10} n^2 \) steps to solve. Consequently, computational complexity theory based on Big O notation does not quite connect to the physical reality of computation with limited available resources.

Using an ensemble of algorithmic thermodynamics, a cost-to-compute, measured in entropy, can be attributed to carrying out a computation using finite resources, provided that the system is at equilibrium (i.e. for the Lagrange multipliers \( o_k, o_f, \ldots \)). Knowing which program or which manifest was computed from a small set of options is informative, and knowing which program/manifest was computed from the set of all possible programs/manifests is maximally informative. It is by this connection to entropy that statistical ensembles of programs are connect to the physical reality of computation under limited resources.

2. **Reservoirs of computing resources:**

It is common in statistical physics to appeal to various reservoirs such as a thermal reservoir or a particle reservoir, etc. The typical Gibbs ensemble in physics is \( Z(Q, \beta) = \sum_{q \in Q} \exp(-\beta E[q]) \). It’s average energy is given by \( \bar{E} = -\partial \ln Z / \partial \beta \) and its fluctuations are \( (\Delta E)^2 = \partial^2 \ln Z / \partial \beta^2 \). To justify
that fluctuations are possible and compatible with the laws of conservation of energy, the system is claimed/idealized to be in contact with a thermal reservoir. In this idealized case, both the system and the reservoir have the same temperature and they can exchange energy. The reservoir is considered large enough that the fluctuations of the smaller system are negligible to its description. Mathematically, the reservoir has infinite heat capacity. Thus, the reservoir abstractly represents an infinitely deep pool of energy at a given, constant temperature.

A similar analogy can be supported for a system of natural computing, in which the computing resources are provided to the system in the form of reservoirs. For instance, instead of a thermal reservoir, we may have runtime and tape reservoirs. These reservoirs have mathematically infinite runtime and tape capacities and thus acts as infinitely deep pools of computing resources. Computing is made possible by the interaction of the reservoirs with the system, and the intensity of the exchanges is calibrated by the computing repetency and the computing frequency, instead of by the temperature.

Maximizing the entropy quantifies the maximum amount of algorithmic information a user can obtain out of the system by selecting the manifest of programs to run within the available resources. By considering that the group of reservoirs are a 'supercomputer', the analogy is completed and algorithmic thermodynamics describes the dynamics of computation in equilibrium with the resources made available by the supercomputer.

3. Reference manifest

In statistical physics, knowing which micro-state happens to be occupied at any given time is inconsequential by design (the entropy is maximized to erase knowledge of which micro-state is occupied). However, algorithmic thermodynamics, unlike statistical physics, admits a reference manifest and therefore the fact that a specific micro-state is 'occupied', is consequential. Indeed, as the computation of an actual manifest is carried out and eventually completed, the production of Shannon information associated with the random selection of the reference manifest from the set of all possible manifests is a new emergent characteristic. The entropy of the system, before the computation is carried out, is information after the computation is carried out.

So far so good; but where is quantum mechanics, where is the qubit, where is the geometry of space-time...?

3.5 Hint 1: Seth Lloyd

In 2002, Lloyd[15] calculated the total number of bits available for computation in the universe, as well as the total number of operations that could have occurred since the universe’s beginning.
For both quantities (the quantity of bits stored in the universe and the quantity of operations made on those bits), Lloyd obtains the number $\approx 10^{122}k_B[\text{bit}]$. This number is consistent with other approaches; for instance, the Bekenstein-Hawking entropy\(^\text{[16, 17]}\) of the cosmological horizon (also $\approx 10^{122}k_B[\text{bit}]$), and the entropy of the holographic surface at the cosmological horizon suggested by Susskind\(^\text{[18]}\) (also $\approx 10^{122}k_B[\text{bit}]$).

How did Lloyd derive these numbers? First, he calculated the value for these quantities while ignoring the contribution of gravity and he obtained $\approx 10^{90}k_B[\text{bit}]$. It is only by including the degrees of freedom of gravity that the number $\approx 10^{122}k_B[\text{bit}]$ is obtained, which he does in the second part. As we are interested in the totals, we will go directly to the calculations that include the contribution of gravity. We state Lloyd’s main result and note that the details of the calculation can be reviewed in his paper. Lloyd obtains a relation between time and number of operations for the universe:

$$\#\text{ops} \approx \frac{\rho_c c^5 t^4}{\hbar} \approx \frac{t^2 \bar{c}^3}{G \hbar} = \frac{1}{t_p^2} t^2$$

(55)

where $\rho_c$ is the critical density and $t_p$ is the Planck time and $t$ is the age of the universe. With present-day values of $t$, the result is $\approx 10^{122}k_B[\text{bit}]$. Lloyd concludes that his results are consistent with the Bekenstein bound and the holographic principle. He states:

"Applying the Bekenstein bound and the holographic principle to the universe as a whole implies that the maximum number of bits that could be registered by the universe using matter, energy, and gravity is $\approx \frac{c^2 t^2}{t_p^2}$."

which is also $\approx 10^{122}k_B[\text{bit}]$. A particularly interesting consequence of this result is that these relations appear to imply conservation of both information and operations in space-time (the numerical quantity of $10^{122}$ is obtained by summing over all available degrees of freedom in space-time). So with this hint, we are now looking for a fundamental relationship between entropy, information, operations, and... space-time.

We will generalize the intuition to the following statement: A deformation of the geometry entails a modification of the entropy.

### 3.6 Hint 2: Entropy and space-time

A relation between entropy and space-time has been anticipated (or at least hinted at) since probably the better part of four decades. The first hints were provided by the work of Bekenstein\(^\text{[19, 20, 21]}\) regarding the similarities between black holes and thermodynamics, culminating in the four laws of black hole thermodynamics. The temperature, originally introduced by analogy, was soon augmented to a real notion by Hawking\(^\text{[16]}\) with the discovery of the Hawking temperature derived from quantum field theory on curved space-time. We note
the discovery of the Bekenstein-Hawking entropy, connecting the area of the surface of a horizon to be proportional to one fourth the number of elements with Planck area that can be fitted on the surface: \( S = k_Bc^3/(4\hbar G)A \).

We mention Ted Jacobson\[22\] and his derivation of the Einstein field equation as an equation of state of a suitable thermodynamic system. To justify the emergence of general relativity from entropy, Jacobson first postulated that the energy flowing out of horizons becomes hidden from observers. Next, he attributed the role of heat to this energy for the same reason that heat is energy that is inaccessible for work. In this case, its effects are felt, not as "warmth", but as gravity originating from the horizon. Finally, with the assumption that the heat is proportional to the area \( A \) of the system under some proportionality constant \( \eta \), and some legwork, the Einstein field equations are eventually recovered.

Recently, Erik Verlinde\[23\] proposed an entropic derivation of the classical law of inertia and those of classical gravity. He compared the emergence of such laws to that of an entropic force, such as a polymer in a warm bath. Each law is emergent from the equation \( T \, dS = F \, dx \), under the appropriate temperature and a posited entropy relation. His proposal has encouraged a plurality of attempts to reformulate known laws of physics using the framework of statistical physics. Visser\[24\] provides, in the introduction to his paper, a good summary of the literature on the subject. The ideas of Verlinde have been applied to loop quantum gravity (\[25\]), the Coulomb force (\[26\]), Yang-Mills gauge fields (\[27\]), and cosmology (\[28\], \[29\], \[30\]). Some criticism has, however, been voiced\[31\], \[32\], \[33\], \[34\], \[35\], including by Visser\[24\].

Even more recently, a connection between entanglement entropy and general relativity has been supported by multiple publications\[36\], \[37\], \[38\], \[39\], \[40\], \[41\], \[42\], \[43\], \[44\], \[45\], \[46\], \[47\], \[48\], \[49\], \[50\], \[51\], \[52\], \[53\], \[54\], \[55\], \[56\], \[57\], \[58\], \[59\], \[60\], \[61\], \[62\], \[63\], \[64\], \[65\], \[66\], \[67\], \[68\], \[69\], \[70\], \[71\], \[72\], \[73\].

Finally, we mention the body of work of George Ellis regarding the evolving block universe hypothesis detailed in \[54\], \[55\], \[56\] and the connection between space-time events, general relativity and quantum mechanics.

We are now ready to investigate our second attempt at a solution.

### 3.7 Attempt 2: The search for a suitable ensemble

Our second series of attempts could be grouped under a simple concept: we attempted to construct a specific system of statistical physics having a double interpretation; one, as a system of algorithmic thermodynamics admitting an equation of state involving bits and operations, and second, that said equation of state be interpretable as a physical system of space-time (perhaps as a solution to general relativity).

Finding a specific system of statistical physics means attributing an implementation to the thermodynamic observable functions \( (O_{\alpha}(p), O_{\beta}(p), \text{etc.}) \) used in the partition function. This is essentially the approach used by Ted Jacobson and Erik Verlinde in the context of connecting general relativity and classical gravity, respectively, to entropy. In each of their papers, the degrees of freedom of space are assumed to be quadratic (i.e. they grow as an area law). Consequently, the thermodynamic observables are quadratic degrees of freedom. Attempting to
expand upon these ideas, we have investigated the emergence of many physical laws, including a toy model of a cosmology emergent from quadratic degrees of freedom. However, in the end, we felt that there was a general problem with this approach.

The problem with this approach, even if it successfully lead to some set of laws, is that any results would be specific to the constructed ensemble. One would still have to justify why this specific ensemble and not another happens to be the one that describes the World. But of course, picking the ensemble via postulation would negate any possibility of a satisfying answer. Specifically, we were unable to justify by natural argument why we would pick this ensemble over any other. Such ensemble would thus suffer from the artificial model fallacy which is precisely what we are trying to avoid in this manuscript.

Furthermore, we were missing out on the full potential of statistical physics as a general framework. Indeed, statistical physics can produce conservation equations on the broadest of scales. As a typical example, we refer to the fundamental relation of thermodynamics involving the conservation of energy over a change in thermodynamic observables:

\[
d\bar{E} = T\,d\bar{S} + p\,d\bar{V} - \mu\,d\bar{N}
\] (56)

To capture this generality, our retained solution was not to define a specific system of statistical physics, but instead to increase the generality of thermodynamics; in the present case, with a non-commutative algebra applied to the thermodynamic observables. In this generalization, which we call \textit{geometric thermodynamics}, the general conservation relation above becomes a special case of an even more general conservation relation that, surprisingly, has the suitable properties.

We will now introduce the retained solution: geometric thermodynamic. First, as a sketch, then rigorously as geometric statistical physics in section 3.8.

### 3.8 Geometric Thermodynamics (as a sketch)

We identified the potential to generalize statistical physics with a non-commutative algebra as we attempted to create thermodynamic cycles that are consistent with the symmetries of space-time. By doing so, we realized that such cycles could be produced if the relevant thermodynamic observables obeyed a non-commutative algebra. With this insight, we have "reverse engineered" the type of partition function along with a suitable microscopic object of study which would eventually produce cycles with suitable properties.

To understand in more detail, let us investigate a hypothetical cycle involving several thermodynamic observables. Let's name them $X$, $Y$ and $Z$. Such quantities would be extensive, have the meter as their unit, and would be conjugated to a Lagrange multiplier $\tilde{k}$ having the inverse units ($m^{-1}$). The equation of state of such a system would be:
\[ \hat{k}^{-1} dS = dX + dY + dZ \] (57)

For a change over the quantities \( X, Y \) and \( Z \) to be consistent with the symmetries of Euclidean space, one would expect that the change in entropy along two paths of equal distance, say a path going in a straight line from \((0, 0, 0)\) to \((0, 5, 0)\) and a path going in a straight line from \((0, 0, 0)\) to \((3, 4, 0)\), to be equal. Indeed, the Euclidean distance along either path is the same: in this case, 5 meters. Since the paths are related to one another via rotation of the frame of reference, the entropic cost of the transformation should only depend on the Euclidean length of the path, and not on the orientation of the frame of reference.

One can enforce this property by demanding that the thermodynamic observables obey a suitable non-commutative algebra. Let’s see with an example. As the first step, we add the generators of an algebra, say we name them \( \{\sigma_1, \sigma_2, \sigma_3\} \), to each quantity. We get:

\[ \hat{k}^{-1} dS = \sigma_1 dX + \sigma_2 dY + \sigma_3 dZ \] (58)

We note that in this expression, the entropy becomes a vector, and so this will be addressed rigorously in section 4. As we will now see, this entropy will become a real number by squaring it.

The second step is to verify that the entropy conforms to the Euclidean distance. We can investigate if this is the case by squaring the equation of state. We obtain:

\[ \hat{k}^{-2}(dS)^2 = \sigma_1^2 (dX)^2 + \sigma_2^2 (dY)^2 + \sigma_3^2 (dZ)^2 \]

\[ + (\sigma_1 \sigma_2 + \sigma_2 \sigma_1) dX dY + (\sigma_1 \sigma_3 + \sigma_3 \sigma_1) dX dZ + (\sigma_2 \sigma_3 + \sigma_3 \sigma_2) dY dZ \] (59)

In the case where \( \sigma_1, \sigma_2 \) and \( \sigma_3 \) are commutative, the cross terms \( \sigma_1 \sigma_2 + \sigma_2 \sigma_1, \sigma_1 \sigma_3 + \sigma_3 \sigma_1 \) and \( \sigma_2 \sigma_3 + \sigma_3 \sigma_2 \) do not cancel, but if they are, say matrices, that obey the following relations:

\[ \sigma_1^2 = 1 \] (60)
\[ \sigma_2^2 = 1 \] (61)
\[ \sigma_3^2 = 1 \] (62)
\[ \sigma_1 \sigma_2 + \sigma_2 \sigma_1 = 0 \] (63)
\[ \sigma_1 \sigma_3 + \sigma_3 \sigma_1 = 0 \] (64)
\[ \sigma_2 \sigma_3 + \sigma_3 \sigma_2 = 0 \] (65)
Then, the cross-terms cancel and we obtain:

\[ \tilde{k}^{-2}(dS)^2 = (dX)^2 + (dY)^2 + (dZ)^2 \]  \hspace{1cm} (66)

The entropy, here, is a real number again.

The resulting equation of state has the mathematical form of the Euclidean distance \( d^2 := \tilde{k}^2(dS)^2 \). The entropy, as demanded, is invariant under rotation of the Euclidean frame of reference. As we will see, if one uses the flexibility of geometric algebra, one can generalize this argument to space-times of any dimensions, any signature, and even including arbitrarily curved space-times.

For instance, a thermodynamic system of special relativity would have \( X, Y, Z \) and \( T \) as its thermodynamic quantities. The equation of state, using the generators \( \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\} \) is:

\[ \tilde{k}^{-1} dS = \tilde{k}^{-1} f_0 dT + \gamma_1 dX + \gamma_2 dY + \gamma_3 dZ \]  \hspace{1cm} (67)

Here, both \( \tilde{k} \) and \( f \) are Lagrange multipliers. \( T \) is an extensive quantity with units \( s \) and it is conjugated with \( f \) having units \( s^{-1} \). Squaring the equation of state gives:

\[
\begin{align*}
\tilde{k}^{-2}(dS)^2 &= \tilde{k}^{-2} f^2_0 (dT)^2 + \gamma^2_1 (dX)^2 + \gamma^2_2 (dY)^2 + \gamma^2_3 (dZ)^2 \\
&+ \tilde{k}^{-1} f(\gamma_0 \gamma_1 + \gamma_1 \gamma_0) dT dX + \tilde{k}^{-1} f(\gamma_0 \gamma_2 + \gamma_2 \gamma_0) dT dY + \tilde{k}^{-1} f(\gamma_0 \gamma_3 + \gamma_3 \gamma_0) dT dZ \\
&+ (\gamma_1 \gamma_2 + \gamma_2 \gamma_1) dX dY + (\gamma_1 \gamma_3 + \gamma_3 \gamma_1) dX dZ \\
&+ (\gamma_2 \gamma_3 + \gamma_3 \gamma_2) dY dZ
\end{align*}
\]  \hspace{1cm} (68)

The cross-terms cancel provided that the generators obey the following relations:

\[
\begin{align*}
\gamma^2_0 &= 1 \\
\gamma^2_1 &= -1 \\
\gamma^2_2 &= -1 \\
\gamma^2_3 &= -1 \\
\gamma_0 \gamma_1 + \gamma_1 \gamma_0 &= 0 \\
\gamma_0 \gamma_2 + \gamma_2 \gamma_0 &= 0 \\
\gamma_0 \gamma_3 + \gamma_3 \gamma_0 &= 0 \\
\gamma_1 \gamma_2 + \gamma_2 \gamma_1 &= 0 \\
\gamma_1 \gamma_3 + \gamma_3 \gamma_1 &= 0 \\
\gamma_2 \gamma_3 + \gamma_3 \gamma_2 &= 0
\end{align*}
\]  \hspace{1cm} (69) \hspace{1cm} (70) \hspace{1cm} (71) \hspace{1cm} (72) \hspace{1cm} (73) \hspace{1cm} (74) \hspace{1cm} (75) \hspace{1cm} (76) \hspace{1cm} (77) \hspace{1cm} (78) \hspace{1cm} (79)

We also pose \( c := \tilde{k}^{-1} f \), then, the equation of state is:
\[ \hat{k}^{-2}(dS)^2 = c^2(dT)^2 - (dX)^2 - (dY)^2 - (dZ)^2 \] (80)

Here, the entropy becomes a real number again.

Geometric thermodynamics is quite easy to construct, yet it is incredibly powerful. In the general case, one begins by defining an arbitrary non-commutative basis as follows:

\[ e_\mu \cdot e_\nu = \frac{1}{2}(e_\mu e_\nu + e_\nu e_\mu) = g_{\mu\nu} \] (81)

To define geometric thermodynamics as a system of statistical physics, one first defines \( n \) thermodynamic observables using the geometric basis. The statistical priors, such as \( E = \sum_{q \in \mathcal{Q}} E[q] \rho[q] \), are now simply multiplied with a generator \( e_i \) of the geometric algebra, yielding \( n \) equations:

\[ e_i X_i = \sum_{q \in \mathcal{Q}} e_i X_i[q] \rho[q] \] (82)

Then, by maximizing the entropy with these priors as the constraints and by using the method of the Lagrange multipliers, one will obtain a generalized non-commutative thermodynamics conservation relation instead of equation (32):

\[ dS = \hat{k} e_1 dX_1 + \cdots + \hat{k} e_n dX_n \] (83)

We note that had we instead selected a geometric algebra such that the generators are commutative, then one would recover, as a special case, the traditional conservation relation of energy found in statistical physics. Explicitly, posing the properties of the generators \( e_1, ..., e_n \) to be commutative:

\[ e_i^2 = 1 \] (84)
\[ e_i e_j = e_j e_i \] (85)

one obtains the relation \( dS = \hat{k} dX_1 + \cdots + \hat{k} dX_n \), which is of the same mathematical form as equation (32). Therefore, geometric thermodynamics is indeed a generalization of thermodynamics; a fact quite important to what we are trying to achieve. Indeed, statistical physics as long been considered by many to be our physical theory least likely to be falsified within its domain of applicability. Deriving the laws of physics as a consequence of generalizing it thus benefits from increased robustness.
3.9 Recap: Geometric algebra

A geometric algebra $\mathcal{G}$ is a ring equipped with algebraic generators that satisfy the generator relation:

$$e_\mu \cdot e_\nu = \frac{1}{2}(e_\mu e_\nu + e_\nu e_\mu) = g_{\mu\nu}$$  \hspace{1cm} (86)

The generators form a basis that includes the generators themselves and all arrangements of their wedge products. For instance, an algebra of four generators $\{e_0, e_1, e_2, e_3\}$ form the complete basis:

<table>
<thead>
<tr>
<th>basis elements</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>${1,}$</td>
<td>grade-0</td>
</tr>
<tr>
<td>$e_0, e_1, e_2, e_3,$</td>
<td>grade-1</td>
</tr>
<tr>
<td>$e_0e_1, e_0e_2, e_0e_3, e_1e_3, e_2e_3.$</td>
<td>grade-2</td>
</tr>
<tr>
<td>$e_0e_1e_2, e_0e_1e_3, e_0e_2e_3, e_1e_2e_3,$</td>
<td>grade-3</td>
</tr>
<tr>
<td>$e_0e_1e_2e_3}$</td>
<td>grade-4</td>
</tr>
</tbody>
</table>

Poly-vectors of $\mathcal{G}$ can be constructed as a linear combination of these basis elements. For instance:

**Vectors and poly-vectors:**

<table>
<thead>
<tr>
<th>example</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v:= 1$</td>
<td>0-vector, or scalar</td>
</tr>
<tr>
<td>$v:= 3e_0 + 4e_1$</td>
<td>1-vector, or vector</td>
</tr>
<tr>
<td>$v:= 3e_0e_3 + 2e_2e_1$</td>
<td>2-vector, or bivector</td>
</tr>
<tr>
<td>$v:= 5e_0e_1e_2$</td>
<td>3-vector, or trivector</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$v:= 2e_0e_1 \ldots e_k$</td>
<td>k-vector</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$v := 1 + 2e_0 + 5e_2e_3$</td>
<td>poly-vector</td>
</tr>
</tbody>
</table>

We note that the k-vectors are a linear combination of basis elements of the same grade, whereas a poly-vector mixes different grades.

If the scalars multiplying the basis elements of the poly-vectors are elements of the reals, then the algebra is called a real geometric algebra $\mathcal{G}(\mathbb{R})$, and if they are complex then the algebra is called a complex geometric algebra $\mathcal{G}(\mathbb{C})$. For instance:

$$v := r + r_0e_0 + r_1e_1 + r_2e_2 + r_{01}e_0e_1 + \ldots \hspace{1cm} \text{where } r, r_0, r_1, r_{01}, \ldots \in \mathbb{R}$$  \hspace{1cm} (98)
is a real algebra $\mathcal{G}(\mathbb{R})$, and

$$v := z + z_0 e_0 + \ldots \quad \text{where } z, z_0, \cdots \in \mathbb{C} \quad (99)$$

is a complex algebra $\mathcal{G}(\mathbb{C})$.

