Abstract

We present a mathematical model formalizing the practice of science in nature. The model, which we name *formal science*, is constructed within the frameworks of algorithmic information theory and that of theoretic computer science. Formal Science is a significant improvement over the informal practice of science as well as a 'clarification tool par excellence' for the foundation of physics, and as such it is able to derive the corpus of physics as a theorem. Formal science reveals that nature and the laws that govern it, far from being arbitrary, are in fact mathematically extremely special; they are, quite minimally, emergent as the 'substance' that formally verifies the experiments enumerated by the observer. After we present the model, we then begin the long program to derive all known physics using formal science as the sound, free of physical baggage, mathematical foundation of physics.
# Technical Introduction

4.1 Recap: Statistical Physics ........................................ 16
4.2 Recap: Algorithmic Statistical Physics ........................ 19
4.3 Applicability to Formal Science ................................. 22
4.4 Attempt 1: Algorithmic Thermodynamics ...................... 22
4.5 Hint 1: Seth Lloyd ................................................. 27
4.6 Hint 2: Entropy and Space-time ................................. 28
4.7 Attempt 2: The Search for a Suitable Ensemble ............... 29
4.8 Geometric Thermodynamics (as a sketch) ...................... 30
4.9 Recap: Geometric Algebra ....................................... 33
4.10 Recap: Quantum Thermodynamics .............................. 36

# Geometric Statistical Physics

5

# Results (Space-time)

6

6.1 Law of Inertia .................................................... 52
6.2 Arrow of Time .................................................... 54
6.3 Action ............................................................. 55
6.4 Fermi-Dirac statistics of events ............................... 56

# Results (Quantum mechanics)

7

7.1 Quantum Mechanics and Measurements ....................... 59
7.2 Decoherence at the Time-like/Space-like Boundary ........... 61
7.3 Non-relativistic limit ............................................ 64
7.4 Measurement ...................................................... 66

# Results (Cosmology)

8

8.1 Beckenstein-Hawking Entropy .................................. 67
8.2 de Sitter Space .................................................. 68
8.3 Cosmological Pressure .......................................... 70
8.4 Cosmological Inertia ............................................ 71
8.5 Entropic Derivation of $\Lambda$CDM ............................ 71

# Results (Standard model)

9

9.1 The Length of Poly-vectors .................................... 73
9.2 Potentials and taxi-cab terms .................................. 81
9.3 Area terms and electromagnetism .............................. 81

# Results (Sketches)

10

10.1 Normalization .................................................. 83
10.2 Fields (physics) .................................................. 83

# Discussion

11

11.1 Mathematical description of the observer .................... 84
11.2 An observer-centric model of reality ......................... 86
11.3 Interpretation of quantum mechanics ........................ 90
Part I
Science

1 Introduction

Formal science is a mathematical theory formalizing the practice of science in nature. Formal 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. Formal science introduces (and requires the use of) 'natural models', in which the laws of physics are derived as theorems from the description of nature, as distinct from an 'artificial model', in which the description of nature (the solutions) is obtained from postulated laws of physics. Formal science is at least as general as informal science and it introduces very strong constraints on a theory to be scientific in the formal mathematical sense.

Unlike a usual physical theory containing only a physics part, formal 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 this scope, formal 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?". Formal 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.

Formal science is constructed using the formalism of theorectic 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 formal 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 formal 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. Formal science is universal in the computer theoretic sense and, intuitively, in the 'physical/experimental' sense.

Formal science is philosophically extremely robust, and it exceeds the robust-
ness of informal science. It is because of both its robustness and its universality that formal 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 formal 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, and in their generality. Formal science reveals that nature and the laws that govern it, far from being arbitrary, are in fact mathematically extremely special; they are, quite minimally, emergent as the ‘substance’ that formally verifies the experiments enumerated by the observer. For these reasons and because it is a formalization of the practice of science, formal science is a candidate model to serve as the most fundamental description of nature possible — this is, in fact, its intended application.

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 formal 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 formal 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 formal 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 formal 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 essentially starts with raw data, and essentially compresses it into much shorter (ideally elegant) axioms. 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, understanding the raw data via short axioms 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 \( F = G \frac{m_1 m_2}{r^2} \) is a law of nature. We can compress an even bigger chunk of this data by adding a few more laws; such as aerodynamics laws, weather patterns, etc. Finally, armed with a set of initial conditions, and a set of deterministic laws, the initial conditions can be unpacked into the data that was initially used to justify the axiomatic re-organization (for completeness, we also note the quantum mechanical case, in which the unpacked data would come out as a superposition of solutions associated to a probability distribution which is approached by real world probabilities under repeated measurements over multiple copies of the same experimental preparation).

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, it is common to re-organize (flip) the presentation of the previously enumerated statements by using the postulated laws as the basis of the argument, as follows:

1. The theory of the metric expansion of the universe implies (predicts) the astronomical redshift.
2. The Big Bang theory implies (predicts) the CMB.
3. The theory of inflation implies (predicts) the homogeneity of the CMB temperature.
4. The theory of 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. The theory of gravity implies (predicts) that objects will fall from trees, should their attachment fail.
We note that the direction of the natural argument is flipped by the mathematisation of the artificial model. Essentially, these artificial models present their theorems as true statements implied by their axioms; consequently, there exists no proof within these models that they could eventually be falsified in the wild. However, as practitioners of science, aware of the justificatory origins of the model, we appropriately reduce the certainty suggested by this presentation; we are in fact quite aware that the theorems of these models are mere predictions, not necessarily true in the wild, and we do welcome and even expect the discovery of confirmatory or refuting evidence of these models. Consequently, if an artificial model is not consistent with all future raw data, then the model will eventually make incorrect predictions and will be falsified. Remarkably, since we are aware of this possibility but the model isn't, we are therefore operating using a model of reality closer to the truth than any artificial models that we use. Formal science corrects this "inequality".

Formal science, as a framework, connects 'raw data' (the axioms) to 'laws of physics' (the theorems) without requiring a preliminary axiomatic re-organization of the raw data. In formal science, unlike a typical physical theory, the direction of the natural argument is maintained:

\[
\begin{align*}
\text{Axioms} & & \text{Theorems} \\
\text{Formal science:} & data & \implies & model & (1) \\
\text{Typical physical theory:} & model & \implies & data & (2)
\end{align*}
\]

The central tenet of formal science is to construct a framework sufficient to derive physics, yet mathematically formulated in the natural direction. As an example, 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 an unsound implication: in all cases the model is simply falsified should it fail to account for the broken egg. Consequently, formal 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 concepts such as 'causality' can, in principle, be used as an artificial model for a subset of all observations, formal science shuns their introductions as postulates. Within the framework of formal 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 would thus be more fundamental within formal science to state that the past (if it exists — again, must be proven) 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. This is of course only a specific example amongst many, but paying special attention to these details will be paramount within the framework.
1.1 Hint 1: John A. Wheeler

We will now investigate two hints; the first by John A. Wheeler regarding the 'participatory-universe' hypothesis, then by Gregory Chaitin regarding the limits of mathematical formalism and how mathematics may deeply/subtly connect to science. Using these hints, we will eventually develop a method to describe nature universally (the raw data) without using physical baggage (forces, particles, etc.).

We summarize John A. Wheeler’s participatory universe hypothesis as follows. First, for any experiments, regardless of its simplicity or complexity, the registration of counts (in the form binary yes-or-no alternatives, the bit) is taken as a common book-keeping tool unifying the practice of science. Further to that, John A. Wheeler suggests (in the aphorism "it from bit" [1, 2]) that what we consider to be the "it" is simply one out of many possible mixture of theoretical glue that binds the "bits" together. Essentially, the 'bit' is real and the 'it' is derived. John A. Wheeler states;

"It from bit symbolizes the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a participatory universe"

Here, John A. Wheeler implies that the bit is the anchor to reality. The bit would come into being in the final act, so to speak, and then constraints the possible "it"s, whose theoretical formulation must of course be consistent with all bits generated (and not erased) thus far. Furthermore, he mentions that the bit is registered following an equipment-evoked response. To further illustrate his point of view, John A. Wheeler gives the photon as an example of the theme:

"With polarizer over the distant source and analyzer of polarization over the photodetector under watch, we ask the yes or no question, "Did the counter register a click during the specified second?" If yes, we often say, "A photon did it." We know perfectly well that the photon existed neither before the emission nor after the detection. However, we also have to recognize that any talk of the photon "existing" during the intermediate period is only a blown-up version of the raw fact, a count."

For John A. Wheeler, it makes little sense to speak of the photon existing (or not existing) until a detector registers a count. But he goes further and suggests that even after the registration of a count, deducing that the photon existed in between the counts is a "blown-up version of the raw fact, a count". Here, John A. Wheeler implies that the counts are what is real, not the theory that explains the counts. The theory is one hypothesis among many alternative and is, at best, a mathematical tool to make some sense of the counts, which by themselves define the world irrespectively of the theory.
In *Frontiers of time* (about a decade before 'it from bit'), John A. Wheeler lays out multiple attempts to derive some form of physical behavior/law from the study of experimentally-derived bits, but his approaches suffer from introducing physical baggage to get them started. Taking a specific example, on page 150, he reasons that time should emerge out of entropy. So far so good, but then, he argues that because the universe goes from Big Bang, to Big Stop, to Big Crunch, the statistics of entropy must be time symmetric. Therefore, he concludes that the acceptable rules of statistics to describe the dynamics of this entropy are those that he calls "double-ended statistics" which works in both directions of time (pages 150-155). The argument has of course an obvious fatal error; if time is derived from the bits, then so should the cosmos — why would one not be allowed to refer to time a-priori (it must be derived from entropy) but be allowed to refer to the cosmos' hypothetical future time-reversal to justify some properties on the bits. In fact, 39 years later, the results of [3], the Planck Collaboration, indicate a critical density consistent with flat topology and eternal expansion, possibly contradicting Wheeler's argument relying upon the necessity of some upcoming future cosmological reversal. Obviously, the eventual correct approach is only appealing if all physical statements (the 'its') follow from the bits. John A. Wheeler's book presents a myriad of similarly constructed arguments. John A. Wheeler does understand this to be a problem, and in his defense he does present "double-ended statistics" only as a example of what might be done. In fact, some 11 years later he corrects his approach (to what we refer to in this paper as his later definition of the participatory-universe hypothesis).

In *Information, physics, quantum: The Search For Links*, he provides general guidance on how to rectify this. It is there that he introduces the core idea that the bits are the result of the registering of equipment-evoked responses. With this, John A. Wheeler discards the idea of referring to the cosmos at all to enforce any kind of properties on the distribution of the bits, and instead refers to equipment evoked responses exclusively. After-all, evidence for both time and the cosmos are derived from the information provided to us by experimental devices (including the biological senses).