We use numbered indices to denote the number of generators of $\mathcal{G}$. For instance if $\mathcal{G}$ has four generators $\{e_0, e_1, e_2, e_3\}$ we denote the algebra by $\mathcal{G}_4$ generally, or $\mathcal{G}_4(\mathbb{C})$ if the algebra is complex with four generators, or $\mathcal{G}_4(\mathbb{R})$ if the algebra is real with four generators.

Furthermore, if the generator relation is orthogonal;

$$\gamma_i \cdot \gamma_j = \frac{1}{2} (\gamma_i \gamma_j + \gamma_j \gamma_i) = \eta_{ij} \quad (100)$$

where $\eta_{ij}$ is the signature of the generator relation, for instance:

$$\eta_{ij} = \text{diag}(+, -, -, -) \quad (101)$$

then,

$$\gamma_0 \gamma_0 = 1 \quad (102)$$
$$\gamma_1 \gamma_1 = -1 \quad (103)$$
$$\gamma_2 \gamma_2 = -1 \quad (104)$$
$$\gamma_3 \gamma_3 = -1 \quad (105)$$
$$\gamma_0 \gamma_1 + \gamma_1 \gamma_0 = 0 \quad (106)$$
$$\gamma_0 \gamma_2 + \gamma_2 \gamma_0 = 0 \quad (107)$$
$$\gamma_0 \gamma_3 + \gamma_3 \gamma_0 = 0 \quad (108)$$
$$\gamma_1 \gamma_2 + \gamma_2 \gamma_1 = 0 \quad (109)$$
$$\gamma_1 \gamma_3 + \gamma_3 \gamma_1 = 0 \quad (110)$$
$$\gamma_2 \gamma_3 + \gamma_3 \gamma_2 = 0 \quad (111)$$

For real algebras, we add an additional indice $Cl_{n,m}(\mathbb{R})$, where $n$ is the number of generators squaring to 1, and $m$ is the number of generators squaring to $-1$. In the case of signature $\text{diag}(+, -, -, -)$, the algebra is $Cl_{1,3}(\mathbb{R})$.

The geometric product of two poly-vectors $v$ and $u$ is:

$$vu = v \cdot u + v \wedge u \quad (112)$$

It can be calculated quite simply by expanding the product and applying the generator relation to simplify the expression. For instance, consider the following 1-vectors:

$$v := a e_0 + b e_1 \quad (113)$$
$$u := c e_0 + d e_1 \quad (114)$$
Then, the geometric product is

\[ \mathbf{vu} = (ae_0 + be_1)(ce_0 + de_1) \]  
\[ = (ace_0 + ade_0e_1 + be_1ce_0 + be_1de_1) \]  
\[ = ac + ade_0e_1 - db e_0e_1 + bd \]  
\[ = v \cdot u + v \wedge u \]  

(115)

(116)

(117)

(118)

(119)

One can construct higher grades of the basis using the antisymmetrization. Using the gamma matrices \( \{ \gamma_0, \gamma_1, \gamma_2, \gamma_3 \} \) as a representation, the complete basis contains:

1. The identity matrix: 1
2. 4 matrices: \( \gamma_i \)
3. 6 matrices \( \sigma_{\mu\nu} = \frac{1}{2} [\gamma_\nu, \gamma_\mu] \)
4. 4 matrices \( \sigma_{\mu\nu\rho} = \frac{1}{6} [\gamma_\mu, \gamma_\nu, \gamma_\rho] \)
5. 1 matrix \( \sigma_{\mu\nu\rho\delta} = \frac{1}{24} [\gamma_\mu, \gamma_\nu, \gamma_\rho, \gamma_\delta] \)

3.10 Recap: Quantum thermodynamics

Since the generators of any finite geometric algebra \( \mathcal{G} \) have matrix representations, we will find it useful to recall the thermodynamics of quantum observables (a.k.a partition functions with matrices and operators). We consider the case of finite operators.

Let \( \hat{H} \) be a self-adjoint operator. If \( \hat{H} \) can be represented by an \( n \times n \) matrix, then \( \hat{H} \) can be diagonalized to an \( n \times n \) matrix:

\[ \hat{H} = U \begin{pmatrix} E_1 & 0 & \ldots & 0 \\ 0 & E_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & E_n \end{pmatrix} U^\dagger \]  

(120)

We say that \( E_1 \) to \( E_n \) are the eigenvalues of \( \hat{H} \) and we note its eigenbasis as \( |\phi_1\rangle \) to \( |\phi_n\rangle \).

A linear superposition of the eigenbasis is a pure quantum state:

\[ |\psi\rangle = a_1 |\phi_1\rangle + \ldots a_n |\phi_n\rangle , \quad \text{where } a_1, \ldots, a_n \in \mathbb{C} \]  

(121)

The complex coefficients \( a_1, \ldots, a_n \) are probability amplitudes. The corresponding probability \( \rho_i \) is obtained by taking the magnitude of the probability amplitude:
\[ \rho_i := (a_i)(a_i)^* \] (122)

We note that, as a probability, we require that \( 1 = \sum_{i=0}^{n} \rho_i \) and that \( \forall \rho_i (\rho_i \geq 0) \). The density matrix \( \hat{\rho} \) for a pure state is:

\[ \hat{\rho} := |\psi\rangle \langle \psi| \] (123)

or explicitly,

\[
\hat{\rho} = \begin{pmatrix}
(a_0)(a_0)^* & (a_0)(a_1)^* & \cdots & (a_0)(a_n)^* \\
(a_1)(a_0)^* & (a_1)(a_1)^* & \cdots & (a_1)(a_n)^* \\
\vdots & \vdots & \ddots & \vdots \\
(a_n)(a_0)^* & (a_n)(a_1)^* & \cdots & (a_n)(a_n)^*
\end{pmatrix}
\] (124)

The Von Neumann entropy is:

\[ S = \text{Tr}[\hat{\rho} \ln \hat{\rho}] \] (125)

The trace has cyclic invariance \( \text{Tr}[ABC] = \text{Tr}[CAB] = \text{Tr}[BAC] \). The matrix logarithm of a diagonalizable matrix is \( \ln A = U \text{Diag}(\ln D) U^\dagger \), where \( \text{ln} D \) is the diagonal element-by-element logarithm. Using these identities, we can calculate the entropy by diagonalizing \( \hat{\rho} = U \hat{d} U^\dagger \):

\[
S = \text{Tr}[U \hat{d} U^\dagger \text{ln}[\hat{d}] U^\dagger] \\
= \text{Tr}[U^\dagger U \hat{d} U^\dagger \text{ln}[\hat{d}]] \\
= \text{Tr}[\hat{d} \text{ln} \hat{d}]
\] (126)

The entropy of \( \hat{\rho} \) is the entropy of its eigenvalues:

\[ S = \sum_{i=1}^{n} \lambda_i \ln \lambda_i \] (129)

The Von Neumann entropy of a pure state is 0. Thus, it is often said that the Von Neumann entropy measures the informational departure of a mixed state from a pure state.

Measurement-entropy: A projective (‘collapse-causing’ measurement) of \( \hat{H} \) on \( |\psi\rangle \) projects \( |\psi\rangle \) to one eigenbasis \( |\phi_i\rangle \) in \( \{ |\phi_1\rangle, \ldots, |\phi_n\rangle \} \) with probability \( \rho_i \). Since the projective measurement involves the random selection of one element \( |\phi_i\rangle \) out of a set of possible measurement outcomes \( \{ |\phi_1\rangle, \ldots, |\phi_1\rangle \} \), it fits the
definition of an information-bearing message in the Shannon sense. The Shannon entropy, in this case, quantifies the amount of information gained by knowing which eigenbasis was randomly selected by the act of measurement. The Shannon entropy of a projective measurement on $|\psi\rangle$ is thus given by:

$$H = - \sum_{i=1}^{n} ((a_i)(a_i)^*) \ln[(a_i)(a_i)^*]$$  \hfill (130)

This Shannon entropy agrees with the density matrix approach. Indeed, post-measurement, the density matrix $\hat{\rho}$ is a mixture:

$$\hat{m} = \begin{pmatrix}
(a_1)(a_1)^* & 0 & \ldots & 0 \\
0 & (a_2)(a_2)^* & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & (a_n)(a_n)^*
\end{pmatrix}$$  \hfill (131)

The difference in entropy between the pre-measurement pure state $\hat{\rho}$ and the post-measurement mixture $\hat{m}$ is equal to:

$$H = -(\text{Tr}[\hat{m} \ln \hat{m}] - \text{Tr}[\hat{\rho} \ln \hat{\rho}])$$  \hfill (132)

$$= - \text{Tr}[\hat{m} \ln \hat{m}]$$  \hfill (133)

$$= - \sum_{i=1}^{n} ((a_i)(a_i)^*) \ln[(a_i)(a_i)^*]$$  \hfill (134)

This is the same as the Shannon entropy obtained by equation 130.

Unitary transformations: One can change the state $|\psi\rangle$ by applying a unitary transformations $U$. $U$ is unitary if its conjugate transpose is also its inverse $U^* = U^{-1}$. By convention, we denote the inverse of $U$ as $U^\dagger$. The properties are $U^\dagger U = UU^\dagger = \hat{1}$. A general $2 \times 2$ unitary transformation is:

$$U = \begin{pmatrix}
\alpha & \beta \\
-e^{i\varphi} \beta^* & e^{i\varphi} \alpha^*
\end{pmatrix}$$  \hfill (135)

Applying it to $|\psi\rangle = a_1 |\phi_1\rangle + a_2 |\phi_2\rangle$, we get:

$$|\psi'\rangle = U |\psi\rangle$$  \hfill (136)

$$= \begin{pmatrix}
\alpha & \beta \\
-e^{i\varphi} \beta^* & e^{i\varphi} \alpha^*
\end{pmatrix} |\psi\rangle$$  \hfill (137)

$$= \begin{pmatrix}
\alpha & \beta \\
-e^{i\varphi} \beta^* & e^{i\varphi} \alpha^*
\end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$  \hfill (138)

$$= (\alpha a_1 + \beta a_2) |\phi_1\rangle + (-e^{i\varphi} \beta^* a_1 + e^{i\varphi} \alpha^* a_2) |\phi_2\rangle$$  \hfill (139)
The Shannon entropy of a measurement of $|\phi'\rangle$ along the eigenbasis is:

$$H = -(\alpha a_1 + \beta a_2)(\alpha a_1 + \beta a_2)^* \ln[(\alpha a_1 + \beta a_2)(\alpha a_1 + \beta a_2)^*]$$

$$+ (-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)^* \ln[(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)^*]$$

(140)

Let us now show that this entropy agrees with the matrix density approach. The density matrix of $|\phi'\rangle$ is:

$$\hat{\rho} = \begin{pmatrix} (\alpha a_1 + \beta a_2)(\alpha a_1 + \beta a_2)^* & (\alpha a_1 + \beta a_2)(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)^* \\ (-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)(\alpha a_1 + \beta a_2)^* & (-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)^* \end{pmatrix}$$

(141)

Post-measurement, the density matrix is:

$$\hat{\rho}_m = \begin{pmatrix} (\alpha a_1 + \beta a_2)(\alpha a_1 + \beta a_2)^* & 0 \\ 0 & (-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)(-e^{i\Phi} a_1 + e^{i\Phi} \alpha^* a_2)^* \end{pmatrix}$$

(142)

The entropy of $\hat{\rho}_m$ is equal to $H$.

**Thermal states:** In the case of a thermally prepared state, the probability measure is:

$$\rho_i = \frac{1}{Z} e^{-\beta E_i}$$

(143)

It then follows that:

$$a_i = \sqrt{e^{i\Phi} \frac{1}{Z} e^{-\beta E_i}} = e^{i\frac{\Phi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_i}}$$

(144)

where $e^{i\frac{\Phi}{2}}$ is a complex phase, such that:

$$(a_i)(a_i)^* = e^{i\frac{\Phi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_i} e^{-i\frac{\Phi}{2}}} \sqrt{\frac{1}{Z} e^{-\beta E_i}}$$

$$= \frac{1}{Z} e^{-\beta E_i}$$

(145)

Thus, the thermal quantum state is written as:

$$|\psi_{\text{thermal}}\rangle = \left( e^{i\frac{\Phi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_1}} |\phi_1\rangle + \cdots + e^{i\frac{\Phi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_n}} |\phi_n\rangle \right)$$

(147)
Injecting $\rho_i = \frac{1}{Z} \exp(-\beta E_i)$ into the Boltzmann definition of entropy one obtains the quantum version of the thermodynamic equation of state:

$$S = -k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) \ln \left( \frac{1}{Z} e^{-\beta E_i} \right)$$

(148)

$$= -k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) (-\beta E_i - \ln Z)$$

(149)

$$= k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) (\beta E_i + \ln Z)$$

(150)

$$= k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \beta E_i \right) + k_B \ln Z \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right)$$

(151)

$$= k_B \beta \sum_{i=1}^{n} \left( E_i \frac{1}{Z} e^{-\beta E_i} \right) + k_B \ln Z$$

(152)

Posing $\overline{E} := \sum_{i=1}^{n} E_i \exp(-\beta E_i)/Z$, then:

$$S = k_B (\ln Z + \beta \overline{E})$$

(153)

Using the Von Neumann formalism, it is possible to obtain the same result, as follows. First, the partition function is defined as:

$$Z = \text{Tr} \left[ e^{-\beta \hat{H}} \right]$$

(154)

and the entropy as:

$$S = -\text{Tr} [\hat{\rho} \ln \hat{\rho}]$$

(155)

In the case of a thermal state, the density matrix is:

$$\hat{\rho} = \frac{1}{Z} e^{-\beta \hat{H}}$$

(156)

Then, injecting $\hat{\rho}$ into $S$, we get

$$S = -\text{Tr} \left[ \left( \frac{1}{Z} e^{-\beta \hat{H}} \right) \ln \left( \frac{1}{Z} e^{-\beta \hat{H}} \right) \right]$$

(157)

$$= -\text{Tr} \left[ \frac{1}{Z} e^{-\beta \hat{H}} (-\beta \hat{H} - \ln Z) \right]$$

(158)

$$= \text{Tr} \left[ \frac{1}{Z} e^{-\beta \hat{H}} (\beta \hat{H} + \ln Z) \right]$$

(159)

$$= \beta \text{Tr} \left[ \frac{\hat{H}}{Z} e^{-\beta \hat{H}} \right] + \ln Z \frac{1}{Z} \text{Tr} \left[ e^{-\beta \hat{H}} \right]$$

(160)
Posing \( \langle \hat{H} \rangle := \text{Tr} \left[ \hat{H} e^{\beta \hat{H}} / Z \right] \), the entropy is:

\[
= \beta \langle \hat{H} \rangle + \ln Z \tag{161}
\]

It has the same mathematical form as the fundamental relation of thermodynamics (Equation 32).

4 Geometric statistical Physics

Our goal with geometric statistical physics is to recover the structure of space-time (Lorentz invariance, general invariance, speed of light, metric interval, etc.) strictly using the facilities of statistical physics (entropy, partition function, observables, etc.). We would then say that the structure of space-time is an emergent bulk property of the appropriately described statistical ensemble. In this optic, even the speed of light will not be taken as an axiom, but will instead be a property emergent from the construction. How will we do that? First, we have to interpret the speed of light as a tool to hide information in space-time. Specifically, the speed of light hides information regarding events whose interval to the observer are time-like or space-like. Interpreted as such, we can then use the entropy in statistical physics to achieve the same purpose as the speed of light (hide information), provided that we "place" this entropy at the appropriate position in the system. In this context, the speed of light will then be emergent as a Lagrange multiplier, constant throughout the ensemble, and as a result of maximizing the appropriately positioned entropy.

We note that attributing an entropy to events separated by a horizon to connect to thermodynamics has been done since at least 1973 by J.D. Bekenstein\cite{19}. Furthermore, from G. W. Gibbons and S. W. Hawking’s 1977 article\cite{57}, I quote:

"An observer in these models will have an event horizon whose area can be interpreted as the entropy or lack of information of the observer about the regions which he cannot see."

The part missing to complete a full entropic picture of space-time, we suggest, is to apply the same line of reasoning to configurations of time-like and space-like separated events. For instance, we can imagine an observer \( O \) whose "visibility" is defined by the usual light cone of special relativity. We can describe this light cone entirely using notions of statistical physics by analyzing the number of configurations of events outside the light-cone and associating it to an entropy. Indeed, to prevent faster-than-light communication, all possible configurations of events outside the light-cone must be of maximal entropy (i.e., \( \approx \) equally likely within the priors) to be void of information from the perspective of \( O \). This entropy thus hides events that \( O \) cannot see.

The same reasoning can be applied to the future of \( O \). Indeed, to prevent \( O \) from knowing its future, future events must also be void of information from \( O \)’s perspective and thus be at maximal entropy.

35
Table 4: The physical quantities of the geometric ensemble

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x[q] )</td>
<td>space</td>
<td>[meter]</td>
<td>extensive</td>
</tr>
<tr>
<td>( \tilde{k} )</td>
<td>entropic repetency</td>
<td>[1/meter]</td>
<td>intensive</td>
</tr>
<tr>
<td>( \overline{x} )</td>
<td>bulk space</td>
<td>[meter]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>( t[q] )</td>
<td>time</td>
<td>[second]</td>
<td>extensive</td>
</tr>
<tr>
<td>( f )</td>
<td>entropic frequency</td>
<td>[1/second]</td>
<td>intensive</td>
</tr>
<tr>
<td>( \overline{t} )</td>
<td>bulk time</td>
<td>[second]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>( c := f/\tilde{k} )</td>
<td>entropic speed</td>
<td>[meter/second]</td>
<td>intensive</td>
</tr>
</tbody>
</table>

Using this strategy, we can construct an ensemble of statistical physics that recovers the structure of space-time in the bulk. We will now describe the physical quantities relevant to geometric thermodynamics.

**Definition 8** (Physical quantities). *As we derive an ensemble of events, two physical quantities will be introduced as Lagrange multipliers. They are: 1) the entropic repetency \( \tilde{k} \) and 2) the entropic frequency \( f \). Specifically, \( \tilde{k} = k/2\pi = 1/\lambda \), where \( k \) is the wave-number and \( \lambda \) is the wavelength.*

These quantities are the conjugated variables to a distance \( x \) and time \( t \), respectively. By convention, we prefix the Lagrange multipliers with the word "entropic", and its averaged conjugated quantity will be prefixed with the word "bulk". \( \tilde{k} \) and \( f \) are both intensive properties, whereas \( x \) and \( t \) are extensive. Indeed, a process taking 1 min followed by a process taking 2 min takes a total of 3 min (extensive). For the \( x \) quantity; walking 1 meter followed by walking 2 meters implies one has walked a total of 3 meters (extensive). Adding or removing clocks from a group of clocks ticking at a frequency \( f \) (say once per second) has no impact on the frequency of the other elements of the group (intensive). The same argument applies to the entropic repetency (intensive). The units of \( \tilde{k} \) are \( m^{-1} \), the units of \( x \) are the meters, the units of \( t \) are the seconds, and the units of \( f \) are \( s^{-1} \). Finally, we define the speed of light as the ratio of the Lagrange multipliers \( c := f/\tilde{k} \). These quantities are summarized in Table 4.

We note that the temperature \((k_B T = 1/\beta)\) has no central role in geometric statistical physics. In fact, unlike the speed of light, space-time in general (excluding horizons) does not have a constant temperature and therefore describing space-time as a thermodynamic system (using temperature, energy, and entropy) would be inappropriate as the system would be outside equilibrium. However, with our strategy, it is precisely because the speed of light is constant and that faster-than-light communication is impossible that the speed of light can take the role normally assumed by the temperature as the Lagrange multiplier of the ensemble. By using the speed of light instead of the temperature as the Lagrange multiplier, the ensemble applies an entropy to all of space-time, with or without horizons, and thus determines its complete structure.

36
To apply geometric statistical physics to the axioms of science, we will pose our final assumption:

**Assumption 4** (The fundamental assumption of 'geometric substance'). We will equip the ensemble of experiments with the observables of geometric thermodynamics (Definition 13). We will call the experiments of this new ensemble (Definition 14), geometric events \( p \) (Definition 10).

The quantities of statistical physics that will be augmented to poly-vectors will be prefixed with the term *geometric*. Geometric quantities contain basis elements within their expression to enforce the suitable non-commutative relation between the quantities of the expression. After we pose basic definitions, we will apply the usual machinery of statistical physics in order to maximize the entropy of the ensemble of events by using the method of the Lagrange multipliers. Specifically, we will derive:

- The geometric density \( g[p] \) (Definition 11).
- The geometric entropy \( S \) (Definition 12).
- The geometric equation of state \( dS \) (Theorem 1).
- The geometric measure and the partition function \( Z \) (Theorem 2).

First, let us define a space-time event.

**Definition 9** (Space-time event). A space-time event is, in flat space-time, a 1-vector of \( Cl_{1,3}(\mathbb{R}) \):

\[
p := \gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3
\]

(162)

The quantities \( \{X_0, X_1, X_2, X_3\} \) are elements of \( \mathbb{R} \). By convention, the first term \( X_0 \) denotes the time dimension and the next 3 terms denote the space dimensions. In curved space-time, whose generators are \( \{e_0, e_1, e_2, e_3\} \), then, a space-time event is:

\[
p := e_0 X_0 + e_1 X_1 + e_2 X_2 + e_3 X_3
\]

(163)

Now, let us define a geometric event.

**Definition 10** (Geometric event). A geometric event is a generalization of a space-time event. In the general case, a geometric event is a poly-vector. Using the basis of \( Cl_{1,3}(\mathbb{R}) \):

\[
\{1, e_0, e_1, e_2, e_3, e_0e_1, e_0e_2, e_0e_3, e_1e_2, e_1e_3, e_2e_3, e_0e_1e_2, e_0e_2e_3, e_1e_2e_3
\}

(164)

(165)

(166)

(167)

(168)
where

$$g_{\mu \nu} = \frac{1}{2} (e_\mu e_\nu + e_\nu e_\mu)$$

(169)

the most general geometric event in $Cl_{1,3}(\mathbb{R})$ is:

$$p := G + X^i e_i + A^{ij} e_i e_j + V^{ijk} e_i e_j e_k + U e_0 e_1 e_2 e_3 \equiv \sum_{i=1}^{2^n} E_i X_i$$

(170)

We note:

- Geometric events do not need to use all elements of the basis to qualify as such. For instance, space-time events are specific types of geometric events.
- A geometric event can be expressed as a $2^n$-tuple; where $n$ is the number of basis elements of the algebra:

$$(G, X_0, X_1, X_2, X_3, A_{01}, A_{02}, A_{03}, A_{12}, A_{13}, A_{23}, V_{012}, V_{013}, V_{023}, V_{123}, U)$$

(171)

**Definition 11** (Geometric density). Let $g[p]$ be a poly-vector valued function:

$$g : D \rightarrow G \quad p \mapsto p$$

(172)

Here, $g$ takes an element of $D$ and maps it to a geometric event. If the sum of $g[p]$ over all $p$ elements of $D$ converges:

$$\Omega = \sum_{p \in D} g[p]$$

(173)

where $\Omega \in G$, then $g[p]$ is a geometric density. Finally, we note by $\hat{g}[p]$ the matrix representation of $g[p]$.

**Definition 12** (Geometric entropy). Let

$$p = \left\{ g : D \rightarrow G \mid \sum_{p \in D} g[p] = \Omega \right\}$$

(174)

We define the geometric entropy as:

$$S : \{D\} \times p \rightarrow G \quad (D, g) \mapsto -\Omega^{-1} k_B \sum_{p \in D} g[p] \ln g[p]$$

(175)
The term $\Omega^{-1}$ is added to normalize the entropy.

**Definition 13** (Geometric bulk). The constraints of the geometric ensemble are:

$$E_i X_i = \sum_{p \in D} E_i X_i[p] g[p]$$  \hspace{1cm} (176)

Here we note that we have simply taken the usual expression of a prior of statistical physics (Equation 25) and we have equipped it with a basis of $G$: noted as $E_i$. For instance, the geometric algebra $Cl_{1,3}(\mathbb{R})$ admits $2^{1+3} = 16$ basis elements $\{E_1, \ldots, E_{16}\}$. Therefore, an ensemble constructed from this algebra admits $2^n$ statistical priors. Finally, we note that the functions $X_i[p]$ are maps $X_i : D \rightarrow \mathbb{R}$ where $X_i[p]$ returns the value of the $i$th element of the $2^n$-tuple of $g[p]$. The terms $X_i$ are averages.

**Definition 14** (Geometric ensemble). A geometric ensemble $\mathcal{E}_G$ is a probability pseudo-space composed of the 2-tuple:

$$\mathcal{E}_G := \{D, g\}$$  \hspace{1cm} (177)

where $D$ is the sample space and $g : D \rightarrow G$ is the geometric density.

Remarks:

- A case of special interest is the geometric ensemble in $3 + 1$ space-time over the reals: $\mathcal{E}_{Cl_{1,3}(\mathbb{R})}$.

- Since the set of all experiments is countable (Definition 3: $D = \text{Dom}[UTM]$), and the set of all events is uncountable, then some events are not experiments. At most, only events constructed from the computable reals are experiments.

- Perhaps counter-intuitively, there exist countable sets with well-defined notions of continuity. For instance, the computable real numbers form a real closed field. Consequently, notions of continuity, derivative and anti-derivative are definable on the computable numbers. Intuitively, as there exists a computable number between any two computable numbers as well as a distance function $d = |x - y|$ on the computable numbers, then there exists maps from open sets to open sets within some neighborhood $\epsilon$ and thus the notion of continuity is well-defined, even if said set is countable.

- To define, say, partial differential equations (PDE) of experiments (from a countable set containing at most the computable numbers) in a rigorous manner, we would need to use computable analysis, as opposed to real analysis. For our purposes, however, the difference is merely a formality. Indeed, computable analysis is very similar to real analysis (the identities relevant to physics regarding derivative/anti-derivative are the same).
• Going forward, we simply keep in mind that most real numbers (the non-computable numbers) and some (most?) solutions of PDE (non-computable solutions) which might mathematically exist in real analysis, are in practice experimentally 'unreachable/non-producible' in a finite number of steps, and would thus not exist as solutions using computable analysis. The difference is inconsequential for the kind of theorems we prove in this manuscript.

• From the notational standpoint and for our purposes, the difference between defining the entropy over the computable reals, versus the reals, is merely the difference in using $S = -\sum \rho \ln \rho$ (for the countable set) versus $S = -\int \rho \ln \rho d\rho$ (for the uncountable set) to define the entropy.

• In this manuscript, to stay consistent with our computer theoretic origins, we will sum over the computable reals, unless otherwise stated.

To derive the geometric ensemble, we will assume the following is permitted: Instead of creating an ensemble of $\mathbb{M}$ (a manifest) selected over $\mathbb{W}$ (the set of all manifests), we create $n$ ensembles of $p$ (an experiment) selected over $\mathbb{D}$ (the domain of science). In this case, the ensemble $\mathbb{M} \in \mathbb{W}$ is the grand canonical ensemble to $n$ canonical ensembles $p \in \mathbb{D}$. At any point, should we prefer to work with $M \in \mathbb{W}$, rather than with $n$ systems of $p \in \mathbb{D}$, we can redress to a grand-canonical ensemble simply by introducing $\mu N(M)$ as a thermodynamic observable in the grand-canonical ensemble and summing $\mathbb{M} \in \mathbb{W}$ instead of $p \in \mathbb{D}$. Specifically, the assumption is that $\mu N(M)$ is a valid thermodynamic observable of a manifest. As this assumption is about experiments, and geometric events are experiments equipped with additional structure, then we will also inherit this assumption for geometric ensembles. Using geometric events, we will now create a canonical ensemble by summing over $p \in \mathbb{D}$.

**Theorem 1** (Geometric equation of state). *In the general case, the equation of state of the geometric ensemble is:*

$$dS[g] = \lambda d\Omega[g] + \sum_{i=1}^{2^n+1} \lambda_i E_i dX_i[g]$$  \hspace{1cm} (178)

**Proof.** We use the method of the Lagrange multipliers to find the maximum of the geometric entropy. To make use of matrix tools, we will work with the matrix representations of the geometric vectors. Therefore, we will note $S, \Omega, \hat{g}, \hat{E}_i$ as the matrix representation of $S, \Omega, \hat{g}, \hat{E}_i$, respectively.

1. There are $2^n + 1$ constraints:

$$\hat{\Omega} = \sum_{p \in \mathbb{D}} \hat{g}[p] \hspace{1cm} (179)$$

$$\hat{E}_i \hat{X}_i = \sum_{p \in \mathbb{D}} \hat{E}_i \hat{X}_i[p] \hat{g}[p] \hspace{1cm} (180)$$

40
2. The Lagrange equation to maximize is:

\[ \tilde{L}[\hat{g}, \lambda, \lambda_1, \ldots, \lambda_{2^n}] = \]

\[ -\hat{\Omega}^{-1} k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] - \lambda \left( -\hat{\Omega} + \sum_{p \in D} \hat{g}[p] \right) - \sum_{i=1}^{2^n} \lambda_i \left( -\hat{E}_i X_i + \sum_{p \in D} \hat{E}_i X_i[p] \hat{g}[q] \right) \]

where \( \lambda, \lambda_1, \ldots, \lambda_{2^n} \) are Lagrange multipliers elements of \( \mathbb{R} \), and \( \hat{\Omega} \) and \( \hat{E}_i X_i \) are matrices elements of \( \mathbb{C}^{n \times n} \).

3. Maximizing \( \tilde{L} \) is done by taking its gradient and posing it equal to \( \tilde{0} \):

\[ \nabla \tilde{L}[\hat{g}, \lambda, \lambda_1, \ldots, \lambda_{2^n}] = \tilde{0} \]  

Consequently, the equation to solve is:

\[ \tilde{0} = \nabla \left( -\hat{\Omega}^{-1} k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] + \lambda \hat{\Omega} - \lambda \sum_{p \in D} \hat{g}[p] + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i - \sum_{i=1}^{2^n} \sum_{p \in D} \lambda_i \hat{E}_i X_i[p] \hat{g}[q] \right) \]  

(183)

4. The gradient of the entropy is:

\[ \nabla \left( -\hat{\Omega}^{-1} k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] \right) = \nabla \left( -\hat{\Omega} + \lambda \sum_{p \in D} \hat{g}[p] - \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i + \sum_{i=1}^{2^n} \sum_{p \in D} \lambda_i \hat{E}_i X_i[p] \hat{g}[q] \right) \]

(184)

5. We note immediately that \( \hat{\Omega}, X_i, X_i[p] \) and \( \hat{E}_i \) are not variables of \( \tilde{L}[\hat{g}, \lambda, \lambda_1, \ldots, \lambda_{2^n}] \), consequently, the gradient of these variables vanishes. We obtain:

\[ \nabla \left( -k_B \hat{\Omega}^{-1} \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] \right) = -\hat{\Omega} \nabla \lambda - \sum_{i=1}^{2^n} \hat{E}_i X_i \nabla \lambda_i + \nabla \left( \lambda \sum_{p \in D} \hat{g}[p] + \sum_{i=1}^{2^n} \sum_{p \in D} \lambda_i \hat{E}_i X_i[p] \hat{g}[q] \right) \]

(185)
6. Distributing the gradient over the last term, we obtain:

\[
\nabla \left( -\hat{\Omega}^{-1}k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] \right) = -\hat{\Omega} \nabla \lambda - \sum_{i=1}^{2^n} \hat{E}_i \nabla \lambda_i \\
+ \sum_{p \in D} \hat{g}[p] \nabla \lambda + \lambda \nabla \sum_{p \in D} \hat{g}[p] \\
+ \sum_{i=1}^{2^n} \sum_{p \in D} \hat{E}_i X_i[p] \hat{g}[p] \nabla \lambda_i + \sum_{i=1}^{2^n} \lambda_i \nabla \left( \sum_{p \in D} \hat{E}_i X_i[p] \hat{g}[p] \right) 
\]

(186)

7. We then make the following replacements:

\[
\nabla \left( -\hat{\Omega}^{-1}k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p] \right) \rightarrow \nabla \hat{S}[\hat{g}] 
\]

(187)

\[
\nabla \left( \sum_{p \in D} \hat{E}_i X_i[p] \hat{g}[p] \right) \rightarrow \hat{E}_i \nabla X_i[\hat{g}] 
\]

(188)

\[
\nabla \sum_{p \in D} \hat{g}[p] \rightarrow \nabla \hat{\Omega}[\hat{g}] 
\]

(189)

And we also replace the following expressions by their constraints:

\[
\sum_{p \in D} \hat{E}_i X_i[p] \hat{g}[p] = \hat{E}_i X_i 
\]

(190)

\[
\sum_{p \in D} \hat{g}[p] = \hat{\Omega} 
\]

(191)

Making the replacements, we obtain:

\[
\nabla \hat{S}[\hat{g}] = -\hat{\Omega} \nabla \lambda - \sum_{i=1}^{2^n} \hat{E}_i \nabla \lambda_i + \hat{\Omega} \nabla \lambda + \lambda \nabla \hat{\Omega}[\hat{g}] + \sum_{i=1}^{2^n} \hat{E}_i X_i \nabla \lambda_i + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i \nabla X_i[\hat{g}] 
\]

(192)

Some terms cancel, and we get:

\[
\nabla \hat{S}[\hat{g}] = \lambda \nabla \hat{\Omega}[\hat{g}] + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i \nabla X_i[\hat{g}] 
\]

(193)
8. Since $\hat{S}[\hat{g}]$, $\hat{\Omega}[\hat{g}]$ and $\hat{X}_i[\hat{g}]$ are functions of one variable, then their gradients is equal to their total derivatives:

$$d\hat{S}[\hat{g}] = \lambda d\hat{\Omega}[\hat{g}] + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i d\hat{X}_i[\hat{g}]$$  \hfill (194)
\begin{align*}
\sum_{p \in \mathcal{D}} \hat{E}_i X_i[p] \hat{g}[p] &= \hat{E}_i X_i & (199) \\
\sum_{p \in \mathcal{D}} \hat{g}[p] &= \hat{\Omega} & (200)
\end{align*}

We obtain:

\begin{align*}
\nabla \left( -\hat{\Omega}^{-1} k_B \sum_{p \in \mathcal{D}} \hat{g}[p] \ln \hat{g}[p] \right) &= -\hat{\Omega} \nabla \lambda - \sum_{i=1}^{2^n} \hat{E}_i X_i \nabla \lambda_i + \hat{\Omega} \nabla \lambda + \lambda \nabla \sum_{p \in \mathcal{D}} \hat{g}[p] \\
&+ \sum_{i=1}^{2^n} \hat{E}_i X_i[p] \nabla \lambda_i + \sum_{i=1}^{2^n} \lambda_i \nabla \left( \sum_{p \in \mathcal{D}} \hat{E}_i X_i[p] \hat{g}[p] \right) & (201)
\end{align*}

3. Some terms cancel:

\begin{align*}
\nabla \left( -\hat{\Omega}^{-1} k_B \sum_{p \in \mathcal{D}} \hat{g}[p] \ln \hat{g}[p] \right) &= \lambda \nabla \sum_{p \in \mathcal{D}} \hat{g}[p] + \sum_{i=1}^{2^n} \lambda_i \nabla \left( \sum_{p \in \mathcal{D}} \hat{E}_i X_i[p] \hat{g}[p] \right) & (202)
\end{align*}

4. We note that each function \( \hat{g} \) is dependent upon \( p \) and only \( p \). Thus, taking the gradient of this function is equal to taking its total derivative. Consequently, by distributing the gradient inside the summations, we obtain a differential form with \( |Q| \) differential terms, each independent of one another. The \( |Q| \) independent equations are:

\begin{align*}
-\hat{\Omega}^{-1} k_B \mathcal{D}(\hat{g}[p] \ln \hat{g}[p]) &= \lambda \mathcal{D} \hat{g}[p] + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] \mathcal{D} \hat{g}[p] & (203)
\end{align*}

Furthermore, as per the assumptions of this theorem, \( \hat{g}[p] \) is diagonalizable. Therefore:

\begin{align*}
-\hat{\Omega}^{-1} k_B \mathcal{D}(P^{-1} \hat{g}[q] P \ln[P^{-1} \hat{g}[q] P]) &= \lambda \mathcal{D}(P^{-1} \hat{g}[q] P) + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[q] \mathcal{D}(P^{-1} \hat{g}[q] P) & (204)
\end{align*}
We use the matrix logarithm identity for a diagonalizable matrix $\ln \hat{A} = P^{-1} \ln[A]P$, and we get:

$$-\hat{\Omega}^{-1} k_B \frac{d}{P^{-1}} g[p] PP^{-1} \ln[g[p]] P = \lambda d(P^{-1} g[p] P) + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] d(P^{-1} g[p] P)$$

(205)

$$P^{-1}(P \hat{\Omega}^{-1} P^{-1}) (-k_B d[\hat{g}[p] \ln \hat{g}[p]]) = \lambda d(P^{-1} g[p] P) + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] d(P^{-1} g[p] P)$$

(206)

$$-P \hat{\Omega}^{-1} k_B d[\hat{g}[p] \ln \hat{g}[p]] = \lambda d\hat{g}[p] + \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p] d\hat{g}[p]$$

(207)

Applying the total derivative to the entropy, we get:

$$-P \hat{\Omega}^{-1} k_B (\ln \hat{g}[p] d\hat{g}[p] + d\hat{g}[p]) = \lambda d\hat{g}[p] + \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p] d\hat{g}[p]$$

(208)

Eliminating $d\hat{g}[p]$:

$$-P \hat{\Omega}^{-1} k_B (\ln \hat{g}[p] + 1) = \lambda \hat{\Omega} + \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p]$$

(210)

Finally, solving for $\hat{g}[q]$, we get:

$$\ln \hat{g}[p] + 1 = \frac{\hat{\Omega} P^{-1}}{k_B} \left( -1 + \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p] \right)$$

(211)

$$\hat{g}[p] = \exp \left( -1 + \frac{\hat{\Omega} P^{-1}}{k_B} \left( -1 + \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p] \right) \right)$$

(212)

$$\hat{g}[p] = \exp \left( -1 - \frac{\hat{\Omega} P^{-1}}{k_B} \right) \exp \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i P \hat{E}_i P^{-1} X_i[p] \right)$$

(213)
Redressing the change of basis, we get:

\[ P^{-1} \hat{g}[p] P = P^{-1} \exp \left( -\hat{1} - \frac{P \hat{\Omega} P^{-1}}{\hat{k}_B} \right) \exp \left( -\frac{1}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{\Omega} \hat{E}_i P^{-1} X_i[p] \right) P \]

(214)

\[ \hat{g}[p] = \exp \left( -\hat{1} - \frac{\hat{\Omega}}{\hat{k}_B} \lambda \right) \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] \right) \]

(215)

5. To define \( \hat{Z} \), the partition function, we start as follows:

\[ \hat{\Omega} = \sum_{p \in D} \hat{g}[p] \]

(216)

\[ = \sum_{p \in D} \exp \left( -\hat{1} - \frac{\hat{\Omega}}{\hat{k}_B} \lambda \right) \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] \right) \]

(217)

\[ = \exp \left( -\hat{1} - \frac{\hat{\Omega}}{\hat{k}_B} \lambda \right) \sum_{p \in D} \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] \right) \]

(218)

Therefore,

\[ \hat{\Omega} \left( \sum_{p \in D} \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[p] \right) \right)^{-1} = \exp \left( -\hat{1} - \frac{\hat{\Omega}}{\hat{k}_B} \lambda \right) \]

(219)

which we define as the inverse of the partition function:

\[ \hat{Z}^{-1} := \hat{\Omega} \left( \sum_{p \in D} \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[q] \right) \right)^{-1} \]

(220)

6. Finally, the probability measure is:

\[ \hat{g}[q] = \frac{1}{\hat{Z}} \exp \left( -\frac{\hat{\Omega}}{\hat{k}_B} \sum_{i=1}^{2^n} \lambda_i \hat{E}_i X_i[q] \right) \]

(221)
Remark:

- In the case where $\hat{g}[q]$ is not diagonalizable, a probability measure may still exist, however, its expression is significantly more verbose as it would include an infinite Taylor series of non-commuting terms. Specifically, the step that would fail is $d(\hat{g}[q] \ln \hat{g}[q]) = \ln \hat{g}[q] d\hat{g}[q] + d\hat{g}[q]$. This equality holds only if $\ln \hat{g}[q]$, $\hat{g}[q]$ and $d\hat{g}[q]$ commute, or if they are simultaneously diagonalizable.

**Theorem 3** (Diagonalizable geometric entropy). Assume that the matrix representation of $\hat{g}[q]$ is diagonalizable. Then, the geometric equation of state is:

$$dS[\lambda_1, \ldots, \lambda_{2^n}] = 2^n \sum_{i=1}^{2^n} \lambda_i E_i dX_i[\lambda_1, \ldots, \lambda_{2^n}]$$ (222)

**Lemma 3.1** (Boltzmann entropy). The diagonal form of the geometric entropy is a system of equations, each corresponding to the usual Boltzmann entropy:

$$S_j = -k_B \sum_{p \in D} g_j[p] \ln g_j[p]$$ (223)

where $S_j$ and $g_j$ are the eigenvalues of the diagonalization of the matrix representation of the entropy.