This completes our summary of the core concepts of John A. Wheeler’s participatory universe hypothesis. For the reader interested in more details, we would recommend reading [2] and [1].

So why this brief mention by John A. Wheeler of associating bits to an equipment-evoke response, essential — why can’t bits just stand alone? To understand this, we have to first recognize that the bits only have meaning if they are associated to some logical structure and that bits without it are meaningless. Let see why with the following example:

Let’s say that we were to provide someone with a list of bits:

\[111010110001001110101010101\]

How valuable would this person finds this information? Probably not much —why? As a hint, imagine if we were to tell this person that these bits represent the winning numbers of the next lottery draw. Then, all of sudden and although the sequence of bits stays the same, the bits are much more valuable.
Alternatively, we could have said that these bits are the results of random spin measurements. The bits once again stays the same, but their meaning is now completely different. Thus, some form of logical structure must be associated with any bits that we acquire about the World, otherwise they are without context or sense. This is why the pairing of experimental results (in the form of bits) and the experimental setup (under which the bits are acquired), are both equally crucial to the description.

But how do we describe the very complex world of experimental equipment without invoking physical baggage?

We intuit that this may have been a primary roadblock encountered by John A. Wheeler: formalizing equipment-evoked response seems to require a physical description of said equipment, and as this would contain physical baggage, then the fundamentality of the theory would be compromised. We risk running in circles; the 'it' describes the equipment which produces the 'bit' from which the 'it' is derived, which allows us to describe equipment... and so on.

The solution that we retained was to define an experiment not by the physical devices that are used in it, but instead by the protocol that must be followed to realize it. Specifically, a formal model of science attributes a scientific context to each bits of information that can be acquired about the World by associating it to a well-defined experiment (e.g. a replayable protocol, a series of steps others can follow, etc.). As we will see with the next hint, shifting the description from equipment to protocol is the key to make the endeavor mathematically precise (and without physical baggage).

1.2 Hint 2: Gregory Chaitin

Before we can formalize science within mathematics, we first need to identify a mathematical structure that 'behaves' the same as science does.

Gregory Chaitin summarizes his work on the halting probability[4], the Ω construction, in the book Meta Math![5]. Let U be the set of all universal Turing machines, then:

\[
\Omega : \mathbb{U} \rightarrow [0,1] \quad \text{UTM} \rightarrow \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \tag{4}
\]

The image of Ω are real numbers that are normal, incompressible and provably algorithmically random due to their connection to the halting problem in computer science. The reader may wish to read the first few paragraphs of our technical introduction (Section 4.2) on algorithmic information theory for a more detailed primer on Ω, and then come back to this section.

In the book Meta Math! Gregory Chaitin states that the following is the 'strongest' incompleteness theorem he produced in his career:

"A finitely axiomatic system (FAS) can only determine as many bits of Ω as its complexity."
As we showed in Chapter V, there is (another) constant $c$ such that a formal axiomatic system $\text{FAS}$ with program-size complexity $H[\text{FAS}]$ can never determine more than $H[\text{FAS}] + c$ bits of the value for $\Omega$.

where $H[p]$ is the Kolmogorov complexity of $p$.

This result essentially quantifies the general incompleteness in mathematics (originally identified/proved by Gödel for a specific case: the Gödel sentences in Peano’s axioms) and equates it to the Kolmogorov complexity, measured in quantities of bits, of the axiomatic basis of the finitely axiomatic system.

Gregory Chaitin dedicated a considerable amount of time to consider the implication of his $\Omega$ construction in regards to the philosophy of mathematics. What does such widespread incompleteness mean for mathematics? He concludes the following:

"I therefore believe that we cannot stick with a single finitely axiomatic system, as Hilbert wanted, we’ve got to keep adding new axioms, new rules of inference, or some other kind of new mathematical information to the foundations of our theory. And where can we get new stuff that cannot be deduced from what we already know? Well, I’m not sure, but I think that it may come from the same place that physicists get their new equations: based on inspiration, imagination and on — in the case of math, computer, not laboratory-experiments."

Finally, Gregory Chaitin further suggests:

"So this is a “quasi-empirical” view of how to do mathematics, which is a term coined by Lakatos in an article in Thomas Tymoczko’s interesting collection New Directions in the Philosophy of Mathematics. And this is closely connected with the idea of so-called “experimental mathematics”, which uses computational evidence rather than conventional proof to “establish” new truths. This research methodology, whose benefits are argued for in a two-volume work by Borwein, Bailey and Girgensohn, may not only sometimes be extremely convenient, as they argue, but in fact it may sometimes even be absolutely necessary in order for mathematics to be able to progress in spite of the incompleteness phenomenon..."

In another more recent article[6], Gregory Chaitin provides concrete examples of how the incompleteness phenomenon can enter some fields of mathematics. Specifically, he states:

"In theoretical computer science, there are cases where people behave like physicists; they use unproved hypotheses. $P \neq NP$ is one example; it is unproved but widely believed by people who study time complexity. Another example: in axiomatic set theory, the axiom of projective determinacy is now being added to the usual axioms.
And in theoretical mathematical cryptography, the use of unproved hypotheses is rife. Cryptosystems are of immense practical importance, but as far as I know it has never been possible to prove that a system is secure without employing unproved hypotheses. Proofs are based on unproved hypotheses that the community currently agrees on, but which could, theoretically, be refuted at any moment. These vary as a function of time, just as in physics.

Finally, we note Gregory Chaitin’s Meta-biological theory proposed in [7]; Proving Darwin: making biology mathematical, which references many of these concepts.

If Gregory Chaitin’s suggestion is correct, and that the incompleteness phenomenon induced on the foundations of mathematics by $\Omega$ may necessitate a scientific approach to said foundations, perhaps the appropriate insight for our purposes is in the reverse; we will, in fact, seek to use $\Omega$ and its properties to formalize the practice of science using mathematics.

We will now lay out the basis of the model.

1.3 Notation

The parentheses (example: $2(1 + 2) = 6$) are used to denote the order of operations. To avoid confusing ‘maps with inputs’ with ‘order of operations’ we will elect to use the square bracket to define valued maps. For instance a map $f : X \to \mathbb{R}$ will be written as $f[x]$ for $x \in X$. $S$ will denote the entropy, and $\mathcal{S}$ the action. Sets, unless a prior convention assigns it another symbol, will be written using the latex mathbb typography (ex: $\mathbb{L}, \mathbb{W}, \mathbb{Q}$, etc.). Matrices will have a hat (ex: $\hat{A}$), vectors will be in bold (ex: $\mathbf{a}, \mathbf{A}$) and most other constructions (ex: scalars) will have normal typography (ex. $a, A$). When important, matrices that are diagonal may be represented by the grave symbol, instead of the hat symbol, in order to better keep track of diagonalization (ex. $\grave{a}, \grave{A}$). Finally, the identity matrix is $\hat{1}$ and the null matrix is $\hat{0}$.

2 The Axioms of Formal Science

**Definition 1** (Language). A language $\mathbb{L}$, 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 $\emptyset$:

$$\mathbb{L} := \{\emptyset, s_1, s_2, \ldots\}$$  (5)

For instance, the sentences of the binary language are:

$$\mathbb{L}_0 := \{ \emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots \}$$  (6)
and its alphabet is:

$\Sigma_b = \{0, 1\}$

(7)

The fundamental object of study of formal 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 $\mathbb{L}$. The first sentence, $h$, is called the hypothesis. The second sentence, $TM$, is called the protocol. Let $UTM : \mathbb{L} \times \mathbb{L} \rightarrow \mathbb{L}$ be a universal Turing machine, then we say that the experiment holds if $UTM[TM, h]$ halts, and fails otherwise:

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

(8)

If $p$ holds, we say that the protocol verifies the hypothesis. Finally, $r$, also a sentence of $\mathbb{L}$, 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 $\mathbb{D}$ be the domain ($\text{Dom}$) of formal science. We can define $\mathbb{D}$ in reference to a universal Turing machine $UTM$ as:

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

(9)

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

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

$$M \subset \mathbb{D}$$

(10)

**Definition 5 (Set of all manifests).** Let $\mathcal{P}[\mathbb{A}]$ denote the power set of $\mathbb{A}$. Then the set of all manifests is $\mathcal{P}[\mathbb{D}]$. Of course, $M \in \mathcal{P}[\mathbb{D}]$. 

12
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, a manifest \( \hat{M} \) which corresponds to its state:

\[ \exists ! \hat{M} \tag{11} \]

- \( \hat{M} \) is called the ‘reference manifest’.
- The symbol \( M \) will denote any manifest in \( P[D] \), whereas \( \hat{M} \) specifically denotes the reference manifest corresponding to the present state of affairs.
- We consider the overhead ring symbol to be the designator of ontological existence and to be distinct from mathematical existence referenced by the symbol \( \exists \). For instance, in set theory, all manifests \( M \) exists (\( \exists \)), but in formal science only the state of affairs described by \( \hat{M} \) exists ontologically.

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

Formal science is interested in a specific kind of system; namely the experimentally-verified system:

Definition 6 (Experimentally-verified system). An experimentally verified system comprises a set of experiments \( M \) (a manifest), and a set of verification resources \( \mathcal{N} \) (nature / Definition 7) used to formally verify the experiments.

Definition 7 (Nature). Let \( \mathcal{N} \) be a set of constraints, which we call nature, of the form (see Equation 23 in the technical introduction to statistical physics) :

\[ \mathcal{N} := \{ \]
\[ \bar{O}_1 = \sum_{M \in P[D]} O_1[M] \rho[M], \quad (12) \]
\[ \bar{O}_2 = \sum_{M \in P[D]} O_2[M] \rho[M], \quad (13) \]
\[ \vdots \]
\[ \bar{O}_n = \sum_{M \in P[D]} O_n[M] \rho[M] \]
\[ \} \quad (14) \]

13
Intuitively, these constraints track the ’cost’ to experimentally-verify an experiment. In the case of computer experiments (Section 4.2), for instance, these resources would be computing resources and they may be expressed as a quantity of bits or of operations as required to bring to termination all programs that are elements of the reference manifest on a universal Turing machine. We say that nature $\mathcal{N}$ is ’spent’ to formally verify $\mathcal{M}$.

In an experimentally-verified system both the reference manifest and the resources required to formally verify said manifest exists ontologically:

**Axiom 2** (Existence of nature). *The resources used to verify the reference manifest exist ontologically on the same level as the reference manifest. We may use the overhead notation $\mathcal{N}$ to represent said resources.*

- If the verification resources, elements of $\mathcal{N}$, required to verify $\mathcal{M}$ would not exist on the same level as the reference manifest, it would be philosophically-problematic, even contradictory, to claim that the reference manifest is experimentally-verified, in reality.

- We therefore take it as axiomatic that for an experimentally-verified system, the resources used to perform said verification exists on the same level as the system.

- For instance, in the case of computer experiments, a reference manifest $\mathcal{M}$ which takes, say $10^{122}$ operations to be verified on a universal Turing machine, then implies that the system contains a resource element of $\mathcal{N}$ called ’operations’ whose quantity is $10^{122}$.

As infinitely many manifests $\mathcal{M}$ can be constructed from the elements of $\mathcal{D}$, one may wonder why it is the reference manifest $\mathcal{M}$ that is actual, and not any other. This brings us to the next assumption.

**Assumption 2** (The fundamental assumption of ’nature’). *We adopt the Bayesian principle of insufficient reason: The reference manifest is randomly selected from the set of all possible manifests $\mathcal{P}[\mathcal{D}]$ 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 may 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 $\mathcal{M}$ 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 $\mathcal{P}[\mathcal{D}]$. Using $\rho[\mathcal{M}]$, we will be able to quantify the amount of natural information in the message $\mathcal{M}$. 
Definition 8 (Natural Information). We define natural information as the information one gains by knowing which manifest is randomly selected from $\mathcal{P}[\mathcal{D}]$, according to the probability distribution $\rho[M]$ and under the verification resources constrained by $\mathcal{N}$. Let

$$\mathcal{P} := \left\{ \rho : \mathcal{P}[\mathcal{D}] \rightarrow [0,1] \middle| \sum_{M \in \mathcal{P}[\mathcal{D}]} \rho[M] = 1 \middle| \mathcal{N} \right\}$$

(15)

Then, the entropy of natural information is the functional:

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

(16)

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 may 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 nature are the laws of physics.

Formal 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 $\mathcal{P}[\mathcal{D}]$ as the set of possible messages.

We now solidify the intuitive notions of 'the world' used in Assumption 1 and Axiom 1:

Definition 9 (World). In formal science, we define 'The World' as an experimentally-verified system comprising the reference manifest $\mathcal{M}$, the reference nature $\mathcal{N}$ and spawning the domain of science $\mathcal{D}$. Thus:

$$\mathcal{W} = (\mathcal{D}, \mathcal{M}, \mathcal{N})$$

(17)

• Intentionally, the World is defined as the bare minimum required to maximize the entropy of natural information in nature.

• $\mathcal{M}$ is the state of affairs of $\mathcal{W}$.

• $\mathcal{N}$ is the reference nature spent to formally verify $\mathcal{M}$.
3 Thesis

In short, the thesis of this manuscript is to provide a compelling mathematical argument that; maximizing the entropy of natural information (Definition 8) using the world (Definition 9) as the system, and with the assumptions (Assumptions 1, 2, 3 and 4) implies the corpus\textsuperscript{1} of physics as a theorem.

Part II

Physics

4 Technical Introduction

To precisely quantify the relationship between natural information, entropy, verification resources and how this implies the laws of physics as the bulk description of an experimentally-verified system, we will eventually introduce geometric (or generalized/non-commutative) thermodynamics, but first, we will provide a recap of statistical physics, and then of algorithmic thermodynamics.

4.1 Recap: Statistical Physics

The applicability of statistical physics to a given physical system relies primarily upon two assumptions[8]. Here, we use the form stated by Prof. Victor S. Batista[9] in introductory notes:

1. "The experimental result of a measurement of an observable in a macroscopic system is the ensemble average of such observable."

2. "Any macroscopic system at equilibrium is described by the maximum entropy ensemble, subject to constraints that define the macroscopic system."

The first assumption is responsible for implying a number of fixed macroscopic quantities known as the statistical priors, or observables. Let $Q$ be a set of micro-states and $\mathcal{N}$ be a set of constraints, then set of all probability measures compatible with the constraints is:

$$\mathbb{P} := \left\{ \rho: Q \rightarrow [0,1] \left| \sum_{q \in Q} \rho[q] = 1 \right| \mathcal{N} \right\} \quad (18)$$

Consequently, the observables, in general, are $n$ functions defined as:

\textsuperscript{1}Here, the 'corpus of physics' is defined as an assortment of laws of physics and of physical concepts sufficient to convince a reasonable person that formal science is complete with respect to the laws of physics.
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>$1/\text{Kelvin}$</td>
<td>intensive</td>
</tr>
<tr>
<td>$E$</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/($\text{Kelvin-meter}^3$)</td>
<td>intensive</td>
</tr>
<tr>
<td>$\bar{V}$</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/($\text{Kelvin-kg}$)</td>
<td>intensive</td>
</tr>
<tr>
<td>$\bar{N}$</td>
<td>average number of particles</td>
<td>kg</td>
<td>macroscopic</td>
</tr>
</tbody>
</table>

\[ \mathcal{O}_i : \mathbb{P} \rightarrow \mathbb{R} \]
\[ \rho \mapsto \sum_{q \in \mathbb{Q}} \rho[q] O_i[q] \]  \hspace{1cm} (19)

where $O_i : \mathbb{Q} \rightarrow \mathbb{R}$. Typical thermodynamic quantities are shown in Table 1.

The second assumption is responsible for implying the probability measure which maximizes the entropy:

\[ S : \mathbb{P} \rightarrow [0, \infty[ \]
\[ \rho \mapsto -k_B \sum_{(q \in \mathbb{Q})} \rho[q] \ln \rho[q] \]  \hspace{1cm} (20)

under said constraints. This probability measure, why can be obtained from the method of the Lagrange multipliers, is the Gibbs ensemble:

\[ \rho : \mathbb{Q} \times \mathbb{R}^n \rightarrow [0, 1] \]
\[ (q, \alpha_1, \ldots, \alpha_n) \mapsto Z^{-1} \exp \left( -\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q] \right) \]  \hspace{1cm} (21)

where $\alpha_1, \ldots, \alpha_n$ are Lagrange multipliers. The partition function $Z$ is:

\[ Z : \mathbb{R}^n \rightarrow \mathbb{R} \]
\[ (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{(q \in \mathbb{Q})} \exp \left( -\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q] \right) \]  \hspace{1cm} (22)

The observables form a set $n$ of constraints that we call the thermodynamic bulk state:

\[ \mathcal{O}_i = Z^{-1} \sum_{(q \in \mathbb{Q})} O_i[q] \exp \left( -\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q] \right) \]  \hspace{1cm} (23)

The thermodynamic bulk quantities are also given by the following $n$ relation:
\[ \frac{\partial \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i} = \bar{O}_i \]  

(24)

And the variance by the following \( n \) relations:

\[ \frac{\partial^2 \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i^2} = (\Delta O_i)^2 \]  

(25)

The entropy for this ensemble is:

\[ S[\alpha_1, \ldots, \alpha_n] = k_B (\ln Z + \alpha_1 \bar{O}_1 + \cdots + \alpha_n \bar{O}_n) \]  

(26)

Taking the total derivative of the entropy, we obtain:

\[ dS[\alpha_1, \ldots, \alpha_n] = k_B (\alpha_1 d\bar{O}_1 + \cdots + \alpha_n d\bar{O}_n) \]  

(27)

which the equation of the state of the system.

Thermodynamics is derived from statistical physics. It is concerned primarily by the fundamental relation (27). Thermodynamic changes (and cycles) can be realized by changing the quantities \( \{\alpha_1, \ldots, \alpha_n\} \) and/or by modifications of \( \mathbb{Q} \). Under modification of \( \mathbb{Q} \), usually by cross product: \( \mathbb{Q} \times \mathbb{Q}_1 = \mathbb{Q}_2 \), or by set complement \( \mathbb{Q} \setminus \mathbb{Q}_3 = \mathbb{Q}_4 \), quantities which are invariant \( \{\alpha_1, \ldots, \alpha_n\} \) are called intensive, and quantities which are variant \( \{\overline{A}_1, \ldots, \overline{A}_n\} \) are called extensive.

As an example, replacing the generalized quantities by the typical thermodynamic quantities, in Table 1:

\[ \alpha_1 := \beta \]  

(28)

\[ \alpha_2 := \gamma \]  

(29)

\[ \alpha_3 := \delta \]  

(30)

\[ A_1[q] := E[q] \]  

(31)

\[ A_2[q] := V[q] \]  

(32)

\[ A_3[q] := N[q] \]  

(33)

the partition function would be:

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

(34)

The Gibbs measure is:

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

(35)
The observables are:

\[ E = \frac{1}{Z} \sum_{q \in Q} E[q] \exp(-\beta E[q] - \gamma V[q] - \delta N[q]) \] (36)

\[ V = \frac{1}{Z} \sum_{q \in Q} V[q] \exp(-\beta E[q] - \gamma V[q] - \delta N[q]) \] (37)

\[ N = \frac{1}{Z} \sum_{q \in Q} N[q] \exp(-\beta E[q] - \gamma V[q] - \delta N[q]) \] (38)

The entropy is:

\[ S[\beta, \gamma, \delta] = k_B (\ln Z + \beta E + \gamma V + \delta N) \] (39)

and the equation of state is:

\[ dS[\beta, \gamma, \delta] = k_B (\beta dE + \gamma dV + \delta dN) \] (40)

### 4.2 Recap: Algorithmic Statistical Physics

Many authors\cite{10, 4, 11, 12, 13, 14, 15, 16, 17} 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\cite{18, 19, 20, 16}. First let us introduce \( \Omega \). Let \( U \) be the set of all universal Turing machines, then:

\[ \Omega : U \rightarrow [0, 1] \]

\[ \text{UTM} \rightarrow \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \] (41)

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\cite{4} (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 \cite{21}). 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 quantity of bits of \( \Omega \). As such, Calude\cite{22} calculated the first 64 bits of \( \Omega[utm] \) for some universal Turing machine utm as:
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[16] suggested augmenting \( \Omega \) with a multiplication constant \( D \), which acts as an ‘algorithmic decompression’ term on \( \Omega \).

\[
\Omega[\text{UTM}] = 0.0000001000000100000110\ldots_2
\]

(42)

\[
\text{Chaitin construction} \quad \rightarrow \quad \text{Tadaki ensemble}
\]

\[
\Omega[\text{UTM}] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \quad \rightarrow \quad \Omega[\text{UTM}, D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|} \quad \text{(43)}
\]

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

\[
\text{Gibbs ensemble} \quad \rightarrow \quad \text{Tadaki ensemble}
\]

\[
Z[\beta] = \sum_{q \in Q} e^{-\beta E[q]} \quad \rightarrow \quad \Omega[\text{UTM}, D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|} \quad \text{(44)}
\]

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 \( |p| \):

\[
\overline{|p|} = \sum_{p \in \text{Dom}[\text{UTM}]} |p|2^{-D|p|} \quad \text{(45)}
\]

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

\[
S[\text{UTM}, D] = \ln \Omega + D\overline{|p|} \ln 2 \quad \text{(46)}
\]

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[20] 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:

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

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

(47)

Noting its similarities to the Gibbs ensemble of statistical physics (22), 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."