*Proof.* We start with the definition of the geometric entropy (Definition 12). Its matrix representation is:

$$\hat{S} = -k_B \sum_{p \in D} \hat{g}[p] \ln \hat{g}[p]$$ (224)

Suppose $\hat{S}$ and $\hat{g}$ are $2 \times 2$ matrices. We diagonalize:
\[ P^{-1} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} P = -k_B \sum_{p \in D} P^{-1} \begin{pmatrix} g_1[p] & 0 \\ 0 & g_2[p] \end{pmatrix} P \ln P^{-1} \begin{pmatrix} g_1[p] & 0 \\ 0 & g_2[p] \end{pmatrix} P \] (225)

\[ P^{-1} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} P = -k_B \sum_{p \in D} P^{-1} \begin{pmatrix} g_1[p] & 0 \\ 0 & g_2[p] \end{pmatrix} P P^{-1} \begin{pmatrix} \ln g_1[p] & 0 \\ 0 & \ln g_2[p] \end{pmatrix} P \] (226)

\[ P^{-1} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} P = P^{-1} \left( -k_B \sum_{p \in D} \begin{pmatrix} g_1[p] & 0 \\ 0 & g_2[p] \end{pmatrix} \begin{pmatrix} \ln g_1[p] & 0 \\ 0 & \ln g_2[p] \end{pmatrix} \right) P \] (227)

\[ \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} = \begin{pmatrix} -k_B \sum_{p \in D} g_1[p] \ln g_1[p] & 0 \\ 0 & -k_B \sum_{p \in D} g_2[p] \ln g_2[p] \end{pmatrix} \] (228)

**Lemma 3.2.** The geometric entropy of a diagonalizable \( \hat{g}[q] \) is:

\[ \hat{S}[\lambda_1, \ldots, \lambda_{2^n}] = k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}] + \sum_{i=1}^{2^n} \lambda_i \hat{X}_i[\lambda_1, \ldots, \lambda_{2^n}] E_i \] (229)

**Proof.** For each equation of the system, we inject the diagonal geometric measure \( \hat{g} \) into \( S[g] \):

\[ \langle \hat{S} \rangle_j = -\langle \hat{\Omega}^{-1} \rangle_j k_B \sum_{p \in D} \frac{1}{\langle Z \rangle_j} \exp \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j \hat{X}_i[p] \right) \]

\[ \ln \left[ \frac{1}{\langle Z \rangle_j} \exp \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j \hat{X}_i[p] \right) \right] \] (230)

\[ \langle \hat{S} \rangle_j = -\langle \hat{\Omega}^{-1} \rangle_j k_B \sum_{p \in D} \frac{1}{\langle Z \rangle_j} \exp \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j \hat{X}_i[p] \right) \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j \hat{X}_i[p] - \ln \langle \hat{Z} \rangle_j \right) \] (231)
Since $\sum_{p \in D} \frac{1}{(Z_p)} \exp \left( - \frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j X_i[p] \right) = \langle \hat{\Omega} \rangle_j$, then:

$$\langle \hat{S} \rangle_j = k_B \langle \hat{\Omega}^{-1} \rangle_j \langle \hat{\Omega} \rangle_j \ln \langle Z \rangle_j$$

And since the left term is simply the definition of the average, we get:

$$\langle \hat{S} \rangle_j = k_B \langle \hat{\Omega}^{-1} \rangle_j \langle \hat{\Omega} \rangle_j \ln \langle Z \rangle_j + \langle \hat{\Omega}^{-1} \rangle_j \sum_{p \in D} \frac{1}{(Z_p)} \exp \left( - \frac{1}{k_B} \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j X_i[p] \right) \left( \sum_{i=1}^{2^n} \lambda_i \langle P \hat{\Omega} \hat{E}_i P^{-1} \rangle_j X_i[p] \right)$$

(232)

Finally

$$\hat{S} = \text{diag}(\langle \hat{S} \rangle_1, \langle \hat{S} \rangle_2, \ldots)$$

(234)

and $\hat{S} = P^{-1} \hat{S} P$, therefore:

$$\hat{S} = k_B \ln \hat{Z} + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i \hat{X}_i$$

(235)

**Theorem 4** (Diagonalizable geometric equation of state). The equation of state of:

$$\hat{S}[\lambda_1, \ldots, \lambda_{2^n}] = k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}] + \sum_{i=1}^{2^n} \lambda_i \hat{E}_i \hat{X}_i[\lambda_1, \ldots, \lambda_{2^n}]$$

(236)

is:

$$d\hat{S}[\lambda_1, \ldots, \lambda_{2^n}] = \sum_{i=1}^{2^n} \lambda_i \hat{E}_i d\hat{X}_i[\lambda_1, \ldots, \lambda_{2^n}]$$

(237)

**Proof.** 1. We now take the total derivative of $\hat{S}[\lambda_1, \ldots, \lambda_{2^n}]$:

$$d\hat{S}[\lambda_1, \ldots, \lambda_{2^n}] = d(k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}]) + \sum_{i=1}^{2^n} d(\lambda_i \hat{E}_i \hat{X}_i[\lambda_1, \ldots, \lambda_{2^n}])$$

(238)

Please be warned that taking the total derivative of $2^n + 1$ functions each having $2^n$ variables is quite verbose.
(a) First, we investigate \( d(k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}]) \).

\[
d(k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}]) = \partial_{\lambda_1} k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}] \, d\lambda_1 + \cdots + \partial_{\lambda_{2^n}} k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}] \, d\lambda_{2^n}
\]  

(239)

\[
= \frac{\partial}{\partial \lambda_1} k_B \ln \left[ \sum_{p \in \mathcal{D}} \exp \left( -\frac{1}{k_B} \sum_{i=1}^{2^n} (\lambda_i \hat{E}_i X_i[p]) \right) \right] \, d\lambda_1 + \cdots
\]  

(240)

\[
= -\frac{1}{\hat{Z}} \sum_{p \in \mathcal{D}} \hat{E}_i X_i[p] \exp \left( -\sum_{i=1}^{2^n} (\lambda_i \hat{E}_i X_i[p]) \right) \, d\lambda_1 + \cdots
\]  

(241)

\[
= -\hat{E}_1 X_1 \, d\lambda_1 + \cdots
\]  

(242)

Consequently,

\[
d(k_B \ln \hat{Z}[\lambda_1, \ldots, \lambda_{2^n}]) = -\hat{E}_1 X_1 \, d\lambda_1 - \cdots - \hat{E}_{2^n} X_{2^n} \, d\lambda_{2^n}
\]  

(243)

(b) Second, we investigate \( d(\lambda_i \hat{E}_i X_i[\lambda_1, \ldots, \lambda_{2^n}]) \).

\[
d(\lambda_i \hat{E}_i X_i[\lambda_1, \ldots, \lambda_{2^n}]) = \frac{\partial}{\partial \lambda_1} (\lambda_i \hat{E}_i X_i[\lambda_1, \ldots, \lambda_{2^n}]) \, d\lambda_1 + \cdots + \frac{\partial}{\partial \lambda_{2^n}} (\lambda_i \hat{E}_i X_i[\lambda_1, \ldots, \lambda_{2^n}]) \, d\lambda_{2^n}
\]  

(245)

\[
= \hat{E}_i \left( (\frac{\partial}{\partial \lambda_1} \lambda_i) X_i[\lambda_1, \ldots, \lambda_{2^n}] + \lambda_i \frac{\partial}{\partial \lambda_1} X_i[\lambda_1, \ldots, \lambda_{2^n}] \right) \, d\lambda_1 + \cdots
\]  

(246)

We note that if \( \lambda_i = \lambda_1 \), then \( \frac{\partial}{\partial \lambda_1} \lambda_i = 1 \), and if \( \lambda_i \neq \lambda_1 \), then \( \frac{\partial}{\partial \lambda_1} \lambda_i = 0 \). This is true for all terms of the series \( \lambda_2, \lambda_3, \ldots \), etc. Consequently:

\[
= \hat{E}_i \left( X_i[\lambda_1, \ldots, \lambda_{2^n}] + \lambda_i \frac{\partial}{\partial \lambda_1} X_i[\lambda_1, \ldots, \lambda_{2^n}] \right) \, d\lambda_1 + \cdots
\]  

(247)

We then add the two expressions together, and do so for all \( i \in \{1, \ldots, 2^n\} \). We note that the first expression is completely cancelled by the first half of the second. Then the remaining half of the second expression is simply the definition of the total derivative for the functions:

\[
\bar{X}_1[\lambda_1, \ldots, \lambda_{2^n}], \ldots, \bar{X}_{2^n}[\lambda_1, \ldots, \lambda_{2^n}]
\]  

(249)
Finally, we obtain the equation of state, proving (Theorem 1):

\[ \text{d} \hat{S}[\lambda_1, \ldots, \lambda_{2^n}] = \sum_{i=1}^{2^n} \lambda_i \text{d} \hat{E}_i [\lambda_1, \ldots, \lambda_{2^n}] \] (250)

\[ \square \]

**Definition 15** (Geometric thermodynamics). In some range, the map

\[ S[\lambda_1, \ldots, \lambda_{2^n}] \rightarrow S[X_1, \ldots, X_{2^n}] \] (251)

is invertible. In this range, we can thus think of \( S \) as a function of \( X_1, \ldots, X_{2^n} \), which we then use to define the relations of thermodynamics. The thermodynamic relations are then given by the following \( 2^n \) partial derivatives:

\[ \frac{\partial S}{\partial X_i} \bigg|_{(X_1, \ldots, X_{2^n}) \setminus X_i} = \lambda_i E_i \] (252)

Here, \( S \) is derived with respect to \( X_i \) while holding the other quantities \( \{X_1, \ldots, X_{2^n}\} \setminus X_i \) constant.

As an example, let us consider the case of an ensemble produced from 4 geometric priors (using the \( CI_{1,3}(\mathbb{R}) \) algebra):

\[ \gamma_0 X_0 = \sum_{p \in D} \gamma_0 X_0[p] g[p] \] (253)

\[ \gamma_1 X_1 = \sum_{p \in D} \gamma_1 X_1[p] g[p] \] (254)

\[ \gamma_2 X_2 = \sum_{p \in D} \gamma_2 X_2[p] g[p] \] (255)

\[ \gamma_3 X_3 = \sum_{p \in D} \gamma_3 X_3[p] g[p] \] (256)

The partition function is:

\[ Z = \Omega^{-1} \sum_{p \in D} \exp \left( -\frac{\lambda_0}{k_B} \gamma_0 X_0[p] - \frac{\lambda_1}{k_B} \gamma_1 X_1[p] - \frac{\lambda_2}{k_B} \gamma_2 X_2[p] - \frac{\lambda_3}{k_B} \gamma_3 X_3[p] \right) \] (257)

The entropy is:

51
\[ S = k_B \ln Z + \lambda_0 \gamma_0 \mathbf{X}_0 + \lambda_1 \gamma_1 \mathbf{X}_1 + \lambda_2 \gamma_2 \mathbf{X}_2 + \lambda_3 \gamma_3 \mathbf{X}_3 \]  

(258)

The geometric equation of state is:

\[ dS = \lambda_0 \gamma_0 d\mathbf{X}_0 + \lambda_1 \gamma_1 d\mathbf{X}_1 + \lambda_2 \gamma_2 d\mathbf{X}_2 + \lambda_3 \gamma_3 d\mathbf{X}_3 \]  

(259)

Its matrix representation is:

\[
\begin{pmatrix}
\lambda_0 d\mathbf{X}_0 & 0 & \lambda_3 d\mathbf{X}_3 & \lambda_1 d\mathbf{X}_1 - i\lambda_2 d\mathbf{X}_2 \\
0 & \lambda_0 d\mathbf{X}_0 & \lambda_1 d\mathbf{X}_1 + i\lambda_2 d\mathbf{X}_2 & -\lambda_3 d\mathbf{X}_3 \\
-\lambda_1 d\mathbf{X}_1 - i\lambda_2 d\mathbf{X}_2 & -\lambda_1 d\mathbf{X}_1 + i\lambda_2 d\mathbf{X}_2 & \lambda_3 d\mathbf{X}_3 & 0 \\
-\lambda_3 d\mathbf{X}_3 & -\lambda_1 d\mathbf{X}_1 - i\lambda_2 d\mathbf{X}_2 & 0 & \lambda_0 d\mathbf{X}_0 \\
\end{pmatrix}
\]  

(260)

Finally, the diagonal is:

\[
d\mathbf{S} = \sqrt{(\lambda_0 d\mathbf{X}_0)^2 - (\lambda_1 d\mathbf{X}_1)^2 - (\lambda_2 d\mathbf{X}_2)^2 - (\lambda_3 d\mathbf{X}_3)^2}
\]

(261)

**Definition 16** (Metric-form). In this case, we can ’abuse’ the notation \( \pm \) to group the unique eigenvalues of \( d\mathbf{S} \) as follows:

\[
dS = \pm \sqrt{(\lambda_0 d\mathbf{X}_0)^2 - (\lambda_1 d\mathbf{X}_1)^2 - (\lambda_2 d\mathbf{X}_2)^2 - (\lambda_3 d\mathbf{X}_3)^2}
\]  

(262)

We call this expression the ‘metric-form’ of the poly-vector.

Let us now attribute a physical quantity to the Lagrange multipliers. We insert the quantities \( \tilde{k} \) and \( f \) from Table 4 into the geometric interval by setting \( \lambda_0 = f k_B \) and \( \lambda_1 = \lambda_2 = \lambda_3 = \tilde{k} k_B \). We obtain:

\[
(dS)^2 = (f k_B d\mathbf{X}_0)^2 - (\tilde{k} k_B d\mathbf{X}_1)^2 - (\tilde{k} k_B d\mathbf{X}_2)^2 - (\tilde{k} k_B d\mathbf{X}_3)^2
\]  

(263)

Then, by dividing each side of the equation by \( \tilde{k}^2 k_B^2 \), we obtain:

\[
\tilde{k}^{-2} k_B^{-2} (dS)^2 = (c d\mathbf{X}_0)^2 - (d\mathbf{X}_1)^2 - (d\mathbf{X}_2)^2 - (d\mathbf{X}_3)^2
\]  

(264)

which we identify as the interval of special relativity.

We note that instead of the orthogonal generators \( \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\} \) we could have used arbitrary generators \( \{e_0, e_1, e_2, e_3\} \) and by repeating the same steps, we would have obtained:

\[
\tilde{k}^{-2} k_B^{-2} (dS)^2 = g_{\mu\nu} d\mathbf{X}^\mu d\mathbf{X}^\nu
\]  

(265)
which we identify as the interval of general relativity.

Finally, we note the following new equivalence between the equation of state and the interval $s$:

$$(ds)^2 \equiv \tilde{k}^{-2}k_B^{-2}(dS)^2 \quad (266)$$

Consequently, the equation of state quantifies the interval between events in space-time using entropy.

5 Results (Cosmology)

5.1 Law of inertia

Consider a geometric ensemble constrained by $\sigma_1 X_1, \sigma_2 X_2, \sigma_3 X_3$. Its equation of state is:

$$dS = k_B(\tilde{k}_1 \sigma_1 dX_1 + \tilde{k}_2 \sigma_2 dX_2 + \tilde{k}_3 \sigma_3 dX_3) \quad (267)$$

Its matrix representation is:

$$d\hat{S} = k_B \left( \begin{array}{ccc} \tilde{k}_3 dX_3 & \tilde{k}_1 dX_1 - i\tilde{k}_2 dX_2 \\ \tilde{k}_1 dX_1 + i\tilde{k}_2 dX_2 & -\tilde{k}_3 dX_3 \end{array} \right) \quad (268)$$

The diagonal matrix representation is:

$$d\hat{S} = k_B \sqrt{\tilde{k}_1^2(dX_1)^2 + \tilde{k}_2^2(dX_2)^2 + \tilde{k}_3^2(dX_3)^2} \left( \begin{array}{c} 1 \\ 0 \\ -1 \end{array} \right) \quad (269)$$

Using the $\pm$ notation, we group the eigenvalues as a 0-vector expression:

$$dS = \pm k_B \sqrt{\tilde{k}_1^2(dX_1)^2 + \tilde{k}_2^2(dX_2)^2 + \tilde{k}_3^2(dX_3)^2} \quad (270)$$

Taking $\tilde{k} = \tilde{k}_1 = \tilde{k}_2 = \tilde{k}_3$, we get

$$dS = \pm \tilde{k} k_B \sqrt{(dX_1)^2 + (dX_2)^2 + (dX_3)^2} \quad (271)$$

We now wish to investigate the entropic force emergent from the equation of state. We recall the definition of an entropic force:

$$T dS = F dx \quad (272)$$
For instance, this force could represent the tension felt by a polymer in a warm bath. We are not very far from this definition. The only missing part is to multiply each side by a proportionality constant $T$ with units Kelvin:

$$T \, dS = \pm \hat{T}k_B \sqrt{(d\mathbf{X}_1)^2 + (d\mathbf{X}_2)^2 + (d\mathbf{X}_3)^2}$$  \hspace{1cm} (273)$$

in which case $\hat{T}k_B$ acquires the units of a force $F$.

Which value of $\hat{k}$ and $T$ to use? The natural choices, proposed by Erik Verlinde\cite{59} are of course to take $T$ as the Unruh temperature and $F$ as the law of inertia. Let us solve for $\hat{k}$ using:

$$T_{\text{Unruh}} = \frac{\hbar a}{2\pi c k_B} \hspace{1cm} (274)$$

$$F = ma \hspace{1cm} (275)$$

Then:

$$F = \hat{T}k_B \hspace{1cm} (276)$$

$$ma = \frac{\hbar a}{2\pi c k_B} \hat{k} \hspace{1cm} (277)$$

$$= \frac{\hbar a}{2\pi c} \hat{k} \hspace{1cm} (278)$$

$$\Rightarrow \hat{k} = ma \frac{2\pi c}{\hbar a} \hspace{1cm} (279)$$

$$= 2\pi \frac{mc}{\hbar} \hspace{1cm} (280)$$

We recognize the term $mc/\hbar := \lambda^{-1}$ as the inverse of the reduced Compton wavelength. Here, the Compton wavelength is revealed as a proportionality constant between distance and entropy. Intuitively, an object with a larger Compton wavelength requires less information to specify its position than an object with a small Compton wavelength. Finally, we define:

**Definition 17** (Geometric entropy of inertia).

$$S = 2\pi k_B \frac{mc}{\hbar} \left( \sigma_1 \mathbf{X}_1 + \sigma_2 \mathbf{X}_2 + \sigma_3 \mathbf{X}_3 \right)$$ \hspace{1cm} (281)$$

**Definition 18** (Geometric equation of state of inertia).

$$dS = 2\pi k_B \frac{mc}{\hbar} \left( \sigma_1 d\mathbf{X}_1 + \sigma_2 d\mathbf{X}_2 + \sigma_3 d\mathbf{X}_3 \right)$$ \hspace{1cm} (282)$$

**Definition 19** (Metric equation of state of inertia).

$$dS = \pm 2\pi k_B \frac{mc}{\hbar} \sqrt{(d\mathbf{X}_1)^2 + (d\mathbf{X}_2)^2 + (d\mathbf{X}_3)^2}$$ \hspace{1cm} (283)$$
5.2 Beckenstein-Hawking entropy

Starting from the geometric entropy of inertia, we now consider the case where the Compton wavelength also varies (i.e. the mass varies). Specifically, one considers $S$ to be a function of $(m, X_0, X_1, X_2)$:

$$S[m, X_0, X_1, X_2] = 2\pi k_B \frac{mc}{\hbar} \left( \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 \right) + \ln Z \quad (284)$$

Then the total derivative of $S$ is:

$$dS = 2\pi k_B \frac{c}{\hbar} \left( m (\sigma_1 dX_1 + \sigma_2 dX_2 + \sigma_3 dX_3) + (\sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3) dm \right) \quad (285)$$

Re-arranging, we get:

$$dS = 2\pi k_B \frac{c}{\hbar} \left( \sigma_1 (m dX_1 + X_1 dm) + \sigma_2 (m dX_2 + X_2 dm) + \sigma_3 (m dX_3 + X_3 dm) \right) \quad (286)$$

The metric equation of state is:

$$dS = \pm 2\pi k_B \frac{c}{\hbar} \sqrt{(m dX_1 + X_1 dm)^2 + (m dX_2 + X_2 dm)^2 + (m dX_3 + X_3 dm)^2} \quad (287)$$

We now assign the Schwarzschild radius to the metric: $X_1 = 2Gm/c^2, X_2 = 0, X_3 = 0$, and $dX_1 = 2Gc^{-2} dm, dX_2 = 0, dX_3 = 0$:

$$dS = \pm 2\pi k_B \frac{c}{\hbar} \frac{4Gm}{c^2} dm \quad (288)$$

$$= \pm 2\pi k_B \frac{c}{\hbar} \frac{4Gm}{c^2} dm \quad (289)$$

$$= \pm k_B 8\pi \frac{G}{\hbar c} m dm \quad (290)$$

We now integrate:

$$\int dS = \pm k_B 8\pi \frac{G}{\hbar c} \int m dm \quad (291)$$

$$S = \pm k_B 4\pi \frac{G}{\hbar c} m^2 + C \quad (292)$$

For a black hole, the mass relates to the area as:
\[ A = 4\pi r^2 = 4\pi \left( \frac{2Gm}{c^2} \right)^2 = 4\pi \frac{4G^2m^2}{c^4} \implies m^2 = A \frac{c^4}{16\pi G^2} \] (293)

Replacing \( m^2 \) in our integral result \( S \), we get:

\[ S = \pm k_B 4\pi \frac{G}{\hbar c} A \frac{c^4}{G^2 16\pi} + C \] (294)
\[ = \pm k_B c^3 A \frac{c^4}{hG 16} + C \] (295)

Finally, taking the sign to be positive and \( C = 0 \), we get:

\[ S = k_B c^3 A \frac{c^4}{hG 16} \] (296)

which is the Bekenstein-Hawking entropy.