From equation (47), 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[18, 17].
4.3 Applicability to Formal 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. Formal 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 in order to realize Assumption 3. 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. This will be attempt 1 and yields the equation of state of a ‘classical’ computation.

4.4 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."

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.

---

2 https://link.springer.com/journal/11047
Table 2: Algorithmic quantities of the canonical ensemble of programs

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:

\[
Z : \mathbb{U} \times \mathbb{R}^2 \longrightarrow \mathbb{R} \\
(UTM, o_k, o_f) \longmapsto \sum_{p \in \text{Dom}[UTM]} 2^{-o_k O_x[p] - o_f O_t[p]} \tag{48}
\]

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 of UTM, or \( \mathcal{P}[\mathbb{D}] \). Executing a manifest \( M \in \mathcal{P}[\mathbb{D}] \) 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 \( O_n[M] \), the quantity of programs in the manifest. The quantities of this ensemble are shown in Table 3 and its partition function is:

\[
Z : \mathbb{R}^3 \longrightarrow \mathbb{R} \\
(o_k, o_f, o_\mu) \longmapsto \sum_{M \in \mathcal{P}[\mathbb{D}]} 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \tag{49}
\]

The corresponding probability measure is:

\[
\rho : \mathbb{W} \times \mathbb{R}^3 \longrightarrow \mathbb{R} \\
(M, o_k, o_f, o_\mu) \longmapsto Z^{-1} 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \tag{50}
\]

The probability measure maximizes the entropy subject to the following bulk constraints:
Symbol | Name | Units | Type
--- | --- | --- | ---
\(O_x[M]\) | length of programs in the manifest | [bit] | extensive
\(o_k\) | computing repetency | [1/bit] | intensive
\(\overline{O_x}\) | average tape usage | [bit] | macroscopic
\(O_t[M]\) | running time of programs in the manifest | [operation] | extensive
\(o_f\) | computing frequency | [1/operation] | intensive
\(\overline{O_t}\) | average clock time | [operation] | macroscopic
\(O_n[M]\) | quantity of programs in the manifest | [program] | extensive
\(o_\mu\) | computing overhead | [1/program] | intensive
\(\overline{O_n}\) | average concurrency | [program] | macroscopic

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

\[
\overline{O_x} = \sum_{M \in \mathcal{P}\{2\}} O_x[M]2^{-a_k O_x[M] - a_f O_t[M] - o_\mu O_n[M]} \tag{51}
\]

\[
\overline{O_t} = \sum_{M \in \mathcal{P}\{2\}} O_t[M]2^{-a_k O_x[M] - a_f O_t[M] - o_\mu O_n[M]} \tag{52}
\]

\[
\overline{O_n} = \sum_{M \in \mathcal{P}\{2\}} O_n[M]2^{-a_k O_x[M] - a_f O_t[M] - o_\mu O_n[M]} \tag{53}
\]

The Lagrange multipliers \((o_k, o_f, o_\mu)\) 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.

- The computing overhead \(o_\mu\) counts how many times the average concurrency \(\overline{O_n}\) must be doubled to double the entropy of the ensemble while holding the average clock time \(\overline{O_t}\) and the average tape usage \(\overline{O_x}\) 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:

\[
Z' : \mathbb{U} \times \mathbb{R}^2 \to \mathbb{R} \quad (UTM, o_k, o_f) \mapsto \sum_{p \in \text{Dom}[UTM]} 2^{-a_k O_x[p] - a_f O_t[p]} \tag{54}
\]
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 \( \overline{O}_f \) 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 \( \overline{O}_f \) 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.

2. **Size-cutoff formulation:**

\[
Z'' : U \times \mathbb{R}^2 \rightarrow \mathbb{R} \\
(UTM, o_k, x) \mapsto \sum_{p \in \{q : \text{Dom[UTM]}(q) \leq x \}} 2^{-o_k O_x[p]} \tag{55}
\]

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:**

\[
Z''' : U \times \mathbb{R}^3 \rightarrow \mathbb{R} \\
(UTM, o_k, o_f, t) \mapsto \sum_{p \in \{q : \text{Dom[UTM]}(q) \leq t \}} 2^{-o_k O_x[p] - o_f O_t[p]} \tag{56}
\]

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 cutoff formulation:**

\[
Z'''' : U \times \mathbb{R} \times \{G : f \rightarrow h\} \rightarrow \mathbb{R} \\
(UTM, o_k, g) \mapsto \sum_{p \in \{q : \text{Dom[UTM]}(q) \leq g \}} 2^{-o_k O_x[p]} \tag{57}
\]

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 coincidental 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'.

2. Entropy as a measure of computational distance

Consider an equation of state based on computing resources. A partition function of algorithmic thermodynamics (such as the one of Equation 50), has the following equation of state:

\[
\frac{dS}{d\Omega} = o_k d\Omega_x + o_f d\Omega_t + o_\mu d\Omega_n
\] (58)

Using this equation of state, we can quantify the computing 'distance' between two state of the system using the difference in entropy as the 'meter'. This equation forms a specific type of metric, known as a taxi-cab metric.

3. 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 \( Z(\beta) = \sum_{\eta \in Q} \exp(-\beta E[\eta]) \). Its average energy is given by \( \overline{E} = -\partial \ln Z/\partial \beta \) 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.

By considering that the group of reservoirs are the representation of an idealized 'supercomputer', the analogy is completed and algorithmic thermodynamics describes the dynamics of computation in equilibrium with the resources made available by the 'supercomputer'.

So far so good; but why is the computation classical, and not quantum? Where is quantum mechanics, the qubit, the geometry of space-time... etc.?

4.5 Hint 1: Seth Lloyd

In 2002, Lloyd[23] 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[24, 25] 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[26] (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:

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

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 \( \approx \frac{c^2l^2}{l_p} \approx \frac{l^2}{l_p} \)."

which is also \( \approx 10^{122} k_B \) [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.

The Seth Lloyd calculations are the result that we will use to understand bits and operations as quantities that are conserved in some suitable space at equilibrium, in the same sense that in ordinary statistical physics, the energy or the particle number are conserved quantities in time at equilibrium.

### 4.6 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\(^{27, 28, 29}\) 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\(^{24}\) 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:

\[
S = \frac{k_B c^3}{(4\pi G)} A.
\]

We mention Ted Jacobson\(^{30}\) 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\(^{31}\) 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\(^{32}\) 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 \((33)\), the Coulomb force \((34)\), Yang-Mills
gauge fields ([35]), and cosmology ([36, 37, 38]). Some criticism has, however, been voiced ([39, 40, 41, 42, 43], including by Visser [32].

Even more recently, a connection between entanglement entropy and general relativity has been supported by multiple publications [44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61].

Finally, we mention the body of work of George Ellis regarding the evolving block universe hypothesis detailed in [62, 63, 64] and the connection between space-time events, general relativity and quantum mechanics.

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

4.7 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 \(O_\times[p], O_t[p], \) 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 it is this specific ensemble, and not another, that happens to be the one that describes nature. 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:

\[
dE = T dS + p dV - \mu dN
\] (60)
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 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 5.

4.8 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:

$$\tilde{k}^{-1} dS = dX + dY + dZ \quad (61)$$

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:

$$\tilde{k}^{-1} dS = \sigma_1 dX + \sigma_2 dY + \sigma_3 dZ \quad (62)$$

We note that in this expression, the entropy becomes a vector, and so this will be addressed rigorously in section 5. 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:

\[
\tilde{k}^{-2}(dS)^2 = \sigma_1^2 (dX)^2 + \sigma_2^2 (dY)^2 + \sigma_3^2 (dZ)^2 \\
+ (\sigma_1 \sigma_2 + \sigma_2 \sigma_1) dX dY + (\sigma_1 \sigma_3 + \sigma_3 \sigma_1) dX dZ + (\sigma_2 \sigma_3 + \sigma_3 \sigma_2) dY dZ \\
(63)
\]

In the case were \(\sigma_1, \sigma_2\) 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:

\[
\sigma_1^2 = 1 \quad (64) \\
\sigma_2^2 = 1 \quad (65) \\
\sigma_3^2 = 1 \quad (66) \\
\sigma_1 \sigma_2 + \sigma_2 \sigma_1 = 0 \quad (67) \\
\sigma_1 \sigma_3 + \sigma_3 \sigma_1 = 0 \quad (68) \\
\sigma_2 \sigma_3 + \sigma_3 \sigma_2 = 0 \quad (69)
\]

Then, the cross-terms cancel and we obtain:

\[
\tilde{k}^{-2}(dS)^2 = (dX)^2 + (dY)^2 + (dZ)^2 \\
(70)
\]

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:

\[
\tilde{k}^{-1} dS = \tilde{k}^{-1} f_0 dT + \gamma_1 dX + \gamma_2 dY + \gamma_3 dZ \\
(71)
\]

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:

\[
\tilde{k}^{-2}(dS)^2 = \tilde{k}^{-2} f_0^2 (dT)^2 + \gamma_1^2 (dX)^2 + \gamma_2^2 (dY)^2 + \gamma_3^2 (dZ)^2 \\
+ \tilde{k}^{-1} f_0 (\gamma_0 \gamma_1 + \gamma_1 \gamma_0) dT dX + \tilde{k}^{-1} f_0 (\gamma_0 \gamma_2 + \gamma_2 \gamma_0) dT dY + \tilde{k}^{-1} f_0 (\gamma_0 \gamma_3 + \gamma_3 \gamma_0) dT dZ \\
+ (\gamma_1 \gamma_2 + \gamma_2 \gamma_1) dX dY + (\gamma_1 \gamma_3 + \gamma_3 \gamma_1) dX dZ \\
+ (\gamma_2 \gamma_3 + \gamma_3 \gamma_2) dY dZ \\
(72)
\]
The cross-terms cancel provided that the generators obey the following relations:

\[ \gamma_0^2 = 1 \]  
\[ \gamma_1^2 = -1 \]  
\[ \gamma_2^2 = -1 \]  
\[ \gamma_3^2 = -1 \]  
\[ \gamma_0 \gamma_1 + \gamma_1 \gamma_0 = 0 \]  
\[ \gamma_0 \gamma_2 + \gamma_2 \gamma_0 = 0 \]  
\[ \gamma_0 \gamma_3 + \gamma_3 \gamma_0 = 0 \]  
\[ \gamma_1 \gamma_2 + \gamma_2 \gamma_1 = 0 \]  
\[ \gamma_1 \gamma_3 + \gamma_3 \gamma_1 = 0 \]  
\[ \gamma_2 \gamma_3 + \gamma_3 \gamma_2 = 0 \]  

We also pose \( c := \tilde{k}^{-1} f \), then, the equation of state is:

\[ \tilde{k}^{-2} (dS)^2 = c^2 (dT)^2 - (dX)^2 - (dY)^2 - (dZ)^2 \]  

Here, the entropy becomes are 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:

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

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 \( E = \sum_{q \in Q} E[q] \rho[q] \), are now simply multiplied with a generator \( e_i \) of the geometric algebra, yielding \( n \) equations:

\[ e_i X_i = \sum_{q \in Q} e_i X_i[q] \rho[q] \]  

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 (27):

\[ dS = \tilde{k} e_1 dX_1 + \cdots + \tilde{k} e_n dX_n \]  

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 \( e_1, \ldots, e_n \) to be commutative:

\[ e_i^2 = 1 \]  
\[ e_i e_j = e_j e_i \]  

one obtains the relation

\[ dS = \tilde{k} dX_1 + \cdots + \tilde{k} dX_n, \]

which is of the same mathematical form as equation (27). Therefore, geometric thermodynamic 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. The robustness associated with statistical physics will thus be inherited by the laws of physics derived as a consequence of this generalization.