### 5.3 de Sitter space

We recall that de Sitter space is a hyperboloid defined in 3+1 Minkowski spacetime by the relation:

\[ \alpha^2 = -c^2 X_0^2 + X_1^2 + X_2^2 + X_3^2 \] (297)

with a cosmological horizon at \( r = \alpha \). One may also write \( \alpha \) as:

\[ \alpha = \pm \sqrt{-(cX_0)^2 + X_1^2 + X_2^2 + X_3^2} \] (298)

Let us now compare it to the 1-vector geometric entropy:

\[ S = k_B \ln Z + \lambda_0 \gamma_0 \overline{X}_0 + \lambda_1 \gamma_1 \overline{X}_1 + \lambda_2 \gamma_2 \overline{X}_2 + \lambda_3 \gamma_3 \overline{X}_3 \] (299)

To convert it to its metric form, we first move \( \ln Z \) to the left:

\[ S - k_B \ln Z = \lambda_0 \gamma_0 \overline{X}_0 + \lambda_1 \gamma_1 \overline{X}_1 + \lambda_2 \gamma_2 \overline{X}_2 + \lambda_3 \gamma_3 \overline{X}_3 \] (300)

Then, we diagonalize the matrix representation:

\[
P(\hat{S} - k_B \ln \hat{Z})P^{-1} = \sqrt{-(\lambda_0 \overline{X}_0)^2 + (\lambda_1 \overline{X}_1)^2 + (\lambda_2 \overline{X}_2)^2 + (\lambda_3 \overline{X}_3)^2}
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\] (301)
Using the ± notation we group the unique eigenvalues into the metric form:

\[ S - C = \pm \sqrt{-(\lambda_0 \overline{X}_0)^2 + (\lambda_1 \overline{X}_1)^2 + (\lambda_2 \overline{X}_2)^2 + (\lambda_3 \overline{X}_3)^2} \]  

(302)

Now, we inject the coefficient \(2\pi k_B mc/\hbar\) previously obtained for the equation of state of inertia as the Lagrange multiplier, we get:

\[ S - C = \pm \frac{2\pi k_B mc}{\hbar} \sqrt{-(c \overline{X}_0)^2 + \overline{X}_1^2 + \overline{X}_2^2 + \overline{X}_3^2} \]  

(303)

Squaring it, we get:

\[ (S - C)^2 = \left(\frac{2\pi k_B mc}{\hbar}\right)^2 \left(-c^2 \overline{X}_0^2 + \overline{X}_1^2 + \overline{X}_2^2 + \overline{X}_3^2\right) \]  

(304)

We recover the same mathematical form as the hyperboloid equation in 3+1 space-time characteristic of de Sitter space, except of course that now the quantity \(\alpha\) is associated with the entropy. We will now study this equation in more detail. The entropy is therefore:

\[ S = \pm 2\pi k_B \frac{mc}{\hbar} \alpha + C \]  

(305)

In de Sitter space, \(\alpha\) is the Hubble radius at \(c/H\). Using the Hubble radius, we will derive the cosmological pressure and the cosmological law of inertia, then we will do it again but with added arbitrary thermodynamic quantities. We will call the second case: thermodynamically-deformed de-Sitter space, and we will show that it is equivalent to \(\Lambda CDM\).

### 5.4 Cosmological pressure

We consider that the cosmological horizon bears an entropy for the same reason that the black hole apparent horizon bears an entropy. As we are in de Sitter space, therefore we consider that the cosmological horizon and the Hubble radius are the same. Specifically, we describe the cosmological horizon using the Hubble radius \(r = c/H\) where \(H\) is the Hubble constant and we replace \(a\) by \(cH\) in the Unruh temperature\[30\] \[17\]. With these replacements, the Unruh temperature becomes the cosmological horizon temperature\[30\]:

\[ T_{\text{de-Sitter-horizon}} := \frac{\hbar H}{2\pi k_B} \]  

(306)

Starting with the Bekenstein-Hawking entropy with the minus sign as a starting point (we are inside the cosmological horizon thus we flip the sign), we
then multiply each side by $T_{\text{de-Sitter-horizon}}$ as a proportionality constant and we get:

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG^4} \, dA$$  \hspace{1cm} (307)

Let us now write this equation in terms of volume by replacing $A$ with $V$. With this replacement, the equation will be formally the same as before, but the coefficient now has the units of pressure. Using $A = 4\pi r^2$ and $V = 4/3\pi r^3$ therefore $dA = 2r^{-1} \, dV$, we get:

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG^4} 2r^{-1} \, dV$$  \hspace{1cm} (308)

Simplifying (and using the radius replacement $r \to c/H$), we get:

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG^4} \frac{H}{c} \, dV$$  \hspace{1cm} (309)

$$= -\frac{H^2}{4\pi G} c^2 \, dV$$  \hspace{1cm} (310)

Finally, we rewrite the expression in terms of the critical cosmological density $\rho = 3H^2/(8\pi G)$, and we obtain a negative entropic pressure corresponding to 66% of the total energy of the universe\[30\]:

$$T \, dS = -\frac{2}{3}\rho c^2 \, dV$$  \hspace{1cm} (311)

This result was obtained by Easson in \[30\], where he suggested that the accelerated expansion of the universe could be explained by this entropic negative pressure.

### 5.5 Cosmological inertia

We repeat the same process as was used to derive the cosmological pressure, but instead of rewriting the relation from area to volume, we go from $dA$ to radius $dr$, using: $dA = d(4\pi r^2) = 8\pi r \, dr$. Using this replacement as well as the cosmological horizon temperature and the Bekenstein-Hawking entropy, we get:

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG^4} (8\pi r \, dr)$$  \hspace{1cm} (312)
Then, with $r \to c/H$, we get:

$$T \ dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{\hbar G^4} 8\pi \frac{c}{H} \ dr \ (313)$$

$$= -\frac{c^4}{G} \ dr \ (314)$$

Since the surface gravity of a horizon is equal to $a = c^4/(4G)$, we can rewrite this expression in terms of acceleration, and we get:

$$T \ dS = -4M \ dr \ (315)$$

Using these results, we assign to the energy of the de Sitter universe a 66% negative pressure component (obtained in the previous section) and 25% inertial matter component (the energy content of the cosmos is weighted at one fourth its inertial mass).

5.6 Entropic derivation of $\Lambda$CDM

In the previous case, we have considered that the cosmological horizon is at the Hubble radius (de Sitter space). However, according to present observations, this is not quite the case. The cosmological horizon is slightly beyond the Hubble horizon (at $\approx 5$ giga-parsec, versus 4.1 giga-parsec).

To account for this difference, we interpret the geometric entropy as describing de-Sitter space with a deformation on the position of the Hubble horizon with respect to the cosmological horizon. For generality, we can in fact include any number of scalar thermodynamic quantities $\{\mu_1N_1, \ldots, \mu_nN_n\}$. The equation becomes:

$$S = \pm (2\pi k_B mc/h)\alpha + C - \mu_1\overline{N}_1 - \cdots - \mu_n\overline{N}_n \ (316)$$

This entropy is determined by two competing processes. First, we recall that the entropy is information inaccessible to the observer. Working in the direction contributing to the entropy, $C$ is usually interpreted to be the intrinsic degeneracy of the system (at $T \to 0$, it is, in fact, the degeneracy of the ground state), whereas, $\alpha$ represents, as always, the distance to the Hubble horizon. The contribution from these terms work in the opposite direction to that of the scalar terms $\{\mu_1\overline{N}_1, \ldots, \mu_n\overline{N}_n\}$. Finally, we recall that the Hubble horizon meets the cosmological horizon asymptotically at $t \to \infty$ when all matter has exited the cosmological horizon. Thus, for the universe to be asymptotically de Sitter at $t \to \infty$, it follows that $\lim_{t\to\infty} \alpha \to (2\pi k_B mc/h)^{-1}S$ and for all $i$, $\lim_{t\to\infty} \overline{N}_i \to 0$. 

59
By attributing the role of bookkeeper (of the matter and energy yet to leave the horizon) to \( \{ \mu_1 N_1, \ldots, \mu_n N_n \} \), and by adopting the law of conservation of energy, then the sum-total of all matter and energy leaving the horizon can be summarized as the continuity equation:

\[
dE = (\rho c^2 + p) dV \tag{317}
\]

To equate this relation to the entropy, we must now introduce the temperature at the real cosmological horizon, using the Unruh temperature as the starting point with \( a \rightarrow c^2/r \). We use the relations derived by Easson\[30\], first for the radius:

\[
 r_{\text{cosmological-horizon}} := \frac{c}{\sqrt{H^2 + k/a^2}} \tag{318}
\]

where \( a \) is a scaling factor. Then for the temperature:

\[
 T_{\text{cosmological-horizon}} := \frac{\hbar (c^2/r_{\text{cosmological-horizon}})}{2\pi c k_B} \tag{319}
\]

As before, here we use the temperature as a proportionality constant, and we rewrite the entropy as follows:

\[
 T_{\text{cosmological-horizon}} dS = (\rho c^2 + p) dV \tag{320}
\]

or, more specifically, as:

\[
 T_{\text{cosmological-horizon}} \frac{k_B c^3}{4\hbar G} dA = (\rho c^2 + p) dV \tag{321}
\]

This is the sufficient starting point used by Easson\[30\] (in Annex A of his paper) to recover the Friedman equations of cosmology:

\[
 \dot{H} - \frac{k}{a^2} = -4\pi G \left( \rho + \frac{p}{c^2} \right) \tag{322}
\]

\[
 H^2 = \frac{8\pi G}{3} \rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{323}
\]

as an equivalent representation of (321).

We have come full circle; the Seth Lloyd relations regarding the conservation of bits and operations in the universe, which originally motivated this non-commutative generalization of statistical physics, as allowed us to recover cosmology strictly using the facilities of (geometric) statistical physics.

Entropically emergent cosmology is to geometric statistical physics what the ideal gas law is to statistical physics.
5.7 Arrow of time

We want to find the direction of the maximum rate of change in entropy at point \((0,0)\) in flat Minkowski space-time. Imagine a photon traveling on the absolute edge of the light cone. The change of entropy in this direction is null because the interval along this path is zero. In contrast, straight up (towards the future) the change in entropy is maximal as the absolute value of the interval is not reduced by a change in the \(x,y,z\) coordinates. Finally, towards the past, the gradient of entropy is minimal (negative extremum).

We recall the notion of an entropic force, such as a polymer in a warm bath. The general statistical tendency of a system to fluctuate towards the configuration of higher entropy causes, in these systems, the emergence of an entropic force pointing in the direction of increased entropy.

Here, similar behavior is obtained but instead of just in space and with a force, it is also in time and with a power. Indeed, consider the 1-vector geometric entropy:

\[
dS = \frac{f k_B}{P/T} \gamma_0 dX_0 + \frac{\tilde{k} k_B}{F/T} (\gamma_1 dX_1 + \gamma_2 dX_2 + \gamma_3 dX_3)
\]

(324)

The terms \(f k_B = P/T\) and \(\tilde{k} k_B = F/T\) are simply, at thermodynamic equilibrium, an entropic power, and an entropic force. \(P\) is produced by the gradient of entropy in time, and \(F\) is produced by the gradient of entropy in space. Consequently, the geometric ensemble of space-time always has an arrow of time, powered by entropy, which points towards the maximum of the entropy gradient; that is, in flat Minkowski space, towards the direct future of the observer. Observers, therefore, advances into their future because of the gradient of entropy.

In the case of generally curved space-times, the direction of motion in space-time is the geodesic. In this case, the observer experiences entropic forces along their paths in space-time which "tilt" the direction of maximal entropy.

The use of the more familiar terms power and force is entirely optional. The conclusion is the same had we omitted multiplication by \(T\). In which case, the conjugates emergent from the gradient of entropy would be an entropic frequency \(f\) and an entropic repetency \(\tilde{k}\). Nonetheless, the gradient of entropy would point towards the future of the observer.

Let us parametrize the metric equation of state over a path \(\tau \in [a,b]\) and then integrate:

\[
\int_a^b \frac{\partial S}{\partial \tau} d\tau = \pm 2\pi k_B \frac{mc}{\hbar} \int_a^b \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} d\tau
\]

(325)

We re-arrange as follows:
To recover the dynamics, one merely needs to investigate the change of entropy under an infinitesimal variation of $\delta$.

\[
\frac{\hbar}{2\pi k_B} \int_a^b \frac{\partial S}{\partial \tau} \, d\tau = \pm mc \int_a^b \sqrt{g_{\mu\nu}} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \, d\tau
\]

The path along $\tau$ that extremalize the production of entropy is given in the stationary regime by posing:

\[
\delta \frac{\partial S}{\partial \tau} = 0
\]

and the corresponding equations of motion are:

\[
\delta \sqrt{g_{\mu\nu}} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} = 0
\]

This equation is the Euler-Lagrange equation of motion for a test particle in curved space-time. Expanding it yields the equations of geodesic motion. Consequently, geodesic motion is revealed as the path for which the production of entropy is extremal in space-time.

We have now identified a relation between the action and the entropy. We will use $S$ to denote the action (as we already use $S$ for the entropy). Then, the action relates to the metric equation of state as follows:

\[
S := \pm \frac{\hbar}{2\pi k_B} \int_a^b \frac{\partial S}{\partial \tau} \, d\tau
\]

6 Results (Quantum mechanics)

6.1 Fermi-Dirac statistics of events

We consider that an event can occur at most once (whatever happens to Schrödinger’s cat, for sure, it doesn’t die twice), and thus we will use Fermi-Dirac statistics to study the occupancy distribution of events in space-time.

The 1-vector geometric equation of state of $Cl_{1,3}(\mathbb{R})$ has two unique eigenvalues:

\[
dS = \sqrt{(\lambda_0 dX_0)^2 - (\lambda_1 dX_1)^2 - (\lambda_2 dX_2)^2 - (\lambda_3 dX_3)^2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}
\]
For simplicity, we will consider the 1+1 space-time case and we use the ± notation to group the eigenvalues as a single expression:

$$dS = \pm \sqrt{(\lambda_0 \ d\bar{X}_0)^2 - (\lambda_1 \ d\bar{X}_1)^2}$$

We will now attribute each eigenvalue to a different direction of time. The positive eigenvalue points towards the future, and the negative eigenvalue towards the past.

The Fermi-Dirac distributions for this equation of state are:

$$\langle n \rangle_{\text{future}} = \frac{1}{\exp \sqrt{(\lambda_0 \ d\bar{X}_0)^2 - (\lambda_1 \ d\bar{X}_1)^2} + 1}$$

$$\langle n \rangle_{\text{past}} = \frac{1}{\exp -\sqrt{(\lambda_0 \ d\bar{X}_0)^2 - (\lambda_1 \ d\bar{X}_1)^2} + 1}$$

To attribute the correct eigenvalue to the direction of time over the whole interval, we combine $$\langle n \rangle_{\text{future}}$$ and $$\langle n \rangle_{\text{past}}$$ as a piecewise function:

$$\langle n \rangle = \begin{cases} 
\frac{1}{\exp -\sqrt{(\lambda_0 \ d\bar{X}_0)^2 - (\lambda_1 \ d\bar{X}_1)^2} + 1} & X_0 < 0 \\
\frac{1}{2} & X_0 = 0 \\
\frac{1}{\exp \sqrt{(\lambda_0 \ d\bar{X}_0)^2 - (\lambda_1 \ d\bar{X}_1)^2} + 1} & X_0 > 0
\end{cases}$$

$$\langle n \rangle$$ is shown in Figure (1). As we can see, $$\langle n \rangle$$ has the shape of a light cone in Minkowski space with the observer at the origin (0, 0). Remarkably it achieves the correct shape only by using event occupancy information. With this information, it improves upon the usual description of a light cone, by attributing different properties to the past as it does to the future, in terms of event occupancy statistics. For the future, the occupancy rate of events is depleted at 0% (future events are upcoming and have not occurred). For the past, the occupancy rate of events is saturated at 100% (past events have occurred and will not re-occur). An observer O at point (0,0) evolving towards its future will experience a transfer in the depleted occupancy of future events to a saturation in the occupancy of past events (Figure [1] and [2]). To better illustrate, we introduce the analogy of a tide flooding the past with events as the present advances in space-time. Along with O, the tide advances in space-time at the speed of light towards the direction of the future (Figure [3]). Three distinct regimes of time are described; the past (100% event occupancy), the present (at the inflection point in the occupancy of events) and the future (0% event occupancy).
Figure 1: Graph of the Fermi-Dirac statistics over the occupancy of space-time events. Red means an occupancy rate of 100%, whereas blue means 0% (and with rainbow colors for intermediate values). The black line at $X_0 = 0$ is the hypersurface of the present. The observer is always at point (0, 0). a) The shape of the plot is that of a light cone of special relativity. From the perspective of the present, the occupancy rate in the past is 100% and that of the future is 0%. The white region is indeterminate (or more precisely, complex-valued). b) The image in the middle is a perspective view of the image on the left. c) The image on the right is a cut-out of the white dotted line of the sub-figure a). The shape of the curve is reminiscent of the usual shape of the Fermi-Dirac distribution over energy levels.
6.2 Quantum mechanics and measurements

The probability of occupancy of an event is obtained by applying the Fermi-Dirac statistics to the ensemble. By doing so, we recover both the usual classical probability rules (the probabilities are real-valued functions between zero and one, and they sum to one) and the quantum probability rules (the probabilities are complex amplitudes multiplied by their conjugate). Both 'kinds' of probability play a role in the same ensemble and, notably, are dependent upon which region in space-time (with respect to the observer) the dynamics are described in.

First, we note that, in axiomatic science, movement is not fundamental. Indeed, Axiom 1 together with Assumption 1 assigns manifests to all state of affairs of the world, but it remains silent with respect to any notion of dynamics connecting manifests together in time. In fact, the foundation of axiomatic science does not even mention time itself. Time, along with space-time, is, in the macroscopic/bulk state, considered to be emergent from geometric entropy. To maintain consistency with this setup, we will here introduce movement as an 'interpolation' between a discrete sequence of manifests. Let us explain:

Perhaps the best way to understand these restrictions concretely is to read John A. Wheeler’s participatory universe hypothesis, laid out in Complexity, entropy, and the physics of information\cite{60}. Here, we will briefly state the relevant concepts. In his article, Wheeler considers that the information one obtains about nature is exclusively in the form of detector 'clicks'. For instance, in a photon counting experiment, the detector either 'clicks' or it doesn’t. And in a different setup, the 'clicks' may occur under different circumstances but nonetheless, the basic element contributing to our knowledge of reality remains the 'click'. Now, attributing an ontological existence to a photon in-between the various 'clicks' that are being registered is a "blown-up version" of the simple raw fact that a click was registered. The brute facts are the registrations of clicks on detectors, and the theory 'behind the clicks'; in this case that there exists a photon connecting the clicks is a mere derived hypothesis consistent with the clicks.

Axiomatic science essentially agrees with this interpretation. As the observer evolves forward in time, events are registered as 'clicks' (the occupancy rate goes from 0% to 100%), and their contribution is added to the manifest. Let us now recall that the metric form of a geometric ensemble relates the entropy to the distance, in space-time, between now and some future or past state of affairs. Then, because of the fundamental assumption of science (Assumption 1), we can attribute a manifest to all future or past state of affairs and note their entropic departure from the reference manifest. From this, we interpret the 'illusion' of the flow of time from an entropy basis, as starting with the reference manifest, followed by a sequence of other manifests each one more entropically distant from the present than the previous one. Thus, in axiomatic science, there is no movement, only a sequence of manifests.

As we will see, in this context and with these restrictions, we can still introduce a notion dynamics which corresponds to our understanding of movement such that it is completely compatible with axiomatic science. Consequently, movement
will be introduced by some 'interpolation' between manifests which, ultimately, may leave one 'guessing' as to whether real movement occurs or if it's just an intellectually pleasing gimmick. Using a more robust terminology; the entropy regarding which path was chosen (amongst all possible paths) must be maximized such that the system retains no information as to which path (if any at all) was taken. As we will see, this is precisely the conditions required for an equivalent derivation of quantum mechanics. Indeed, geometric statistical physics assigns the proper statistical weight to each event, as required for a quantum mechanical description of movement using the Feynman path integral.

As mentioned in the description of Figure 1a), the white region outside the light cone corresponds to a complex-valued occupancy rate for events. We can see this as a consequence that the metric contains a square root and thus calculating the interval from the observer to a space-like separated event yields an imaginary value. For instance the space-time interval between \((0,0)\) and, say, \((0,5)\) will be imaginary:

\[
\sqrt{c^2(\Delta t)^2 - (\Delta x)^2} \bigg|_{\Delta t=0, \Delta x=5} = \sqrt{c^2(0)^2 - 5^2} = i 5
\]  

As we will see, the consequences of such are remarkable.

To introduce the dynamics, the idea is to take two geometric events \(p_1\) and \(p_2\) then impose as a restriction that the possible paths between these events are those which obey the entropic action (introduced in Section 5.7). Then, using \(\tau\), we parametrize the interval along a path \(q\) between the events, and we sum over the uncountable set of all paths \(\mathbb{P}\) between them. The partition function previously a "sum over computable points" becomes its higher-dimensional cousin: a "functional integral over paths":

\[
Z[\mathbb{P}] = \int_{q \in \mathbb{P}} \exp \left( \pm \int_q \sqrt{\left( \lambda_0 \frac{\partial X_0}{\partial \tau} \right)^2 - \left( \lambda_1 \frac{\partial X_1}{\partial \tau} \right)^2} \, d\tau \right) \, Dq
\]

or, more generally:

\[
Z[\mathbb{P}] = \int_{q \in \mathbb{P}} \exp \left( \pm \int_q \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} \, d\tau \right) \, Dq
\]  

66
Let us analyse this partition function. For a given path \( q \), parts of \( q \) that are space-like will acquire the imaginary term \( i \) in the action whereas the parts that are time-like will not. We may thus split the action into two parts; the real part of the action as the time-like part and the imaginary part as the space-like part.

\[
\int_{q \in P} \exp \left( \pm \text{Re} \left[ \int_q \sqrt{g_{\mu\nu}} \frac{\partial \lambda^\mu X^\mu}{\partial \tau} \frac{\partial \lambda^\nu X^\nu}{\partial \tau} \, d\tau \right] \right) \pm i \text{Im} \left[ \int_q \sqrt{g_{\mu\nu}} \frac{\partial \lambda^\mu X^\mu}{\partial \tau} \frac{\partial \lambda^\nu X^\nu}{\partial \tau} \, d\tau \right] Dq \quad (339)
\]

When \( \mathcal{O} \) describes purely space-like paths, the real part is eliminated and we recover a formulation very close to the Feynman path integral, including the presence of the imaginary term \( i \) multiplying the action:

\[
Z[P] = \int_{q \in P} \exp \left( \pm i \text{Im} \left[ \int_q \sqrt{g_{\mu\nu}} \frac{\partial \lambda^\mu X^\mu}{\partial \tau} \frac{\partial \lambda^\nu X^\nu}{\partial \tau} \, d\tau \right] \right) Dq 
\quad (340)
\]

6.3 Decoherence at the time-like/space-like boundary

Let us investigate the role of each part of the path and explain why we think that having both a real and an imaginary part leads to a more complete description of the system. Paths may, in the general case, have both a time-like (real) action and a space-like (imaginary) action. As we will see, the part of the path that is space-like gives the normal Feynman path integral, and the part of the path that is time-like gives a decoherent version of the path integral.