### 4.9 Recap: Geometric Algebra

A geometric algebra \( G \) is a ring equipped with algebraic generators that satisfy the generator relation:

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

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:

```

basis elements grade

\{1, e_0, e_1, e_2, e_3, e_0 e_1, e_0 e_2, e_0 e_3, e_1 e_2, e_1 e_3, e_2 e_3, e_0 e_1 e_2, e_0 e_1 e_3, e_0 e_2 e_3, e_1 e_2 e_3, e_0 e_1 e_2 e_3\}

grade-0 grade-1 grade-2 grade-3 grade-4

(91) (92) (93) (94) (95)
```

Poly-vectors of \( G \) 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>( \mathbf{v} := 1 )</td>
<td>0-vector, or scalar (96)</td>
</tr>
<tr>
<td>( \mathbf{v} := 3 \mathbf{e}_0 + 4 \mathbf{e}_1 )</td>
<td>1-vector, or vector (97)</td>
</tr>
<tr>
<td>( \mathbf{v} := 3 \mathbf{e}_0 \mathbf{e}_3 + 2 \mathbf{e}_2 \mathbf{e}_1 )</td>
<td>2-vector, or bivector (98)</td>
</tr>
<tr>
<td>( \mathbf{v} := 5 \mathbf{e}_0 \mathbf{e}_1 \mathbf{e}_2 )</td>
<td>3-vector, or trivector (99)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( \mathbf{v} := 2 \mathbf{e}_0 \mathbf{e}_1 \ldots \mathbf{e}_k )</td>
<td>k-vector (100)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( \mathbf{v} := 1 + 2 \mathbf{e}_0 + 5 \mathbf{e}_2 \mathbf{e}_2 )</td>
<td>poly-vector (101)</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:

\[
\mathbf{v} := r + r_0 \mathbf{e}_0 + r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_{01} \mathbf{e}_0 \mathbf{e}_1 + \ldots \quad \text{where} \quad r, r_0, r_1, r_2, r_{01}, \ldots \in \mathbb{R}
\]

is a real algebra \( \mathcal{G}(\mathbb{R}) \), and

\[
\mathbf{v} := z + z_0 \mathbf{e}_0 + \ldots \quad \text{where} \quad z, z_0, \cdots \in \mathbb{C}
\]

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 \( \{ \mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \} \) 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:

\[
\gamma_i \cdot \gamma_j = \frac{1}{2} (\gamma_i \gamma_j + \gamma_j \gamma_i) = \eta_{ij}
\]

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

\[
\eta_{ij} = \text{diag}(+, -, -, -)
\]

then,
\[\gamma_0 \gamma_0 = 1\]  \hspace{1cm} (106)
\[\gamma_1 \gamma_1 = -1\]  \hspace{1cm} (107)
\[\gamma_2 \gamma_2 = -1\]  \hspace{1cm} (108)
\[\gamma_3 \gamma_3 = -1\]  \hspace{1cm} (109)
\[\gamma_0 \gamma_1 + \gamma_1 \gamma_0 = 0\]  \hspace{1cm} (110)
\[\gamma_0 \gamma_2 + \gamma_2 \gamma_0 = 0\]  \hspace{1cm} (111)
\[\gamma_0 \gamma_3 + \gamma_3 \gamma_0 = 0\]  \hspace{1cm} (112)
\[\gamma_1 \gamma_2 + \gamma_2 \gamma_1 = 0\]  \hspace{1cm} (113)
\[\gamma_1 \gamma_3 + \gamma_3 \gamma_1 = 0\]  \hspace{1cm} (114)
\[\gamma_2 \gamma_3 + \gamma_3 \gamma_2 = 0\]  \hspace{1cm} (115)

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:

\[v u = v \cdot u + v \wedge u\]  \hspace{1cm} (116)

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:

\[v := a e_0 + b e_1\]  \hspace{1cm} (117)
\[u := c e_0 + d e_1\]  \hspace{1cm} (118)

Then, the geometric product is

\[v u = (ae_0 + be_1)(ce_0 + de_1)\]  \hspace{1cm} (119)
\[= ace_0 e_0 + ade_0 e_1 + be_1 e_0 + bee_1 e_1\]  \hspace{1cm} (120)
\[= ac + ade_0 e_1 - dbe_0 e_1 + bd\]  \hspace{1cm} (121)
\[= ac + bd + (ad - db)e_0 e_1\]  \hspace{1cm} (122)
\[= v \cdot u + v \wedge u\]  \hspace{1cm} (123)

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 $\sigma^{\mu\nu\rho\delta} = \frac{1}{24}[\gamma^\mu, \gamma^\nu, \gamma^\rho, \gamma^\delta]$

4.10 Recap: Quantum Thermodynamics

Since the generators of any finite geometric algebra $G$ 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:

$$\hat{H} = U \begin{pmatrix} E_1 & 0 & \cdots & 0 \\ 0 & E_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_n \end{pmatrix} U^\dagger \quad (124)$$

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:

$$|\psi\rangle = a_1 |\phi_1\rangle + \ldots + a_n |\phi_n\rangle, \quad \text{where } a_1, \ldots, a_n \in \mathbb{C} \quad (125)$$

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:

$$\rho_i := (a_i)(a_i)^* \quad (126)$$

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:

$$\hat{\rho} := |\psi\rangle \langle \psi| \quad (127)$$

or explicitly,

$$\hat{\rho} = \begin{pmatrix} (a_0)(a_0)^* & (a_0)(a_1)^* & \ldots & (a_0)(a_n)^* \\ (a_1)(a_0)^* & (a_1)(a_1)^* & \ldots & (a_1)(a_n)^* \\ \vdots & \vdots & \ddots & \vdots \\ (a_n)(a_0)^* & (a_n)(a_1)^* & \ldots & (a_n)(a_n)^* \end{pmatrix} \quad (128)$$
The Von Neumann entropy is:

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

(129)

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

\[
S = \text{Tr}\left[UdU^\dagger \ln[dU^\dagger]\right] = \text{Tr}\left[U^\dagger UdU^\dagger \ln[d]\right] = \text{Tr}[d \ln d]
\]  

(130) \hfill (131) \hfill (132)

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

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

(133)

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:

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

(134)

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

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

(135)
The difference in entropy between the pre-measurement pure state $\hat{\rho}$ and the post-measurement mixture $\hat{m}$ is equal to:

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

(136)

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

(137)

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

(138)

This is the same as the Shannon entropy obtained by equation 134.

Unitary transformations: One can change the state $|\psi\rangle$ by applying a unitary transformation $U$. $U$ 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:

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

(139)

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

$$|\psi'\rangle = U |\psi\rangle$$

(140)

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

(141)

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

(142)

$$= (\alpha a_1 + \beta a_2) |\phi_1\rangle + (-e^{i\varphi}\beta^* a_1 + e^{i\varphi}\alpha^* a_2) |\phi_2\rangle$$

(143)

The Shannon entropy of a measurement of $|\phi'\rangle$ along the eigenbasis is:

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

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

(144)

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

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

(145)

Post-measurement, the density matrix is:
\[ \hat{m} = \begin{pmatrix} (\alpha a_1 + \beta a_2)(\alpha a_1 + \beta a_2)^* & 0 \\ 0 & (-e^{i\varphi}\beta^* a_1 + e^{i\varphi} \alpha^* a_2)(-e^{i\varphi}\beta^* a_1 + e^{i\varphi} \alpha^* a_2)^* \end{pmatrix} \] (146)

The entropy of \( \hat{m} \) is equal to \( H \).

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

\[ \rho_i = \frac{1}{Z} e^{-\beta E_i} \] (147)

It then follows that:

\[ a_i = \sqrt{e^{i\varphi} \frac{1}{Z} e^{-\beta E_i}} = e^{i\frac{\varphi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_i}} \] (148)

where \( e^{i\frac{\varphi}{2}} \) is a complex phase, such that:

\[ (a_i)(a_i)^* = e^{i\frac{\varphi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_i}} e^{-i\frac{\varphi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_i}} \] (149)

\[ = \frac{1}{Z} e^{-\beta E_i} \] (150)

Thus, the thermal quantum state is written as:

\[ |\psi_{\text{thermal}}\rangle = \left( e^{i\frac{\varphi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_1}} |\phi_1\rangle + \cdots + e^{i\frac{\varphi}{2}} \sqrt{\frac{1}{Z} e^{-\beta E_n}} |\phi_n\rangle \right) \] (151)

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

\[ S = -k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) \ln \left[ \frac{1}{Z} e^{-\beta E_i} \right] \] (152)

\[ = -k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) (-\beta E_i - \ln Z) \] (153)

\[ = k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \beta E_i \right) (\beta E_i + \ln Z) \] (154)

\[ = k_B \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \beta E_i \right) + k_B \ln Z \sum_{i=1}^{n} \left( \frac{1}{Z} e^{-\beta E_i} \right) \] (155)

\[ = k_B \beta \sum_{i=1}^{n} \left( E_i \frac{1}{Z} e^{-\beta E_i} \right) + k_B \ln Z \] (156)
Posing $E := \sum_{i=1}^{n} E_i \exp(-\beta E_i)/Z$, then:

$$S = k_B (\ln Z + \beta \bar{E})$$  \hspace{1cm} (157)$$

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

$$Z = \text{Tr} \left[ e^{-\beta \hat{H}} \right]$$  \hspace{1cm} (158)$$

and the entropy as:

$$S = -\text{Tr}[\hat{\rho} \ln \hat{\rho}]$$  \hspace{1cm} (159)$$

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

$$\hat{\rho} = \frac{1}{Z} e^{-\beta \hat{H}}$$  \hspace{1cm} (160)$$

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

$$S = -\text{Tr} \left[ \left( \frac{1}{Z} e^{-\beta \hat{H}} \right) \ln \left( \frac{1}{Z} e^{-\beta \hat{H}} \right) \right]$$  \hspace{1cm} (161)$$

$$= -\text{Tr} \left[ \frac{1}{Z} e^{-\beta \hat{H}} (-\beta \hat{H} - \ln Z) \right]$$  \hspace{1cm} (162)$$

$$= \text{Tr} \left[ \frac{1}{Z} e^{-\beta \hat{H}} (\beta \hat{H} + \ln Z) \right]$$  \hspace{1cm} (163)$$

$$= \beta \text{Tr} \left[ \hat{H} \frac{1}{Z} e^{-\beta \hat{H}} \right] + \ln Z \frac{1}{Z} \text{Tr} \left[ e^{-\beta \hat{H}} \right]$$  \hspace{1cm} (164)$$

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

$$= \beta \langle \hat{H} \rangle + \ln Z$$  \hspace{1cm} (165)$$

It has the same mathematical form as the fundamental relation of thermodynamics (Equation 27).