In the space-like separated region, the system experiences complex interference and it is described by the usual Feynman path integral and with complex amplitudes. However, as the observer advances in time and the paths gradually penetrate the light cone, the probability distribution with complex interference terms abruptly switches to a distribution using only real-valued probabilities. This process occurs continually as the observer advances in time and larger and larger parts of the space-like separated region are integrated within the time-like region of the light cone.

To see the process in the details, it suffices to replace the Lagrange multipliers \( \lambda_i \) by the previously obtained coefficient \( 2\pi k_B mc/\hbar \). Let \( A_E[q] \) be the time-like part of the functional integral and let \( A_S[q] \) be its space-like part:

\[
Z[P] = \int_{q \in P} \exp(A_S[q]) \exp(A_E[q]) Dq \quad (341)
\]

With the coefficient, the space-like part becomes:
\[ A_S[q] = \frac{i}{\hbar} 2\pi \text{Im} \left[ -mc \int q \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} \, d\tau \right] \]  

(342)

The factor \(2\pi\) is attributed to the connection between action and entropy, but otherwise has no impact on the equations of motion. Then for the time-like part of the path, we first multiply the coefficient with \(a/a = 1\) (and with \(a \neq 0\)), then we get:

\[ A_E[q] = -\frac{2\pi c}{\hbar a} \text{Re} \left[ ma \int q \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} \, d\tau \right] \]  

(343)

where

\[ \frac{1}{k_B \beta} = \frac{\hbar a}{2\pi c k_B} = T_{\text{Unruh}} \]  

(344)

A similar process would occur at any horizons, including those of black holes. From O’s point of view the paths inside the horizon are described by complex amplitudes and weighted by the Feynman path integral until they cross the horizon, at which point the system is described by a decoherent sum (in this case thermal) over its energy levels. Remarkably this temperature is the Hawking temperature. Specifically, if we replace \(a \rightarrow c^4/(4MG)\) we get:

\[ \frac{1}{k_B \beta} = \frac{\hbar c^3}{8\pi k_B MG} = T_{\text{Hawking}} \]  

(345)

Thus, a thermally prepared quantum system will enter the light cone of an observer with an energy spectrum at the Unruh temperature (horizon resulting from uniform acceleration) or at the Hawking temperature (horizon resulting from gravity), or even at the cosmological horizon temperature (horizon resulting from the metric expansion of space). In the case of a horizon, as no information can leave it, the time-like radiation of the space-like quantum system is at thermodynamic equilibrium. However, this need not be the case for an unaccelerated observer advancing into the future and capturing a larger sector of the space-like region within its light cone, over time. In flat space-time, the
space-like region is not hidden by a horizon, thus information from the region can eventually enter the light cone of the observer. Consequently, a quantum system with observable $A$ prepared according to the Gibbs measure and with Lagrange multiplier $\alpha$ can enter the light cone of $\mathcal{O}$ with information. Such a system may not be thermal (in the sense that $\beta$ is not a Lagrange multiplier), but it will nonetheless go from a quantum description (in the space-like region) to a decoherent description (in the time-like region) as it enters the light cone. Let us see into more details how the (real) path integral becomes decoherent as it enters the light cone.

We recall the definition of an average observable $\langle O \rangle$, using the path integral formulation:

$$\langle O \rangle = \frac{\int_{q \in \mathcal{P}} O[q] \exp\left(\frac{i}{\hbar}S[q]\right) Dq}{\int_{q \in \mathcal{P}} \exp\left(\frac{i}{\hbar}S[q]\right) Dq}$$ (346)

In our case, we write:

$$\langle O \rangle = \frac{1}{Z} \int_{q \in \mathcal{P}} O[p] \exp\left(-\beta \text{Re}[E[q]] + \frac{i}{\hbar} \text{Im}[-\hbar \beta E[q]]\right) Dq$$ (347)

where $S[q] := \frac{1}{2\pi} \text{Im}[-\hbar \beta E[q]]$ and with:

$$Z := \int_{q \in \mathcal{P}} \exp\left(-\beta \text{Re}[E[q]] + \frac{i}{\hbar} \text{Im}[-\hbar \beta E[q]]\right) Dq$$ (348)

In the case of an information-bearing system (not thermally prepared), we write:

$$\langle O \rangle = \frac{1}{Z} \int_{q \in \mathcal{P}} O[q] \exp\left(-\alpha \text{Re}[A[q]] + i \text{Im}[-\alpha A[q]]\right) Dq$$ (349)

were $\alpha A[q]$ is an arbitrary geometric thermodynamic quantity (or any number thereof) and where $Z := \int_{q \in \mathcal{P}} \exp\left(-\alpha \text{Re}[A[q]] + i \text{Im}[-\alpha A[q]]\right) Dq$. In this non-thermal preparation, measuring the observables over multiple copies of the system gives insight (information) into the preparation of the system (alternatively, we can think of the non-thermal preparation as the free energy being above zero, thus the system is capable of work).

We note:

1. If the system is purely space-like, we obtain the regular path integral:

$$\text{Re}[E[q]] = 0 \implies \langle O \rangle = \frac{1}{Z} \int_{q \in \mathcal{P}} O[p] \exp\left(\frac{i}{\hbar} \text{Im}[-\hbar \beta E[q]]\right) Dq$$ (350)

and $\langle O \rangle$ is the quantum average of the observable.
2. If the system is purely time-like, we obtain the decoherent path integral:

\[ \text{Im}[S[q]] = 0 \implies \mathcal{O} = \frac{1}{Z} \int_{q \in \mathcal{P}} O[q] \exp(-\beta \text{Re}[E[q]]) \, dq \]  \hspace{1cm} (351)

and \( \mathcal{O} \) is here a thermal average of the observable.

Explicitly, the probability of each path in the full space-time region is:

\[ P[q] = \frac{1}{Z} \exp(-\beta \text{Re}[E[q]]) \exp(-i\beta \text{Im}[E[q]]) \]  \hspace{1cm} (352)

and exclusively in the time-like region (\( \text{Im}[E[q]] = 0 \)), the probability reduces to:

\[ P_{\text{time-like}}[q] = \frac{1}{Z} \exp(-\beta \text{Re}[E[q]]) \]  \hspace{1cm} (353)

Probabilities of the type \( P_{\text{time-like}}[q] \), in the Von Neumann density matrix formalism, are mixtures. For illustration, let us suppose \( n \) possible paths denoted as \( P_{\text{time-like}-1}, \ldots, P_{\text{time-like}-n} \), then the density matrix of this ensemble is:

\[ \hat{\rho} = \begin{pmatrix} P_{\text{time-like}-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_{\text{time-like}-n} \end{pmatrix} \]  \hspace{1cm} (354)

The absence of off-diagonal terms indicate that the system has decohered, and that no probability interference will be observed. The sum obeys a classical sum of real-valued probabilities. An observer will therefore interpret ‘coming into causal contact with a quantum system’ as performing a measurement on the system.

### 6.4 Non-relativistic limit

If our claim that a system is quantum in the space-like separated region (and decoherent in the time-like separated region) is correct, then it follows that the non-relativistic limit \( (v \ll c) \) produces, in the space-like separated region, the Schrödinger equation. We also intend to show that in the time-like separated region, the solutions are decoherent tunneling particles at the Unruh temperature.

The non-relativistic limit does not have horizons, nor does it have a time-like or space-like separated regions, however, the proper limit can still be obtained. To achieve this, we will take the limit for each of the two signatures of the metric. The first limit, taken with signature \((-,+,+,+\)) produces the Schrödinger
equation whereas the second limit, taken with signature \((+,-,-,-)\) produces the tunneling solutions. For simplicity, we will work in 1+1 space-time.

For the Schrödinger equation we start with:

\[
Z[\mathcal{P}] = \int_{q \in \mathcal{P}} \exp \left( \frac{2\pi mc}{\hbar} \int_q ds \right) Dq \tag{355}
\]

Using the metric with signature \((-,+,-,+),\) we inject \(ds = \pm \sqrt{dx^2 - dt^2}:\)

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi mc \frac{1}{\hbar} \int_q \sqrt{dx^2 - dt^2} \right) Dq \tag{356}
\]

We factor out dt, and we pose \(dx/dt := \dot{x}:\)

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi mc \frac{1}{\hbar} \int_q dt \sqrt{\dot{x}^2 - 1} \right) Dq \tag{357}
\]

Now we take the Taylor expansion of \(\sqrt{\dot{x}^2 - 1}\) with respect to \(\dot{x}^2,\) we get:

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi i \frac{mc}{\hbar} \int_q dt \left( i - \frac{i\dot{x}^2}{2} + O[\dot{x}]^4 \right) \right) Dq \tag{358}
\]

Factoring \(i\) and neglecting \(O[\dot{x}]^4,\) the non-relativistic limit \(v \ll c\) is:

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi i \frac{mc}{\hbar} \int_q dt \left( -\frac{\dot{x}^2}{2} + 1 \right) \right) Dq \tag{359}
\]

Finally, we absorb the \(+1\) term into a general potential \(V[q].\) We get:

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi i \frac{mc}{\hbar} \int_q dt \left( -\frac{\dot{x}^2}{2} + V[q] \right) \right) Dq \tag{360}
\]

This is the sufficient starting point to derive the Schrödinger equation from the Feynman path integral formulation. We note that the factor \(2\pi\) is there because of the relationship between action and entropy, and thus, the Schrödinger-like equation we obtain will describe the dynamics of the entropy representing the particle instead of the dynamics of the particle itself — the two being related by a factor \(2\pi.\)

To obtain the tunneling solutions, we use the metric with signature \((+,-,-,-).\) Repeating the previous steps (omitted), we eventually obtain:
\[
Z[\mathcal{P}] = \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi \frac{mc}{\hbar} \int_q dt \sqrt{1 - \dot{x}^2} \right) Dq
\]

(361)

Now we take the Taylor expansion of \(\sqrt{1 - \dot{x}^2}\) with respect to \(\dot{x}^2\), we get:

\[
= \int_{q \in \mathcal{P}} \exp \left( \pm 2\pi \frac{mc}{\hbar} \int_q dt \left( 1 - \frac{\dot{x}^2}{2} + O[\dot{x}^4] \right) \right) Dq
\]

(362)

Finally, we note that for an accelerated observer \(a \neq 0\), the tunneling temperature is the Unruh temperature:

\[
= \int_{q \in \mathcal{P}} \exp \left( -\frac{2\pi c}{\hbar a} \int_q dt \left( -\frac{\dot{x}^2}{2} + V[q] \right) \right) Dq
\]

(363)

\[\text{Lagrangian} \quad E[q] \quad \text{E}[q] \quad \text{Thermal states: } -\beta E[q] \]

With the Lagrangian being that of a non-relativistic particle tunneling through a potential.

### 6.5 Measurement

In the non-relativistic limit, the space-like part of the path integral reduces to the Schrödinger equation, which uses probability amplitudes. For instance, a quantum state such as \(|\psi\rangle = \alpha |\phi_1\rangle + \beta |\phi_2\rangle\), has the following density matrix:

\[
\hat{\rho} = \begin{pmatrix}
\alpha \alpha^* & \alpha \beta^* \\
\alpha^* \beta & \beta \beta^*
\end{pmatrix}
\]

(364)

and in the time-like part, the density matrix decoheres to:

\[
\hat{\rho} = \begin{pmatrix}
P_{\text{time-like-1}} & 0 \\
0 & P_{\text{time-like-2}}
\end{pmatrix}
\]

(365)

which is a post-measurement mixture.

In the first case, this density matrix is a pure state also, and consequently, its Von Neumann entropy is 0. But in the second case, the density matrix has no off-diagonal terms, and thus the system has experienced decoherence. Consequently, decoherence occurs at the boundary of the light cone of the
observer. We have thus obtained a mathematical description of the common intuition that a quantum description is a past description of presently available classical information (i.e. informally, the system used to be quantum before it "measured it" as they "interacted/came into causal contact" with each other. More formally, the probability amplitudes became decoherent as the path integral was continued into the time-like region).

We note that the system may behave as if it were classical (loss of interference) and that it is a post-measurement mixture. As the probability distribution of the system inside the light-cone is the same as the probability distribution of the paths over the full space-time, the probabilities are conserved as the system exits the horizon as a decoherent system, and no quantum information should be lost by crossing the boundary (although since the information is thermal, it cannot be used to do work).

The final step, involving the selection of a single specific state of the mixture, is discussed in section 8.1 regarding the interpretation of quantum mechanics.

6.6 Normalization (sketch)

A path integral with both a real thermal part and a imaginary quantum part will exponentially suppress infinite energy terms, and should remain normalizable at high energies (for the same reason that a thermal system has 0% occupancy probability of the infinite energy solutions.). The high energy spectrum is dominated by a specific temperature (Unruh/Hawking), which becomes the normalization condition for the energy levels of the system.

7 Results (Poly-Geometry)

7.1 The length of poly-vectors

So far, we have used the metric form exclusively for geometric partition functions that are 1-vectors. But, how do we define the same in the case of a poly-vector? As will we argue, our definition of the metric form extends the notion of length to that of any poly-vectors.

Let us recall that for a vector its length is defined as the scalar value of its inner product:

$$\|v\|^2 := v \cdot v \quad (366)$$

For instance if $v := \sigma_x x + \sigma_y y$ (of the $Cl_{0,3}(\mathbb{R})$ algebra), then:

$$v \cdot v = x^2 + y^2 \quad (367)$$

In some publications, we find that the length of a poly-vector is defined as a simple extension of the usual inner product but applied to all components of
the poly-vector. For instance, if $\mathbf{u} := a + \sigma_x x + \sigma_y y + \sigma_z b$ (of the $\text{Cl}_{0,3} (\mathbb{R})$ algebra), then its inner product would be defined as:

$$\|\mathbf{u}\|^2 = a^2 + x^2 + y^2 + b^2$$  \hspace{1cm} (368)

The definition might be a valid mathematical norm, but it is physically incorrect. To help us understand what goes wrong with this definition, let us compare it to the results from our definition (Metric-form, Definition [16]).

1. 0-vector: $\mathbf{v} := a$ of the $\text{Cl}_{0,3} (\mathbb{R})$ algebra. Diagonalizing its matrix representation (in this case its already diagonal), we obtain:

$$\|\hat{\mathbf{v}}\| = P^{-1} \hat{\mathbf{v}} P = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = a \mathbf{1}$$  \hspace{1cm} (369)

Its inner-product and metric-form are, respectively:

$$\sqrt{\mathbf{v} \cdot \mathbf{v}} = a \hspace{1cm} \text{inner-product}$$  \hspace{1cm} (370)

$$\|\hat{\mathbf{v}}\| = a \hspace{1cm} \text{metric-form}$$  \hspace{1cm} (371)

In this case, the results are the same.

2. 1-vector: $\mathbf{v} := \sigma_x x + \sigma_y y + \sigma_z z$ of the $\text{Cl}_{0,3} (\mathbb{R})$ algebra. Diagonalizing its matrix representation:

$$\hat{\mathbf{v}} = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} \rightarrow P^{-1} \hat{\mathbf{v}} P = \begin{pmatrix} -\sqrt{x^2 + y^2 + z^2} & 0 \\ 0 & \sqrt{x^2 + y^2 + z^2} \end{pmatrix}$$  \hspace{1cm} (372)

Its inner-product and metric-form are, respectively:

$$\sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x^2 + y^2 + z^2} \hspace{1cm} \text{inner-product}$$  \hspace{1cm} (373)

$$\|\hat{\mathbf{v}}\| = \pm \sqrt{x^2 + y^2 + z^2} \hspace{1cm} \text{metric-form}$$  \hspace{1cm} (374)

where here we have used the $\pm$ notation to group the two eigenvalues as one expression.

In this case, the absolute value of the results is the same, but one of our solutions has a minus sign.

3. 1-vector (≠ 2): $\mathbf{v} := \gamma_x x + \gamma_y y + \gamma_z z + \gamma_t t$ (of the $\text{Cl}_{1,3} (\mathbb{R})$ algebra). Diagonalizing its matrix representation:
\[ \hat{v} = \begin{pmatrix} t & 0 & z & x - iy \\ 0 & t & x + iy & -z \\ -z & -x + iy & -t & 0 \\ -x - iy & z & 0 & -t \end{pmatrix} \]  

(375)

\[ \Rightarrow P^{-1} \hat{v} P = \begin{pmatrix} -\sqrt{-x^2 - y^2 - z^2 + t^2} & 0 & 0 & 0 \\ 0 & -\sqrt{-x^2 - y^2 - z^2 + t^2} & 0 & 0 \\ 0 & 0 & \sqrt{-x^2 - y^2 - z^2 + t^2} & 0 \\ 0 & 0 & 0 & \sqrt{-x^2 - y^2 - z^2 + t^2} \end{pmatrix} \]  

(376)

Its inner-product and metric-form are, respectively:

\[ \sqrt{\hat{v} \cdot \hat{v}} = \sqrt{x^2 + y^2 + z^2 + t^2} \]  

"naive" inner-product  

(377)

\[ ||\hat{v}|| = \pm \sqrt{-x^2 - y^2 - z^2 + t^2} \]  

metric-form  

(378)

where here we have used the \( \pm \) notation to group the two eigenvalues as one expression.

We note that using a suitable inner product, one could obtain the interval of special relativity, but one must adjust the inner product definition to account for the metric signature whenever one changes the geometric algebras. If the inner product is redefined as \( u \cdot v = \eta(u, v) \), then the length is:

\[ \sqrt{\hat{v} \cdot \hat{v}} = \sqrt{-x^2 - y^2 - z^2 + t^2} \]  

(379)

However, using our definition, no change in the definition is required as we go along.

4. poly-vector: Now consider a poly-vector \( u := a + \sigma_x x + \sigma_y y + \sigma_z z \) (of the \( Cl_{0,3}(\mathbb{R}) \) algebra). Diagonalizing its matrix representation:

\[ \hat{u} = \begin{pmatrix} z + a & x - iy \\ x + iy & -z + a \end{pmatrix} \]  

(380)

we get:

\[ P^{-1} \hat{u} P = \begin{pmatrix} a - \sqrt{x^2 + y^2 + z^2} & 0 \\ 0 & a + \sqrt{x^2 + y^2 + z^2} \end{pmatrix} \]  

(381)
Let us now compare the metric-form to the inner-product:

\[
\sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{a^2 + x^2 + y^2} \quad \text{"naive" inner-product} \quad (382)
\]

\[
||\hat{\mathbf{u}}|| = a \pm \sqrt{x^2 + y^2} \quad \text{metric-form} \quad (383)
\]

where we have used the ± notation to group the two eigenvalues as one expression.

Here the departure between the two definitions is quite remarkable; we have a non-Euclidean contribution to the metric-form by the element $a$.

Let us investigate the contribution of $a$ to the metric-form and try to understand why the scalar part finds itself outside the square root. To do so, let us contrast two practical examples:

- Say one draws a Cartesian graph with two axes: $x$ and $y$. One then places a token at the origin $(0, 0)$. Then, say one moves the token 3 units on the $x$-axis, followed by 4 units on the $y$-axis. After these translations, one will find the token at point $(3, 4)$. The total distance the token has moved along the path is $3 + 4 = 7$ units. This is not the shortest path to $(3, 4)$ however. Indeed, since the axes are orthogonal one could have instead moved the token in a straight line to $(3, 4)$. In this case, the token would have moved 5 units along this path.

- Say one draws a Cartesian graph with two axes: the $x$-axis denotes the quantity of apple, and the $y$-axis denotes the number of oranges. Say one wishes to procure 3 apples and 4 oranges. Question: can one abuse the Pythagorean theorem to obtain the desired quantities of fruit by acquiring only 5 units of fruit? The answer is obviously no, and the reason is that the appropriate metric for this situation is, contrary to a graph drawn in the 2d-plane, not the Euclidean metric, but instead the taxi-cad metric. Explicitly, the distance —measured in fruits— between $(0, 0)$ and $(3, 4)$ is given by $d = \Delta x + \Delta y = 3 + 4$ and not $d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$.

Keeping these examples in mind, let us consider the case of the poly-vector $\mathbf{u} := a + b + c + \sigma_x x + \sigma_y y + \sigma_z z$ (of the $\text{Cl}_{0,3} (\mathbb{R})$ algebra). Its metric-form would be:

\[
||\hat{\mathbf{u}}|| = \underbrace{a + b + c}_{\text{taxi-cab element}} \pm \underbrace{\sqrt{x^2 + y^2 + z^2}}_{\text{euclidean element}} \quad (384)
\]

In this example, the metric-form is a combination of euclidean elements (the orthogonal terms) and taxi-cab elements (the scalar terms). The metric explains why one cannot shortcut its way to $a, b, c$ by changing its position given by the Cartesian coordinates $\sqrt{x^2 + y^2 + z^2}$. If one had instead taken the "naive" inner product as the length $\sqrt{a^2 + b^2 + c^2 + x^2 + y^2 + z^2}$, one would have erroneously...
believed the path to 3 apples and 4 oranges can be made shorter by moving along the plane.