5 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. To do that, 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 exceeds the speed of light. 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.

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[27]. Furthermore, from G. W. Gibbons and S. W. Hawking’s 1977 article[65], 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., ~ 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.

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 10 (Physical quantities). As we derive an ensemble of events, two physical quantities will be introduced as Lagrange multipliers. They are: 1) the entropic repetency \( k \) (the wavenumber) and 2) the entropic frequency \( f \).**

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". \( \dot{k} \) and \( \dot{f} \) 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).
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>$\bar{x}$</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>$\bar{t}$</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>

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 a Lagrange multiplier of the ensemble. By using the speed of light instead of the temperature as a Lagrange multiplier, the ensemble applies an entropy to all of space-time, with or without horizons, and thus determines its complete structure.

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

**Assumption 4** (The fundamental assumption of ’geometric substance’). *The ‘geometric substance’ is the media that formally verifies experiments in nature. Consequently, we will equip the ensemble of experiments with the observables of geometric thermodynamics (Definition 15). We will call the experiments of this new ensemble (Definition 16), geometric events $p$ (Definition 12).*

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 13).
- The geometric entropy $S$ (Definition 14).

42
• The geometric equation of state $dS$ (Theorem 1).
• The Gibbs-geometric measure and the partition function $Z$ (Theorem 2).

First, let us define a space-time event.

**Definition 11** (Space-time event). A space-time event is, in flat space-time, a 1-vector of $\text{Cl}_{1,3}(\mathbb{R})$:

\[ \mathbf{p} := \gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 \quad (166) \]

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:

\[ \mathbf{p} := e_0 X_0 + e_1 X_1 + e_2 X_2 + e_3 X_3 \quad (167) \]

Now, let us define a geometric event:

**Definition 12** (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 $\text{Cl}_{1,3}(\mathbb{R})$:

\[
\begin{align*}
\{1, & \\
&e_0, e_1, e_2, e_3, & \quad (168) \\
e_0 e_1, e_0 e_2, e_0 e_3, e_1 e_2, e_1 e_3, e_2 e_3, & \quad (169) \\
e_0 e_1 e_2, e_0 e_1 e_3, e_0 e_2 e_3, e_1 e_2 e_3, & \quad (170) \\
e_0 e_1 e_2 e_3 & \quad (171)
\end{align*}
\]

where

\[ g_{\mu\nu} = \frac{1}{2}(e_\mu e_\nu + e_\nu e_\mu) \quad (173) \]

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

\[ \mathbf{p} := G + \sum_{i}^{4 \text{ terms}} X_i e_i + \sum_{i,j}^{6 \text{ terms}} A^{ij} e_i e_j + \sum_{i,j,k}^{4 \text{ terms}} V^{ijk} e_i e_j e_k + U e_0 e_1 e_2 e_3 \equiv \sum_{i=1}^{\leq 2^n} E_i X_i \quad (174) \]

where

\[ E_i X_i \]

is a compact notation.

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. This is indicated by the notation 'less-than-or-equal' \( \leq \) \( 2^n \) in the sum of the compact notation, granting the user freedom to use less than all \( 2^n \) basis elements.

• A geometric event can be expressed as a \( 2^n \)-tuple; where \( n \) is the number of basis elements of the algebra:

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

(175)

**Definition 13** (Geometric density).

Let \( g[p] \) be a poly-vector valued function:

\[
g : \mathbb{D} \rightarrow \mathcal{G} \\
p \mapsto p
\]

(176)

Here, \( g \) takes an element of \( \mathbb{D} \) and maps it to a geometric event. Using \( g \), we can now define a geometric density \( \rho \) if the following sum converges to a finite value \( \Omega \in \mathcal{G} \):

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

(177)

If \( \Omega \) has an inverse, the normalization condition can be rewritten as:

\[
1 = \Omega^{-1} \sum_{p \in \mathbb{D}} g[p]
\]

(178)

in which case \( \Omega^{-1} \) is the normalization poly-vector.

Let us now define \( \rho \) as the geometric density:

\[
\rho[p] := \Omega^{-1} g[p]
\]

(179)

Then, the normalization condition becomes:

\[
1 = \sum_{p \in \mathbb{D}} \rho[p]
\]

(180)

Other normalization conditions are possible as special cases. For instance, we recall that a geometric algebra has a quadratic form defined as \( Q : \mathcal{G} \rightarrow F \), where \( F \) is a field (usually over the reals, or over the complex). In these cases, we can use a 'square-normalizable' condition:
\[ 1 = \sum_{p \in D} g[p](g[p])^* \] (181)

where \((g[p])^*\) is the complex conjugate of \(g[p]\). We note that in the case where the geometric algebra is \(Cl_{1,0}(\mathbb{R})\), the algebra of the complex numbers, then the normalization condition produces a Hilbert space for the functions \(g\) whose quadratic form is positive-definite. However, not all poly-vectors of \(G\) produces a unique real number by application of the quadratic form, and not all quadratic form a positive-definite, therefore this normalization condition does not hold for all \(g\). Consequently, as we are establishing the general geometric case here, we will proceed with our definition as it handles all geometric algebras.

**Definition 14 (Geometric entropy).** Let

\[ \mathbb{P} = \left\{ \rho : D \rightarrow G \left| \sum_{p \in D} \rho[p] = 1 \right. \right\} \] (182)

We define the geometric entropy as:

\[ S : \mathbb{P} \rightarrow G \rho \mapsto -k_B \sum_{p \in D} \rho[p] \ln \rho[p] \] (183)

**Definition 15 (Geometric bulk).** The constraints \(\mathcal{N}\) of the geometric ensemble are:

\[ \mathbf{E}_i \mathbf{X}_i = \mathbf{E}_i \sum_{p \in D} X_i[p] \rho[p] \] (184)

Here we note that we have simply taken the usual expression of a prior of statistical physics (Equation 23) and we have equipped it with a basis of \(G\): noted as \(\mathbf{E}_i\). For instance, the geometric algebra \(Cl_{1,3}(\mathbb{R})\) admits \(2^{1+3} = 16\) basis elements \(\{\mathbf{E}_1, \ldots, \mathbf{E}_{16}\}\). Therefore, an ensemble constructed from this algebra may admit up to \(2^n\) statistical priors. Finally, we note that the functions \(X_i[p]\) are projection maps \(X_i : D \rightarrow \mathbb{R}\) where \(X_i[p]\) returns the value of the \(i^{th}\) element of the \(2^n\)-tuple associated to \(p\). The terms \(\mathbf{X}_i\) are averages.

**Definition 16 (Geometric ensemble).** A geometric ensemble \(\mathcal{E}_G\) is a pseudo-probability space composed of the 3-tuple:

\[ \mathcal{E}_G := \{D, \mathcal{P}[D], \rho\} \] (185)

where \(D\) is the sample space, and \(\rho : D \rightarrow G\) is the geometric density.
Remarks:

• A case of special interest is the geometric ensemble in 3D space over the reals: $\mathcal{E}_{Cl_3(\mathbb{R})}$. In this case, the norm is positive-definite, consequently this ensemble can be constructed as a probability space, instead of a pseudo-probability space.

• Another case of special interest is the geometric ensemble in $3+1$ space-time over the reals: $\mathcal{E}_{Cl_{1,3}(\mathbb{R})}$. However, since its norm is not positive-definite, (and thus describes a pseudo-Riemannian manifold), consequently, the ensemble it describes is, likewise, a pseudo-probability space.

• Since the set of all experiments is countable (Definition 3: $\mathcal{D} = \text{Dom}[\text{UTM}]$), 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[66], as opposed to real analysis. For our purposes, however, the difference is merely a formality. Indeed, computable analysis is very similar, from a bird’s eye view, 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 $M$ (a manifest) selected over $\mathcal{P}[\mathcal{D}]$ (the set of all manifests), we create $n$ ensembles of $\rho$ (an experiment) selected over $\mathcal{D}$ (the domain of science). In this case, the ensemble $M \in \mathcal{P}[\mathcal{D}]$ is the grand canonical ensemble to $n$ canonical ensembles $\rho \in \mathcal{D}$. At any point, should we prefer to work with $M \in \mathcal{P}[\mathcal{D}]$, rather than with $n$ systems of $\rho \in \mathcal{D}$, we can redress to a grand-canonical ensemble simply by introducing $\mu N(M)$ as a thermodynamic observable and conjugate pair in the grand-canonical ensemble and summing $M \in \mathcal{P}[\mathcal{D}]$ instead of $\rho \in \mathcal{D}$. 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 $\rho \in \mathcal{D}$.

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

$$dS[\rho] = \sum_{i=1}^{<2^n} \lambda_i E_i \, d\rho_i[\rho]$$ (186)

**Proof.** We use the method of the Lagrange multipliers to find the maximum of the geometric entropy.

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

$$1 = \sum_{\rho \in \mathcal{D}} \rho[p]$$ (187)

$$E_i X_i = E_i \sum_{\rho \in \mathcal{D}} X_i[p][\rho]$$ (188)

2. The Lagrange equation to maximize is:

$$\mathcal{L}[\rho, \lambda, \lambda_1, \ldots, \lambda_{2^n}] = S[\rho] - \lambda \left( -1 + \sum_{\rho \in \mathcal{D}} \rho[p] \right) - \sum_{i=1}^{<2^n} \lambda_i \left( -E_i X_i + E_i \sum_{\rho \in \mathcal{D}} X_i[p][\rho] \right)$$ (189)

where $\lambda, \lambda_1, \ldots, \lambda_{2^n}$ are Lagrange multipliers elements of $\mathbb{R}$.

3. Maximizing $\mathcal{L}$ is done by taking its total derivative and posing it equal to 0:
\[ d\mathcal{L}[\rho, \lambda, \lambda_1, \ldots, \lambda_{2^n}] = 0 \]  

Consequently:

\[ 0 = d \left( S[\rho] + \lambda - \lambda \sum_{p \in \mathcal{D}} \rho[p] + \sum_{i=1}^{2^n} \lambda_i E_i X_i - \sum_{i=1}^{2^n} \lambda_i E_i \sum_{p \in \mathcal{D}} X_i[p] \rho[p] \right) \]

which we arrange as:

\[ dS[\rho] = d \left( -\lambda + \sum_{p \in \mathcal{D}} \rho[p] - \sum_{i=1}^{2^n} \lambda_i E_i X_i + \sum_{i=1}^{2^n} \lambda_i E_i \sum_{p \in \mathcal{D}} X_i[p] \rho[p] \right) \]

4. We note that \( X_i, X_i[p] \) and \( E_i \) are not variables of \( \mathcal{L}[\rho, \lambda, \lambda_1, \ldots, \lambda_{2^n}] \), consequently we obtain:

\[ dS[\rho] = -d\lambda + (d\lambda) \sum_{p \in \mathcal{D}} \rho[p] + \lambda d \sum_{p \in \mathcal{D}} \rho[p] \]

\[ - \sum_{i=1}^{2^n} E_i X_i d\lambda_i + \sum_{i=1}^{2^n} E_i \left( \sum_{p \in \mathcal{D}} X_i[p] \rho[p] \right) d\lambda_i + \sum_{i=1}^{2^n} \lambda_i E_i \sum_{p \in \mathcal{D}} X_i[p] \rho[p] \]

5. We then pose the following function definition:

\[ X_i[\rho] := \sum_{p \in \mathcal{D}} X_i[p] \rho[p] \]

We also replace the following expression by its constraint:

\[ \sum_{p \in \mathcal{D}} \rho[p] = 1 \]

Making the replacements, we obtain:

\[ dS[\rho] = -d\lambda + d\lambda + \lambda d(1) - \sum_{i=1}^{2^n} E_i X_i d\lambda_i + \sum_{i=1}^{2^n} E_i X_i[\rho] d\lambda_i + \sum_{i=1}^{2^n} \lambda_i E_i dX_i[\rho] \]
6. Since $X_i[\rho]$ is constrained to $X_i$, then some terms cancel, and we get:

$$dS[\rho] = \sum_{i=1}^{\leq 2^n} \lambda_i E_i \, dX_i[\rho]$$  \hspace{1cm} (197)

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

$$\gamma_0 \bar{X}_0 = \sum_{p \in \mathbb{D}} \gamma_0 X_0[p] \rho[p]$$  \hspace{1cm} (198)

$$\gamma_1 \bar{X}_1 = \sum_{p \in \mathbb{D}} \gamma_1 X_1[p] \rho[p]$$  \hspace{1cm} (199)

$$\gamma_2 \bar{X}_2 = \sum_{p \in \mathbb{D}} \gamma_2 X_2[p] \rho[p]$$  \hspace{1cm} (200)

$$\gamma_3 \bar{X}_3 = \sum_{p \in \mathbb{D}} \gamma_3 X_3[p] \rho[p]$$  \hspace{1cm} (201)

The geometric equation of state is:

$$dS = \lambda_0 \gamma_0 \, d\bar{X}_0 + \lambda_1 \gamma_1 \, d\bar{X}_1 + \lambda_2 \gamma_2 \, d\bar{X}_2 + \lambda_3 \gamma_3 \, d\bar{X}_3$$ \hspace{1cm} (202)

Its matrix representation is:

$$d\hat{S} = \begin{pmatrix}
\lambda_0 \, d\bar{X}_0 & 0 & \lambda_3 \, d\bar{X}_3 & \lambda_1 \, d\bar{X}_1 - i\lambda_2 \, d\bar{X}_2 \\
0 & \lambda_0 \, d\bar{X}_0 & \lambda_1 \, d\bar{X}_1 + i\lambda_2 \, d\bar{X}_2 & -\lambda_3 \, d\bar{X}_3 \\
-\lambda_3 \, d\bar{X}_3 & -\lambda_1 \, d\bar{X}_1 + i\lambda_2 \, d\bar{X}_2 & \lambda_0 \, d\bar{X}_0 & 0 \\
-\lambda_1 \, d\bar{X}_1 - i\lambda_2 \, d\bar{X}_2 & \lambda_3 \, d\bar{X}_3 & 0 & \lambda_0 \, d\bar{X}_0
\end{pmatrix}$$  \hspace{1cm} (203)

Finally, the diagonal is:

$$d\hat{s} = \sqrt{(\lambda_0 \, d\bar{X}_0)^2 - (\lambda_1 \, d\bar{X}_1)^2 - (\lambda_2 \, d\bar{X}_2)^2 - (\lambda_3 \, d\bar{X}_3)^2} \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}$$  \hspace{1cm} (204)

When possible, we will use the ± notation to group the unique eigenvalues of $d\hat{S}$ as follows:

$$dS = \pm \sqrt{(\lambda_0 \, d\bar{X}_0)^2 - (\lambda_1 \, d\bar{X}_1)^2 - (\lambda_2 \, d\bar{X}_2)^2 - (\lambda_3 \, d\bar{X}_3)^2}$$  \hspace{1cm} (205)

which we refer to as the poly-metric.
Definition 17 (Poly-metric). We define the poly-metric as a map, denoted with the double bar notation $\|\cdot\|$, from a poly-vector to the diagonalization of its matrix representation:

$$\|\cdot\| : G \rightarrow F^{n \times n}$$

where $F$ is a field (usually the reals or the complex) and $F^{n \times n}$ denotes a $n \times n$ matrix. To lighten the notation, whenever possible and appropriate, we will use to ± notation to group the unique eigenvalues of $\hat{u}$ as a single expression.

Here, we note that $\lambda_0$ has the units of $s^{-1}$ and $\lambda_1, \lambda_2, \lambda_3$ have the units of $m^{-1}$. We now pose, for the purposes of preserving isotropic principles:

$$\lambda := \frac{\lambda_0}{c} = \lambda_1 = \lambda_2 = \lambda_3$$

where $\lambda$ has the units of $m^{-1}$, and we obtain:

$$\lambda^{-2}(dS)^2 = (c dX_0)^2 - (dX_1)^2 - (dX_2)^2 - (dX_3)^2$$

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:

$$\lambda^{-2}(dS)^2 = g_{\mu\nu} d\bar{X}^\mu d\bar{X}^\nu$$

which we identify as the interval of general relativity (a sketch of the diagonalization process for this interval is offered further in this manuscript, at Equation 361).

Definition 18 (Entropic distance). We now note the following new equivalence between the geometric equation of state and the interval $s$ of relativity:

$$(ds)^2 \equiv \lambda^{-2}(dS)^2$$

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

We are already familiar with the replacement $ds = c d\tau$, where $\tau$ is the proper time read by a clock travelling along the path of an observer. Therefore, we can further pose this replacement:

$$(c d\tau)^2 = \lambda^{-2}(dS)^2$$

which associates a change of entropy to a measurable change of proper time.
We now introduce an explicit probability measure that reproduces the usual thermodynamics relations for the geometric case:

**Theorem 2** (Gibbs-Geometric measure). To simplify the notation let us define the following:

\[
\hat{X}[p] := \sum_{i=1}^{\leq 2^n} \hat{E}_i X_i[p]
\]  

(212)

Now, suppose the following probability measure:

\[
\dot{\rho}[p] = \frac{1}{Z} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right]
\]  

(213)

where:

\[
\dot{Z} = \sum_{p \in \mathcal{D}} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right]
\]  

(214)

where \(\dot{\rho}, \dot{Z}\) and \(\|\hat{X}[p]\|\) are diagonal matrices. Then, the total derivative of the entropy of \(\dot{\rho}\) is the diagonal geometric equation of state:

\[
\dot{S} = k_B \ln \dot{Z} + \lambda \|\hat{X}\|
\]  

(215)

**Proof.** 1. Injecting \(\dot{\rho}\) into the definition of the entropy, we get:

\[
\dot{S} = -k_B \sum_{p \in \mathcal{D}} \frac{1}{Z} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right] \ln \left[\frac{1}{Z} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right] \right]
\]  

(216)

\[
= -k_B \sum_{p \in \mathcal{D}} \frac{1}{Z} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right] \left( \ln \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right] - \ln \dot{Z} \right)
\]  

(217)

The term \(\ln \dot{Z} \sum_{p \in \mathcal{D}} \exp \left[-\lambda/k_B \|\hat{X}[p]\| \right]\) simplifies to \(\ln \dot{Z}\):

\[
= k_B \ln \dot{Z} + \sum_{p \in \mathcal{D}} \frac{1}{Z} \exp \left[-\frac{\lambda}{k_B} \|\hat{X}[p]\| \right] \lambda \|\hat{X}[p]\|
\]  

(218)

definition of the average
and we obtain an average for the remaining term:

\[ = k_B \ln \hat{Z} + \lambda \|\hat{X}\| \]  

(219)

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 \( \langle \hat{g}[q, \ln \hat{g}[q]) = \ln \hat{g}[q] d\hat{g}[q] + d\hat{g}[q] \). This equality holds only if \( \ln \hat{g}[q], \hat{g}[q] \) and \( d\hat{g}[q] \) commute, or if they are simultaneously diagonalizable.

**Definition 19** (Geometric thermodynamics). *In some range, the map

\[ S[\lambda_1, \ldots, \lambda_{2^n}] \to S[X_1, \ldots, X_{2^n}] \]  

(220)

is invertible[20]. 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:

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

(221)

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

6 Results (Space-time)

6.1 Law of Inertia

This first result will allow us to set a specific expression for the Lagrange multipliers which we will use throughout. As one may recall, in usual statistical physics, the Lagrange multiplier \( \beta \) is eventually associated to the temperature by connecting it to well-known thermodynamic equations having the same mathematical form as those derived from statistical physics. Specifically, the following equation (of statistical physics):

\[ \frac{\partial S}{\partial \beta} = \beta k_B \]  

(222)

is equivalent to the following equation (of thermodynamics):
\[ \frac{\partial S}{\partial E} = T^{-1} \]  \hspace{1cm} (223)

provided that we pose \( \beta := 1/(k_B T) \).