Even more complexity to the metric-form is possible. Indeed, complex elements appear when one uses extended elements (e.g. k-blades $e_i \wedge e_j$, $e_i \wedge e_j \wedge e_k$, etc.). Consider these few more examples:

1. **complex-number**: $z := \alpha + \beta i$. The matrix representation is:

$$
\hat{z} = \begin{pmatrix}
\alpha & -\beta \\
\beta & \alpha
\end{pmatrix}
$$

(385)

and the diagonal matrix is:

$$
P^{-1}\hat{z}P = \begin{pmatrix}
\alpha - i\beta & 0 \\
0 & \alpha + i\beta
\end{pmatrix}
$$

(386)

The metric-form is:

$$
\|\hat{z}\| = \alpha \pm i\beta
$$

(387)

2. **quaternions**: $q := a + bi + cj + dk$. Using the basis of $Cl_{0,3}(\mathbb{R})$, a quaternion is $q = a + \sigma_y \sigma_z b + \sigma_x \sigma z c + \sigma_x \sigma_y d$. Then, its matrix representation is:

$$
\hat{q} = \begin{pmatrix}
\alpha + ib & ic + d \\
-ic -d & \alpha - i b
\end{pmatrix}
$$

(388)

and its diagonalization is:

$$
P^{-1}\hat{q}P = \begin{pmatrix}
\alpha - i\sqrt{b^2 + c^2 + d^2} & 0 \\
0 & \alpha + i\sqrt{b^2 + c^2 + d^2}
\end{pmatrix}
$$

(389)

The metric-form is:

$$
\|\hat{q}\| = a \pm i\sqrt{b^2 + c^2 + d^2}
$$

(390)

For instance consider a general poly-vector of $Cl_{0,3}(\mathbb{R})$:

$$
u := U + \sigma_x x + \sigma_y y + \sigma_z z + \sigma_x \sigma_y a_{xy} + \sigma_x \sigma_z a_{xz} + \sigma_y \sigma_z a_{yz} + \sigma_x \sigma_y \sigma_z V
$$

(391)

The matrix representation of this vector is:
\[
\hat{u} = \begin{pmatrix}
ia_{xy} + U + iV + z & ia_{yz} + a_{xz} + x - iy \\
-ia_{yz} - a_{xz} + x + iy & -ia_{xy} + U + iV - z
\end{pmatrix}
\]  
(392)

We diagonalize the matrix to find the eigenvalues:

\[
P^{-1} \hat{u} P = \begin{pmatrix}
(U + iV) + \sqrt{(x + ia_{yz})^2 + (y + ia_{xz})^2} & 0 \\
0 & (U + iV) - \sqrt{(x + ia_{yz})^2 + (y + ia_{xz})^2}
\end{pmatrix}
\]  
(393)

and we find the metric-form:

\[
||\hat{u}|| = (U + iV) \pm \sqrt{(x + ia_{yz})^2 + (y + ia_{xz})^2 + (z + ia_{xy})^2}
\]  
(394)

where here we have used the \pm notation to group the two eigenvalues as one expression.

Finally, we state the general poly-vector of \(Cl_{1,3}(\mathbb{R})\):

\[
v : = G + \gamma_0 t + \gamma_1 x + \gamma_2 y + \gamma_3 z \\
+ \gamma_0 \gamma_1 A_{01} + \gamma_0 \gamma_2 A_{02} + \gamma_0 \gamma_3 A_{03} + \gamma_1 \gamma_2 A_{12} + \gamma_1 \gamma_3 A_{13} + \gamma_2 \gamma_3 A_{23} \\
+ \gamma_0 \gamma_1 \gamma_2 V_{012} + \gamma_0 \gamma_1 \gamma_3 V_{013} + \gamma_0 \gamma_2 \gamma_3 V_{023} + \gamma_1 \gamma_2 \gamma_3 V_{123} \\
+ \gamma_0 \gamma_1 \gamma_2 \gamma_3 U
\]  
(395)

The matrix representation of \(v\) is:

\[
\hat{v} = \begin{pmatrix}
G + iA_{12} - iV_{012} & A_{13} - iA_{23} + V_{013} - iV_{023} & -iU + z + A_{03} - iV_{023} & x - iy + A_{01} - iA_{02} \\
-A_{13} - iA_{23} - V_{013} + iV_{023} & G + t + iA_{21} + iV_{012} & x + iy + A_{01} + iA_{02} & -iU - z - A_{03} - iV_{012} \\
-x + iy + A_{01} + iA_{02} & -iU - z - A_{03} + iV_{012} & G - t + iA_{21} + iV_{012} & A_{13} - iA_{23} - V_{013} + iV_{023} \\
x - iy + A_{01} - iA_{02} & -iU + z + A_{03} + iV_{023} & -A_{13} - iA_{23} + V_{013} + iV_{023} & G - t + iA_{21} - iV_{012}
\end{pmatrix}
\]
(396)

Diagonalizing this matrix is left as an exercise.

### 7.2 Potentials and taxi-cab terms

Potentials can be added to the Lagrangian as scalar terms. For instance, one may consider the movement of a test particle in curved-space under the effect of a potential \(V\). Its action \(S\) would be:

\[
S = \int_a^b \left(-mc \sqrt{g_{\mu\nu} \frac{\partial x^\mu}{\partial \tau} \frac{\partial x^\nu}{\partial \tau}} + V\right) d\tau
\]  
(397)

What part of the geometry of space-time does \(V\) live in...? The question may seem bizarre as \(V\) is not a term of the metric. However, using the metric-form
of a poly-vector, potentials and other scalar terms can be made to live in the
geometry. Consider a geometric ensemble with the following equation of state:

\[ dS = f_{e_0} dX_0 + \hat{k} e_1 dX_1 + \hat{k} e_2 dX_2 + \hat{k} e_3 dX_3 + \lambda_a dU_a + \lambda_b dU_b + \lambda_c dU_c \]  

(398)

The metric-form is:

\[ dS = \pm \hat{k} \sqrt{g_{\mu\nu} dX^\mu dX^\nu} + \lambda_a dU_a + \lambda_b dU_b + \lambda_c dU_c \]  

(399)

By parametrization of the entropy over a path \( t \in [a, b] \), one obtains:

\[ \int_a^b \frac{\partial S}{\partial \tau} d\tau = \pm \int_a^b \left( \hat{k} \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} + \lambda_a \nabla_a + \lambda_b \nabla_b + \lambda_c \nabla_c \right) d\tau \]  

(400)

where \( V := \frac{\partial U}{\partial \tau} \).

As before, and by the calculus of variation \( \delta \), the equations of motions are obtained as those which extremalize the production of entropy in space-time now under the added effect of the potential terms.

This action is then incorporated into a Feynman-type path integral

### 7.3 Area terms and electromagnetism

The bi-vectors of \( Cl_{1,3}(\mathbb{R}) \) are:

\[ \gamma_0 \gamma_1, \gamma_0 \gamma_2, \gamma_0 \gamma_3, \gamma_1 \gamma_2, \gamma_1 \gamma_3, \gamma_2 \gamma_3 \]  

(401)

One can construct an ensemble using all 2-basis generators. The equation of state would be:

\[ dS = \lambda_{E_z} \gamma_0 \gamma_1 d\overline{E}_x + \lambda_{E_y} \gamma_0 \gamma_2 d\overline{E}_y + \lambda_{E_z} \gamma_0 \gamma_3 d\overline{E}_z + \lambda_{B_x} \gamma_1 \gamma_2 d\overline{B}_x + \lambda_{B_y} \gamma_1 \gamma_3 d\overline{B}_y + \lambda_{B_z} \gamma_2 \gamma_3 d\overline{B}_z \]  

(402)

where \( \overline{E}_x, \overline{E}_y, \overline{E}_z, \overline{B}_x, \overline{B}_y, \overline{B}_z \) are the constraints, and where \( \lambda_{E_x}, \lambda_{E_y}, \lambda_{E_z}, \lambda_{B_x}, \lambda_{B_y}, \lambda_{B_z} \) are the Lagrange multipliers.

Posing \( \lambda_E := \lambda_{E_z} = \lambda_{E_y} = \lambda_{E_x} \) and \( \lambda_B := \lambda_{B_x} = \lambda_{B_y} = \lambda_{B_z} \), the matrix representation of \( d\hat{S} \) is:

\[
\begin{pmatrix} 
-iB_x \lambda_B & (-iB_x - B_y) \lambda_B & E_z \lambda_E & (E_x - iE_y) \lambda_E \\
(-iB_x + B_y) \lambda_B & iB_z \lambda_B & (E_x + iE_y) \lambda_E & -E_x \lambda_E \\
E_z \lambda_E & (E_x - iE_y) \lambda_E & -iB_z \lambda_B & (-iB_x - B_y) \lambda_B \\
(E_x + iE_y) \lambda_E & -E_z \lambda_E & (iB_z + B_y) \lambda_B & iB_x \lambda_B 
\end{pmatrix}
\]  

(403)
Using diagonalization, the metric-form is:

\[(dS)^2 = \lambda_E^2(dE_x^2 + dE_y^2 + dE_z^2) - \lambda_B^2(dB_x^2 + dB_y^2 + dB_z^2) \pm 2i\lambda_E\lambda_B(dE_x dB_y + dE_y dB_z + dE_z dB_x)\] (404)

We can rewrite it as:

\[(dS)^2 = \lambda_E^2||dE||^2 - \lambda_B^2||dB||^2 \pm 2i\lambda_E\lambda_B(dE \cdot dB)\] (405)

These are the Lorentz invariants of electromagnetism. The Maxwell equation is obtained under the constraint \(dS \to 0\). We recall the familiar equality between the tensor representation and the invariant representation, as follows:

\[\frac{1}{2}F_{ab}F^{ab} = ||B||^2 - \frac{1}{c^2}||E||^2\] (406)

\[\frac{1}{4}\epsilon^{abcd}F_{ab}F_{cd} = \frac{1}{2}B \cdot E\] (407)

We have thus obtained a purely geometric representation of electromagnetism, in which the invariant is the metric-form of a 2-vector (just as the invariant of special/general relativity is obtained as the metric-form of a 1-vector).

As a sketch, and by repeating the steps previously done to produce the Feynman path integral, we can develop a quantized version of electromagnetism. The idea is to take two events \(q_1\) and \(q_2\), then to convert the sum to a sum over paths between these events. Here, the event \(q_1\) and \(q_2\) are geometric events expressed in the 2-basis. The result is a path integral using the ‘interval/metric-form’ of the 2-basis elements as the action, which corresponds to electromagnetism.

### 7.4 Nambu-Goto action

Using the flexibility of geometric algebra, even more variety of ensembles using the k-basis elements of \(Cl_{1,3}(\mathbb{R})\) can be also constructed. For instance, consider an ensemble with the following constraint:

\[(e_\tau \wedge e_\sigma)\vec{A} = \sum_{q \in Q}(e_\tau \wedge e_\sigma)A[q]\hat{g}[q]\] (408)

Using the machinery of geometrical statistical physics, the following equation of state is eventually obtained:

\[d\hat{S} = \lambda(e_\tau \wedge e_\sigma) d\vec{A}\] (409)

where \(\lambda\) is a Lagrange multiplier with the appropriate units. The entropic-metric, applicable to k-vectors, is thus:
The vectors $e_\tau$ and $e_\sigma$ are tangent to the surface $X$ of some manifold $M$. Paramaterized using $(\tau, \sigma)$, the vectors can be written as:

$$e_\tau = \frac{\partial X[\tau, \sigma]}{\partial \tau}$$
$$e_\sigma = \frac{\partial X[\tau, \sigma]}{\partial \sigma}$$

With said parameterization, the term $d\bar{A}$ can be written as:

$$d\bar{A} = A[X[\tau, \sigma]] \, d\tau \, d\sigma$$

Consequently, $d\hat{S}$ can be rewritten as:

$$dS = \pm \lambda d\bar{A}[X[\tau, \sigma]] \sqrt\left(\frac{\partial X[\tau, \sigma]}{\partial \tau} \land \frac{\partial X[\tau, \sigma]}{\partial \sigma}\right)^2 \, d\tau \, d\sigma$$

**Theorem 5.** The entropic 2-metric is the Nambu-Goto action.

**Proof.** The term:

$$\sqrt\left(\frac{\partial X[\tau, \sigma]}{\partial \tau} \land \frac{\partial X[\tau, \sigma]}{\partial \sigma}\right)^2$$

can be rewritten as:

$$\sqrt\left(- \frac{\partial^2 X[\tau, \sigma]}{\partial \tau^2} \frac{\partial^2 X[\tau, \sigma]}{\partial \sigma^2} + \left(\frac{\partial X[\tau, \sigma]}{\partial \tau} \cdot \frac{\partial X[\tau, \sigma]}{\partial \sigma}\right)^2\right)$$

The proof is taken from [61]. First, the following common identities are noted by the author:

$$e_\tau \land e_\sigma = e_\tau e_\sigma - e_\tau \cdot e_\sigma$$
$$e_\tau \land e_\sigma = -e_\tau \land e_\sigma$$
$$e_\tau \cdot e_\sigma = e_\sigma \cdot e_\tau = \frac{1}{2}(e_\tau e_\sigma + e_\sigma e_\tau)$$

Then the proof given by [61] goes as follows:
\[(e_\tau \wedge e_\sigma)^2 = -(e_\tau \wedge e_\sigma)(e_\sigma \wedge e_\tau) \tag{421}\]
\[= -(e_\tau e_\sigma - e_\tau \cdot e_\sigma)(e_\sigma e_\tau - e_\sigma \cdot e_\tau) \tag{422}\]
\[= -e_\tau e_\sigma e_\tau e_\sigma + e_\tau e_\sigma (e_\sigma \cdot e_\tau) + (e_\tau \cdot e_\sigma) e_\tau e_\sigma - (e_\tau \cdot e_\sigma)(e_\sigma \cdot e_\tau) \tag{423}\]
\[= -e_\tau^2 e_\sigma^2 + (e_\sigma \cdot e_\tau)(e_\tau e_\sigma + e_\sigma e_\tau) - (e_\sigma \cdot e_\tau)^2 \tag{424}\]
\[= -e_\tau^2 e_\sigma^2 + (e_\sigma \cdot e_\tau)^2(e_\sigma \cdot e_\tau) \tag{425}\]
\[= -e_\tau^2 e_\sigma^2 + (e_\sigma \cdot e_\tau)^2 \tag{426}\]

Let \(\mathcal{A}[X[\tau, \sigma]] = 1\), then:

\[dS = \pm \lambda \sqrt{-\frac{\partial^2 X[\tau, \sigma]}{\partial \tau^2} \frac{\partial^2 X[\tau, \sigma]}{\partial \sigma^2} + \left(\frac{\partial X[\tau, \sigma]}{\partial \tau} \cdot \frac{\partial X[\tau, \sigma]}{\partial \sigma}\right)^2} d\tau d\sigma \tag{427}\]

Using the action-to-entropy relation, we finally obtain:

\[\mathcal{A} = \pm \frac{T_0}{c} \iiint \sqrt{-\frac{\partial^2 X[\tau, \sigma]}{\partial \tau^2} \frac{\partial^2 X[\tau, \sigma]}{\partial \sigma^2} + \left(\frac{\partial X[\tau, \sigma]}{\partial \tau} \cdot \frac{\partial X[\tau, \sigma]}{\partial \sigma}\right)^2} d\tau d\sigma \tag{428}\]

where we define \(T_0 \equiv c \hbar \lambda\), with units \(\text{[J/m]}\) as the tension.

The same derivation can be generalized to multiple dimensions using a k-metric to derive the action of d-branes.

\[\Box\]

8 Discussion

We will now apply axiomatic science to unresolved problems within the foundation of physics. As stated in the introduction, unlike artificial models containing only a ‘physics’ part, axiomatic science contains both a ‘science’ part and a ‘physics’ part. Axiomatic science is thus able to explain the origins of the laws of physics using science as the starting point, whereas for an artificial model, the origin of such laws is a blind spot. It will then be within this newly illuminated blind spot that we will find the solutions proposed by axiomatic science.

As a disclaimer, we state that in this paper, we are only beginning to scratch the surface of axiomatic science. Therefore our intention is not to completely resolve all of these problems, but rather to present the case made by axiomatic science and also to encourage future research.

8.1 Interpretation of quantum mechanics

What is generally considered ‘physical/quantum randomness’; that is the random selection of a state from a post-measurement mixture, is not an add-on in
axiomatic science, but its very foundation. The quantum measurement problem is thus identified and corrected within the 'science' part of the framework before the laws of physics are even derived. First, let us recall what the quantum measurement is.

In quantum physics, the unitary evolution of the wave function is deterministic, but the notion breaks down if measurements are occurring in the system (or are performed on the system). In the Von Neumann scheme, a measurement of the second kind, for a quantum object with wave function $|\psi\rangle$ and a quantum apparatus with wave function $|\phi\rangle$, is defined as:

$$|\psi\rangle|\phi\rangle \rightarrow \sum_{n} c_{n}|\chi_{n}\rangle|\phi_{n}\rangle$$

After the measurement the system is in one of eigenstates $|\chi_{n}\rangle$ with probability $|c_{n}|^{2}$. That the otherwise deterministic unitary evolving system adopts the "undeterministic" initiative to collapse itself randomly in one of multiple states after measurement is quite the mystery. Nothing in quantum physics predicts that such a thing would occur. Consequently, the notion of the measurement is introduced into quantum mechanics, formally, as a full-blown axiom not derivable from the Schrodinger equation itself, or any of the other axioms of quantum mechanics.

The theory of quantum decoherence is a modern take on Von Neumann measurements. De-coherence from the environment $|e\rangle$ is introduced as follows:

$$|\psi\rangle|\phi\rangle|e\rangle \rightarrow \sum_{n} c_{n}|\chi_{n}\rangle|\phi_{n}\rangle|e_{n}\rangle$$

Under contact with an environment having multiple degrees of freedom, any interference pattern normally observable from $|\psi\rangle$ will be smudged by the environment beyond the ability of instruments to detect it. De-coherence explains why a quantum superposition of eigenstates is unlikely to be observed macroscopically as interaction with the environment very quickly causes the system to evolve towards a classical probability distribution. However, de-coherence is ultimately of no help in regards to explaining why one eigenstate out of many is randomly selected for the system to be in, post-measurement.

In the results of this manuscript, we have presented an alternative take on decoherence; specifically, we have shown that decoherence occurs at the time-like/space-like boundary of the path integral over the metric form of the ensemble: as the quantum system becomes causally connected to the observer, its probability distribution decoheres/loses-its-ability-to-interfere, and the system becomes a mixture of states instead of a pure state.

But, how does the system go from a mixture of states to selecting a specific state from the mixture, post-measurement? The measurement postulate, as a law, is derived empirically and it is introduced purely so that quantum physics predicts a single macroscopic world (not a superposition of many worlds), consistent with
observations. The two primary competing interpretations of this behavior are a) the Copenhagen interpretation and b) the Everett many-worlds interpretation, but there exists at least half a dozen others. None of these interpretations are, however, considered satisfactory by mainstream physics and thus the question remains unsettled; in the first case the collapse is simply postulated but no mechanisms are generally accepted for it, and in the second case it is postulated that the observer becomes coupled with a specific result of the measurement causing the appearance of collapse, but no mechanism to account for this coupling is generally accepted either. The interpretational problem is retained in all extensions of quantum theory from the Dirac equation to quantum field theory, etc. As one is generally free to apply any of the compatible interpretations to quantum theory, deciding which one is correct, if any, is often criticized as a non-falsifiable problem.

Axiomatic science proposes to address the problem from the other direction. The primary idea is that the quantum measurement is quantified, in axiomatic science, by natural information. We recall that in Shannon’s theory of information, entropy quantifies the amount of information one gains by knowing which message is randomly selected from a set of possible messages. In the present case, axiomatic science postulates that the reference manifest (Axiom 1), describing the state of affairs of the world (Assumption 1), is randomly selected from the set of all possible manifests (Assumption 2). Natural information is then quantified by the entropy associated with the message (Definition 5). This is the ‘science’ part.

How and why do the laws of physics acquire a quantum measurement problem? To derive the ‘physics’ part from the ‘science’ part, one must at some step of the proof maximize the entropy of natural information. Maximizing the entropy of natural information has the consequence of erasing natural information. This renders natural information unavailable to the laws of physics, derived afterward and as a consequence of the erasure. One who then uses said derived laws of physics to find solutions/manifests will unavoidably encounter this entropy within the solutions. Indeed, post erasure, the meaning of natural information is clarified: essentially, natural information represents the total amount of information about nature that cannot be derived by the laws of physics. Intuitively, the observer, having access to natural information, ‘sees’ the reference manifest, but the laws of physics, derived from axiomatic science as the consequence of maximizing the entropy of natural information, ‘recover’ solutions only up to natural entropy. Consequently, there is an information gap between what is known to the observer (the reference manifest) and what is derivable by the laws of physics (the set of all manifests). The gap is precisely the sum-total of all quantum measurements required to connect the predictions of physics to the actual manifest.

The quantum measurement problem is acquired as a problem only when one constructs an artificial model of nature because such models are blind to the ‘science’ part, and thus cannot account for the origin of the laws of physics. We recall that an artificial model is produced when one postulates the laws of physics, then solves for manifests, and a natural model is produced when one postulates the manifests then solves for the laws of physics. In the artificial case,
one obtains a plurality of manifests as possible solutions; only one of which is the reference manifest. Since one intuitively expects that the laws of physics ought to explain the reference manifest, one may then become baffled as to why the laws of physics produce a plurality of manifests as their solutions. The culprit is identified by axiomatic science: natural information must be erased to derive the laws of physics using science.

So, why do we need to erase natural information to recover the laws of physics? The fundamental motivation is to release the description from the shackles of natural information to facilitate formulating the broadest possible patterns about nature. However, one cannot form a pattern from a single existing candidate (there is only one reference manifest), unless one invents hypothetical alternatives (the set of all manifests). For example, I can say "I am a physicist, but I could have been a doctor instead", or I could say "I measured the spin up, but it could have been down". Although neither violates the laws of physics, in reality, one happened and the other didn’t. It is precisely because natural information is erased from the laws of physics that the claim 'both alternatives are compatible with the laws of physics' can be made. Consequently, the laws of physics will recover both alternatives as possible solutions but would be unable to determine which of the two occurred without access to natural information. Consider if I would have instead said: "I am a physicist, but I could have been superman". How credible is that claim? Clearly, being superman violates the laws of physics, whereas being a doctor doesn’t. Do we then want our laws of physics to rule out superman, but not the doctor, even though in reality, we got the physicist? Remarkably, we want our laws of physics to permit, not only the reference manifest but also all other possible manifests.

In the case of axiomatic science, this concept is taken to its maximum. The description of the state of affairs as a manifest is the most general description of such possible, and the entropy of natural information is maximized to generate the broadest possible rules. These rules are the laws of physics.

To better understand why the laws of physics are not derivable without erasing natural information, it helps to attempt the following challenge; can we derive the laws of physics without erasing natural information, perhaps in such a way that the theory remains aware of the reference manifest? One can try, but one will not recover the laws of physics; instead, one will obtain a manifest theory:

Definition 20 (Manifest theory). A manifest theory is a program \( p \) that outputs \( M \) when run on a universal Turing machine. Thus,

\[
\text{UTM}(p) = M
\]

We further qualify a manifest theory as elegant if it is the shortest program that outputs the reference manifest when run on said universal Turing machine.

The manifest theory is pure computation with no insight or patterns. Contrary to the solutions of the laws of physics, for the manifest theory, all alternatives
(e.g., I being a doctor instead of a physicist, or even being superman instead of a physicist) are equally impossible simply because they did not happen and therefore will not be outputted by the program. Consequently, the manifest theory does not understand "it could have happened, but it didn’t" as for it, "it did not happen = it cannot happen".

In contrast, the laws of physics, explicitly derived by maximizing natural entropy, are the broadest patterns one can formulate about nature. The patterns (law of physics) emerge as a fundamental relation, precisely because we erase the shackles of natural information which would otherwise fix the theory to a manifest theory.

Picking the laws of physics as our explanatory tool of choice, rather than the manifest theory, is a choice made by physicists who prefer understanding the world by patterns rather than by brute computation. This 'preference' is formalized implicitly by Assumption 2, the fundamental assumption of 'nature'. Indeed, when one is presented with a world that exists brutely, one is free to assume that it is randomly selected from the set of all possible worlds, provided that one can list all the possible worlds (having access to a universal Turing machine or general intelligence is required for this step). One identifies all alternatives to the present state of nature, then formulates a law that holds for all alternatives and including the present state.

In axiomatic science, only the reference manifest exists brutely. Listing the other manifests as hypothetical alternatives, a step necessary to identify a pattern is an algorithmic operation performed on a 'chalkboard' and does not grant the status of ontological existence to these alternative manifests. This is where the distinction between the interpretation of quantum mechanics offered by axiomatic science and the others, occurs. For axiomatic science, neither the collapse of the wave-function interpretation nor the many-world interpretation are acceptable interpretations, as there is never a situation where more than one solution has the property of existence. Axiomatic science correctly predicts that one solution is actual, the reference solution, knowable to the 'observer' as the reference manifest, but that any pattern identified from the erasure of natural information (e.g. the laws of physics) will not have this knowledge; that is, unlike the 'observer', the laws of physics sees natural entropy in lieu of natural information.

Due to the importance of this point, we will re-iterate it more rigorously:

Theorem 6 (QM interpretation proposed by axiomatic science). Axiomatic science states that there is no collapse (thus it rejects the Copenhagen interpretation), and also that the system was never in a superposition of many-worlds, to begin with (thus it rejects the many-world interpretation). Axiomatic science states that all alternative manifests are mathematical creations used to facilitate the formulation of the laws of physics as patterns, and thus, have no ontological properties. Axiomatic science predicts the discrepancy between what is observed, and what the laws of physics offers as solutions, without the introduction of ad hoc postulates, and quantifies the discrepancy using the entropy of natural information.
Proof. First, let us investigate standard quantum mechanics (QM). QM is an artificial model. As such, the laws of physics are postulated, then solutions are found. Due to this construction, there is ambiguity regarding the ontological status of its solutions and the model is open to falsification. For instance, let us suppose a quantum theory QM, with input \( h \), and solved for solutions. Here, \( h \) may be interpreted as a description of the specific physical system QM is solved for. One may write:

\[
\text{Solve}[\text{QM}, h] = \alpha_1 \langle \text{solution}_1 \rangle + \alpha_2 \langle \text{solution}_2 \rangle + \alpha_3 \langle \text{solution}_3 \rangle + \ldots
\] (432)

However, experimentalists report to us that they observe only one element of \( \{\langle \text{solution}_1 \rangle, \langle \text{solution}_2 \rangle, \langle \text{solution}_3 \rangle, \ldots\} \) with probability \( \alpha_1^* \alpha_1, \alpha_2^* \alpha_2, \alpha_3^* \alpha_3, \ldots \), respectively. The quantum measurement problem is concerned with explaining the discrepancy: solving to get a superposition of solutions is not the same as reporting one solution randomly selected from the set of solutions. Thus, the predictions do not match the observations. One then asks: are the other solutions real, and if so, where do they "go" post-observation? Were they ever real to begin with? etc.

In axiomatic science, the ontological existential qualifier is granted to the reference manifest \( \dagger M \) by Axiom 1, but no other manifest \( M \) receives this "award". The other non-existing manifests are not derived by observation, but by enumeration on a 'chalkboard' using a Turing machine. Thus, the set of all manifests is a "man-made/intellectual" construction. Knowing which manifest is the reference manifest from the set of all possible manifests according to the probability distribution \( \rho[M] \) defines natural information. To derive the laws of physics, one then erases natural information as one maximizes natural entropy. We recall that the mathematical definition of the world (Definition 6) is \( \dagger W = (\dagger W, \dagger M) \), where \( \dagger W \) is the set of all possible manifests, and \( \dagger M \) is the reference manifest. The laws of physics do not represent what the reference manifest obeys, but instead what the world obeys (not too far off, but still not identical). Incorrectly thinking that they do is the source of the confusion regarding quantum mechanical interpretations. Explicitly, it is \( \dagger W \) that implies the laws of physics, and not \( \dagger M \) by itself.

The discrepancy between what is reported by observers (the reference manifest) and what is obtained by solving the laws of physics (the set of all possible manifests) is not a surprise worthy of its own postulate, but instead, a theorem fully predicted and derivable from axiomatic science.

Specifically, in axiomatic science (AS), one can solve the implied physical laws for solutions:

\[
\text{Solve}[\text{AS}, h] = \alpha_1 \langle M_1 \rangle + \cdots + \alpha_n \langle M_n \rangle
\] (433)

But in this case, it is known to axiomatic science that one solution will not be like the others:

\[
\exists! M_i \in \{M_1, \ldots, M_n\} [M_i = \dagger M]
\] (434)
We note that had we solved the same problem but with added natural information $NI$, the algorithm would be able to output exactly what is observed:

$$\text{Solve}[AS, h, NI] = |\hat{M}\rangle$$  \hspace{1cm} (435)

In this case $\text{Solve}[AS, h, NI]$ is a manifest theory. At no point is there confusion about what the other non-existing solutions are or aren’t, and at no point does a quantum measurement postulate need to be inserted into the axiomatic foundation. With axiomatic science as the foundation, experimentalists no longer surprise us when they report observing only one element of the set of solutions, as this is exactly what we expect them to tell us in the first place.

All laws of physics that we have derived in this manuscript, from the Feynman path integral, to decoherence, to the Fermi-Dirac statistics of events, to de Sitter space, to the Bekenstein-Hawking entropy, to electromagnetism, to special relativity and general relativity, are here derived in the same manner: we postulate the reference manifest, list all alternatives (in principle), then maximize the entropy of natural information and we obtain the law as the equation of state. None of those laws could have been derived from the first principles of axiomatic science without first erasing natural information.

8.2 Classification of scientific theories

**Definition 21** (Classification of scientific theories). We can classify scientific theories based on the amount of natural information they contain:

- **Fundamental**: A scientific theory is fundamental if it is constructed as a consequence of maximizing the entropy of natural information. Its set of solutions is the set of all possible manifests, recovered up to the entropy of natural information. By definition, all laws of physics, and only laws of physics, are fundamental theories (Assumption 3).

- **Manifest**: A scientific theory is manifest (Definition 20) if it produces $M$ as its only result. The manifest theory is valid only for the current manifest. It expires as soon as the observer evolves slightly forward in time (at which point the reference manifest is replaced by a new slightly more entropic manifest, and consequently it requires a new manifest theory).

- **Empirical**: A scientific theory is empirical if it contains some but not all of natural information. Let $NI^*$ be a subset of $NI$. Its solutions are obtained as $\text{Solve}[AS, h, NI^*]$. The empirical theory also recovers a plurality of solutions, only one of which is actual. Because it contains some natural information, the solutions are a subset of the set of all possible manifests. The empirical theory is more tolerant to falsification than the manifest theory but less so than the fundamental theory.
As a toy example, let us consider natural selection and the tree of life using the various classifications. Understanding life by detailing the tree of life is conceptually similar to a manifest theory of life, whereas understanding it as a solution to the theory of natural selection is akin to an empirical theory. Remarkably, natural selection, as an empirical theory, will contain a measurement 'pseudo-problem' (i.e. a non-fundamental version of the measurement problem).

Let us briefly recall the axioms of natural selection:

1. Organisms within a population have different traits.
2. Offspring inherit a mixture of the traits of their parents.
3. The probability of survival is trait-dependant (and not all offspring survive).
   • (Conclusion): Consequently, a population will acquire the traits most likely for survival.

These three axioms and one conclusion assume a lot of physical baggage (organisms, reproduction, parents/offspring, survival, etc.). The presence of this physical baggage is a strong indicator that the theory is empirical. Consequently, it is aware of some natural information to the exclusion of some manifests. For instance, the tree of life of those manifests with Boltzmann brains are not solutions of natural selection.

The axioms of natural selection are not formal. However, if we suppose that a formalization of such exists, then we can use formal symbols for the investigation. Let $\text{NS}$ be the formalization of natural selection. We can now ask; how many trees of life are compatible with $\text{NS}$? The set of trees of life consistent with $\text{NS}$ is quite large, but should still be countable, and may even be finite. We can re-write the many-trees as a solution to $\text{NS}$:

$$\text{Solve}[\text{NS}, h, \text{NI}^*] = \alpha_1 |\text{tree}_1\rangle + \cdots + \alpha_n |\text{tree}_n\rangle$$

Should biologists now start wondering what happens to the other members of the many-trees whenever a new fossil is dug up? Surely, digging up the fossil of a lizard implies that lizard-less trees vanishes away from the set of remaining solutions. Perhaps the vanishing of incompatible solutions (as fossils are dug up) is not a concern because the trees of life can’t produce interference patterns in double-slit experiments — but alas, there is no requirement that they have to, as the probability amplitudes $\{\alpha_1, \ldots, \alpha_n\}$ can be constrained to the reals. Perhaps it is because the correct value of the probability amplitudes cannot be computed from $\text{NS}$ and thus the theory is not predictive — alas, we won’t know the true probability distribution until we discover many other trees of life. So, no: The measurement pseudo-problem of empirical theories, although present, is not considered fundamental, simply because the theory is not fundamental. In this case, the measurement pseudo-problem of an empirical theory occurs because the theory is a strongly compressed representation of reality. Digging up fossils contributes natural information to the theory which slowly brings it closer
to being a manifest theory, as the alternative solutions not set by the axioms are eliminated by the contribution.

Axiomatic science clarifies the relationship between natural information, theory, and solutions. Starting with a group of empirical theories, one can add natural information to bring the group closer to a manifest theory, or remove natural information to bring it closer to a fundamental theory. The process is likely to converge (in either direction) via a series of discrete steps which is why informally practicing science in the wild (may) eventually produce the correct fundamental theory, with or without knowledge/usage of axiomatic science. Indeed, the manifest theory is immediately falsified by the next event, however an empirical theory which contains some natural information (but not all) will be expected to survive falsification longer. Finally, when natural information is completely eliminated, the measurement problem does not go away; rather it reaches a maximum — which is where the laws of physics live.

8.3 Time

If the world exists as a brute manifest (Axiom 1), why does an observer believe he/she has a future and a past? Recall that both the past and the future would be associated with a different state of affairs of the World and thus would be described by their manifests $M$ (Assumption 1), different than $\dot{M}$, and thus having no ontological properties. Why then is the observer not 'stuck' in a singular and static $\dot{M}$? How and why does time appear to flow from the 'past' to the 'future'?

We also recall that in the introduction we stated that something as 'trivial' as postulating that the present is caused by the past cannot be done in axiomatic science as it is an artificial argument; the past or the future must be derived as logically implied by the 'raw data' as a natural argument before it can be integrated into the framework.

Geometric thermodynamics offers the past and the future as natural arguments. Specifically, it suggests a purely entropic model of space-time in which the build-up of natural information determines the structure of space-time (its geometric substance). Let’s see how this transpires in the details. We have introduced a non-commutative extension to statistical physics granting it the ability to associate an entropy to arbitrary space-times. Generally speaking, statistical physics connects a set $Q$ of 'microstates' (a.k.a the microscopic description) to a set of functions on $Q$ (a.k.a. the bulk state) by the use of Lagrange multipliers, and under the principle of maximum entropy. In the present case, the 'microscopic' object of study is the space-time event, the bulk state is the space-time structure, and the Lagrange multiplier is the speed of light. A partition function of such events can be constructed by using generators of the geometric algebra. These generators ascribe geometric properties to the equation of state of such a system. The geodesic equations of motions are obtained by applying the principle of stationary action to the equation of state.

We state immediately that the geometric equation of state does not attribute an entropy to space (this is done by the Bekenstein-Hawking entropy and requires
a horizon), but to space-time. In space-time, like in space, the role of the entropy is to hide a region of it from the observer. In space-time, the hidden region defines the future of the observer. Specifically, the future is hidden from the present by entropy. It is as a result of this entropy that the observer can attribute a future to itself.

A change in the entropy of natural information is quantified by \( \tilde{k}^{-2}k_B^{-2}(dS)^2 \) which we have equated to the space-time interval \((ds)^2\) (in Theorem 1). We are already familiar with the replacement \((ds)^2 = c^2(d\tau)^2\), where \(\tau\) is the proper time. We will now call \(\tilde{k}^{-2}k_B^{-2}(dS)^2\) the entropic distance. Consequently, a change in proper time will always be quantified by a change in entropy. The entropic distance quantifies the informational departure of some hypothesized future or past manifest from the reference manifest. The near future, with a short entropic distance, is very similar to the present, and the far future, with a long entropic distance, is proportionally dissimilar. The arrow of time is upgraded to an arrow of space-time.

Let us begin by comparing the Euclidean case to the Minkowski case. In Euclidean space, the geometric equation of state is:

\[
\tilde{k}^{-2}k_B^{-2}(dS)^2 = (dx)^2 + (dy)^2 + (dz)^2 \quad (437)
\]

In Minkowski space-time, the geometric equation of state is:

\[
\tilde{k}^{-2}k_B^{-2}(dS)^2 = (dx)^2 + (dy)^2 + (dz)^2 - c^2(dt)^2 \quad (438)
\]

Finally, in an arbitrary space-time, the geometric equation of state is:

\[
\tilde{k}^{-2}k_B^{-2}(dS)^2 = \sum_{\mu,\nu} g_{\mu\nu} dX^\mu dX^\nu \quad (439)
\]

In the Euclidean case, we note that there is no nullsurface. All events (points) have a non-zero entropic distance from the origin. The case is degenerate. In the Minkowski space-time and the general space-time, however, we do note the presence of a nullsurface (light cone) on which all events have zero entropic distance from the origin. The presence of this nullsurface and the fact that it is a surface of zero entropic distance will allow us to define an extended present using entropy.

**Definition 22 (Present).** In space-times with a nullsurface, we can define an extended present for the observer as the set of all events of zero entropic distance:

\[
0 = \tilde{k}^{-2}k_B^{-2}(dS)^2 \quad (440)
\]

**Definition 23 (Past/Future).** We can further define a position in time away from the present and quantified by the entropy as:

\[
\text{Future} \quad ds = \tilde{k}^{-1}k_B^{-1}dS \quad (441)
\]

\[
\text{Past} \quad ds = -\tilde{k}^{-1}k_B^{-1}dS \quad (442)
\]
We note that if the geometric equation of state admits a nullsurface, then the entropy will be constant along the surface, and the entropic distance between all events on the surface will be zero. We thus conjecture that all events on the nullsurface are known/knowable to the observer because they have an entropic distance of zero from it and thus none of these events are hidden by entropy. The information available in the reference manifest is the information available on the nullsurface. Consequently, when we point a telescope towards the sky, the information we have access to is what’s on the nullsurface (i.e. the photons that hit the telescope have zero interval between their emission point at the telescope; the nullsurface here represents the sets of all such photons).

With entropic definitions for the past, the present and the future, it may become attractive to picture the world as an evolving block universe\[56\]. Here the relation $ds = \tilde{k}^{-1}k_B^{-1}dS$ quantifies the informational departure of a future event from the present. As stated, this entropy hides from an observer’s knowledge of future events. This entropy is eliminated if and when the gap between the observer and the entropically-distant event is reduced to 0 (that is when the observer moves forward in time until the event occurs). Thus, the future is hidden from the present by entropy. The past is attributed to the opposite description: Indeed, a reduction in entropy as measured from the present characterizes the past. The past is, therefore 'over-determined' by the present.

For any reference manifests, time appears to flow towards the future because geometric thermodynamics attributes a positive entropic distance to a region in space-time (the future) and a negative entropic distance to another (the past). The gradient of space-time entropy always peaks in the direction of the observer’s future in its frame of reference. Therefore, a space-time system seeking to statistically increase its entropy is powered to evolve forward in time by entropy. Indeed at thermodynamic equilibrium, the entropic frequency does become an entropic power as $k_B f = P/T$.

What is the role of the speed of light? In geometric thermodynamics, the speed of light is not fundamental, but instead emergent (in the same sense that thermodynamic pressure and temperature are emergent) from statistical physics. The speed of light, usually taken as an axiom in special relativity, is here defined as the ratio of the Lagrange multipliers of the partition function $c := f/\tilde{k}$. The speed of light here fulfills a role similar to the role fulfilled by the temperature in standard thermodynamics. Essentially, the speed of light is the "temperature" of space-time. The speed of light is a property emergent from the random selection of events from a larger set under the principle of maximum entropy. The speed of light is constant in an ensemble of space-time events at equilibrium for the same reason that the temperature is constant in a system at thermodynamic equilibrium. Indeed, when $f$ is the inverse of the Planck time, and $\tilde{k}$ the inverse of the Planck length, we get $c$, the speed of light:

\[
\frac{f}{\tilde{k}} = \left( \sqrt{\frac{c^3}{\hbar G}} \sqrt{\frac{\hbar G}{c^5}} \right)^{-1} = c
\]  

(443)
The speed of light appears constant because all observers on the nullsurface agree as to the results of their measurements (guaranteed because all events have zero entropic distance). Consequently, a theory of 'apparent communication' (information synchronicity along the nullsurface) replacing literal communications, follows from the speed of light being emergent. Indeed, communication (light) has an apparent speed because full measurement agreement between all observers is only achieved on the nullsurface and distances on this surface are measured proportionally to the speed of light.

Finally, a notion of dynamics used to interpolate the reference manifest to some future or past manifest via equations of motion yields the quantum mechanical equations of motions. This behavior is governed by a functional integral which, because of its metric form, admits both a time-like part (classical probabilities) and space-like part (quantum probabilities). Decoherence of the probability distribution occurs at the boundary of the nullsurface. The observer attributes its evolution in time to a sequence of manifests, each containing more information than the previous one, as events within its immediate space-like vicinity are consumed.

8.4 Mathematical description of the observer
One can think of an observer as a universal Turing machine undergoing a random walk in algorithmic space.

8.5 Geometric substance
The assumption of 'geometric substance' (Assumption 4), which assumes that all experiments have the observables of geometric thermodynamics (and are thus events), may appear to, by seemingly-arbitrarily picking 'geometry' as the extension to statistical physics, introduce a choice into the derivation. Consequently, this would, from this step forward, break the derivation so far consistent as a necessarily true argument. To remedy the situation, the question to answer then becomes: why, out of all possible mathematical concepts, should it be geometry that is associated with information, in reality?

We will attempt a partial answer:
Intuitively, we may assume that the existence of geometry implies the existence of information as one can use line, dots and curves to represent information (in the traditional manner known as writing). However, mathematically, information is not symbols on a sheet of paper, but rather it involves the random selection of elements from a set according to a probability distribution (Shannon entropy). Symbols written on paper contain no information if they are written according to a deterministic process (beyond the elegant representation of said deterministic process), and they do if they are the result of a random process (in which case the quantity of information goes to infinity if we let the random process run indefinitely). In this context, geometry by itself does not imply information because it has no concept of random selection, ...so, lets investigate the other way around; perhaps it is information that logically implies geometry?
Since the foundation of axiomatic science is in information (specifically, natural information: Definition 7), can we therefore show that Shannon entropy implies, say, the Pythagorean theorem?

It is possible but we must instead look at generalizations of the Shannon entropy, such as the density matrix approach of Von Neumann, or preferably the geometric density approach that we have introduced in this manuscript, then we can indeed prove geometric theorems 'in the bulk'. In this case, instead of assuming —a priori— the existence geometric observables, we would instead be assuming that the 'true probability rules of reality' are in fact said generalizations, and consequently, geometry is implied by said generalizations. This may be more palatable for certain inclinations. Consequently, 'reality' would be random in the microscopic detail, however, the special rules governing the addition/interference of probability amplitudes would produce a specific kind of interference pattern, a 'geometric' interference pattern, responsible for bulk geometric properties.

9 Conclusion

Is there a connection between 'consciousness' and quantum measurement? Yes, but it is almost comical. The mind does not 'collapse the wave-function', rather it imagines the alternative scenarios required to identity the equations which limits its participation in nature.

References