To find an expression for \( \lambda \), here a similar strategy is adopted. We will manipulate our equations until they are mathematically of the same form as some familiar laws of physics and then we use the equivalence to assign the correct expression to \( \lambda \) such that the two are equated. Let us begin:

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

\[ dS = \lambda (\sigma_1 dX_1 + \sigma_2 dX_2 + \sigma_3 dX_3) \]  \hspace{1cm} (224)

Its matrix representation is:

\[ d\hat{S} = \lambda \begin{pmatrix} dX_3 & dX_1 - i dX_2 \\ dX_1 + i dX_2 & -dX_3 \end{pmatrix} \]  \hspace{1cm} (225)

The diagonal matrix representation is:

\[ d\hat{S} = \lambda \sqrt{(dX_1)^2 + (dX_2)^2 + (dX_3)^2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \]  \hspace{1cm} (226)

Using the \( \pm \) notation, we group the eigenvalues as a poly-metric:

\[ \|dS\| = \pm \lambda \sqrt{(dX_1)^2 + (dX_2)^2 + (dX_3)^2} \]  \hspace{1cm} (227)

We now wish to investigate this equation as an entropic force emergent from the equation of state, connecting the entropy to the distance quantified by the poly-metric. We recall the definition of an entropic force:

\[ dS = \frac{F}{T} \, dx \]  \hspace{1cm} (228)

As a specific example; the tension within the chain of a polymer in a warm bath is an entropic force.

We are not very far from this definition of a general entropic force. In fact, the units of \( \lambda \) are the same as those of \( F/T \).

Which value of \( F \) and \( T \) to use? The natural choices, proposed by Erik Verlinde[67] are to take \( T \) as the Unruh temperature and \( F \) as the law of inertia. Let us solve for \( \lambda \) using:
\[ T_{\text{Unruh}} = \frac{\hbar a}{2\pi ck_B} \]  
\[ F = ma \]  

Then:

\[ \lambda = FT^{-1} \]  
\[ = ma \frac{2\pi ck_B}{\hbar a} \]  
\[ = 2\pi k_B \frac{mc}{\hbar} \]

We recognize the term \( mc/\hbar \) 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 20** (Geometric equation of state of inertia).

\[ dS = 2\pi k_B \frac{mc}{\hbar} \left( \sigma_1 d\bar{X}_1 + \sigma_2 d\bar{X}_2 + \sigma_3 d\bar{X}_3 \right) \]

**Definition 21** (Poly-metric equation of state of inertia).

\[ \|dS\| = \pm 2\pi k_B \frac{mc}{\hbar} \sqrt{(d\bar{X}_1)^2 + (d\bar{X}_2)^2 + (d\bar{X}_3)^2} \]

### 6.2 Arrow of Time

Let us now consider a geometric ensemble constrained by \( \gamma_0 \bar{X}_0, \gamma_1 \bar{X}_1, \gamma_2 \bar{X}_2, \gamma_3 \bar{X}_3 \). Its equation of state is:

\[ dS = \lambda (c\gamma_0 d\bar{X}_0 + \gamma_1 d\bar{X}_1 + \gamma_2 d\bar{X}_2 + \gamma_3 d\bar{X}_3) \]

Its poly-metric is:

\[ \|dS\| = \pm \lambda \sqrt{(c d\bar{X}_0)^2 - (d\bar{X}_1)^2 - (d\bar{X}_2)^2 - (d\bar{X}_3)^2} \]

We want to find the direction of the maximum rate of change in entropy. 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, a behavior similar to that of an entropic force is also obtained but instead of just in space and with a force, it is also in time and with a power. Indeed, the terms $\lambda c = P/T$ and $\lambda = F/T$ are simply, at a given equilibrium temperature, 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 the gradient of entropy of space-time points towards their future, for essentially the same reason that a polymer in a warm bath is entropically favored to stretch itself towards the direction of increased 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.

### 6.3 Action

The metric equation of state associated to the interval of general relativity is:

$$||dS|| = \pm 2\pi k_B \frac{m c}{\hbar} \sqrt{g_{\mu\nu} \, d\bar{X}^{\mu} d\bar{X}^{\nu}}$$

(238)

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

$$\int_a^b \left| \frac{\partial}{\partial \tau} S[\tau] \right| d\tau = \pm 2\pi k_B \frac{m c}{\hbar} \int_a^b \sqrt{g_{\mu\nu} \frac{\partial \bar{X}^\mu}{\partial \tau} \frac{\partial \bar{X}^\nu}{\partial \tau}} \, d\tau$$

(239)

We re-arrange as follows:

$$\frac{\hbar}{2\pi k_B} \int_a^b \left| \frac{\partial}{\partial \tau} S[\tau] \right| d\tau = \pm m c \int_a^b \sqrt{g_{\mu\nu} \frac{\partial \bar{X}^\mu}{\partial \tau} \frac{\partial \bar{X}^\nu}{\partial \tau}} \, d\tau$$

(240)

To recover the dynamics, one merely needs to investigate the change of entropy under an infinitesimal variation of $\delta$.

$$\frac{\hbar}{2\pi k_B} \int_a^b \delta \left| \frac{\partial}{\partial \tau} S[\tau] \right| d\tau = \pm m c \int_a^b \delta \sqrt{g_{\mu\nu} \frac{\partial \bar{X}^\mu}{\partial \tau} \frac{\partial \bar{X}^\nu}{\partial \tau}} \, d\tau$$

(241)
The path along \( \tau \) that extremalize the production of entropy is given in the stationary regime by posing:

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

(242)

and the corresponding equations of motion are:

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

(243)

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 a change of entropy as follows:

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

(244)

6.4 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:

\[
d\hat{S} = \lambda \sqrt{(c d\bar{X}_0)^2 - (d\bar{X}_1)^2 - (d\bar{X}_2)^2 - (d\bar{X}_3)^2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}
\]

(245)

For simplicity, we will consider the 1+1 space-time case and we use the \( \pm \) notation to group the eigenvalues as a single expression:

\[
\|dS\| = \pm \lambda \sqrt{(c d\bar{X}_0)^2 - (d\bar{X}_1)^2}
\]

(246)

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 \left[ \lambda \sqrt{(c d X_0)^2 - (d X_1)^2} \right] + 1} + 1 \quad (247)

\langle n \rangle_{\text{past}} = \frac{1}{\exp \left[ -\lambda \sqrt{(c d X_0)^2 - (d X_1)^2} \right] + 1} \quad (248)

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 \left[ -\lambda \sqrt{(c d X_0)^2 - (d X_1)^2} \right] + 1} & X_0 < 0 \\
\frac{1}{2} & X_0 = 0 \\
\frac{1}{\exp \left[ \lambda \sqrt{(c d X_0)^2 - (d X_1)^2} \right] + 1} & X_0 > 0
\end{cases} \quad (249)

\langle n \rangle is shown in Figure (1) using a 2-dimentional heat map. 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 of the light cone only by using event occupancy information. The usual description of a light cone is thus augmented, using occupancy statistics, with a different description for the past than for the future. 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). Interestingly, the occupancy probability of events nonetheless describes a probability space (the occupancy rate varies between 0\% to 100\% within the light-cone), even if we initially started with a pseudo-probability space for the geometric ensemble. 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 1a and 1b). In other words, the observer evolves forward in time by filling its past with events. To 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 1c). 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).

Of specific interest, we also note that events outside the light-cone of the observer (the white region in the figure) have a complex occupancy rate. This will become significant when we derive quantum mechanics in the next section.
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.
7 Results (Quantum mechanics)

7.1 Quantum Mechanics and Measurements

The probability of occupancy of an event is obtained by applying the Fermi-Dirac statistics to the ensemble (Section 6.4). By doing so, we recover real-valued occupancy rates but also complex-valued occupancy rates. Both the real probabilities and the complex occupancy rates play a role in the same ensemble and, notably, are dependant upon which region in space-time (with respect to the observer) the system is described.

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:

\[
\sqrt{c^2(\Delta t)^2 - (\Delta x)^2}\bigg|_{\Delta t=0,\Delta x=5} = \sqrt{c^2(0)^2 - 5^2} = i5
\]

We will make use of this imaginary number to recover a formulation of quantum mechanics using the path integral approach.

But first, let us investigate how the notion of dynamics can be introduced into formal science. We note that, in formal science, movement is not fundamental. Indeed, Assumption 1 assign a manifest 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 formal 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[2]. We have touched upon this in the introduction, but here, we will briefly state the relevant concepts as it relates to movement. 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.
Formal 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 poly-metric 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 formal 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 but we must restrict it to an interpolation tool 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.

To introduce the dynamics, the idea is to take two geometric events \( p_1 \) and \( p_2 \) then impose as a restriction that the interval associated to the possible paths between these events is given by the poly-metric. Unlike our previous probability distribution, the paths are no longer necessarily in a straight line and may now include paths following any curves (which is why we are now integrating over the path). Then, using \( \tau \), we parametrize the interval along a path \( q \) between the events, and we sum over the uncountable set of all paths \( P \) between them. The partition function previously a "sum over computable points" becomes its cousin of higher cardinality: a "functional integral over paths" (as we create an 'interpolation', we are free to release the restriction to the computable reals, and thus integrate over the uncountable set of paths):

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

Or more generally;

**Definition 22** (Geometric path integral),

\[
\tilde{Z}[P] = \int_{q \in P} \exp \left[ -\frac{\lambda}{k_B} \int_q \frac{\partial}{\partial \tau} \left( \sum_{i=1}^{<2^n} \hat{E}_i X_i[q[\tau]] \right) \, d\tau \right] \, Dq
\]
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.

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

(253)

When $O$ 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:

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

(254)

### 7.2 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 connecting two events gradually penetrate the light cone of the observer, 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 observer.

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:

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

(255)

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

Thermally prepared action: \( iS[q]/\hbar \)

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:

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

Thermal states: \(-\beta E[q]\)

where

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

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:

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

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 — Section 8.5). 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 un-accelerated 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 $O$ 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:

$$
\langle O \rangle = \frac{\int_{q \in P} O[q] \exp \left[ iS[q] \right] Dq}{\int_{q \in P} \exp \left[ iS[q] \right] Dq}
$$

(260)

In our case, we write:

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

(261)

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

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

(262)

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

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

(263)

where $\alpha A[q]$ is an arbitrary geometric thermodynamic quantity (or any number thereof) and where $Z := \int_{q \in P} \exp \left[ -\alpha \text{Re}[A[q]] + i \text{Im}[\alpha A[q]] \right] Dq$. 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 Feynman path integral:

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

(264)
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:

\[
\text{Im}[S[q]] = 0 \implies \overline{O} = \frac{1}{\mathcal{Z}} \int_{q \in \mathcal{P}} O[q] \exp \left[ -\beta \text{Re}[E[q]] \right] \text{D}q \tag{265}
\]

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

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

\[
P[q] = \frac{1}{\mathcal{Z}} \exp \left[ -\beta \text{Re}[E[q]] \right] \exp \left[ -i\beta \text{Im}[E[q]] \right] \tag{266}
\]

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

\[
P_{\text{time-like}}[q] = \frac{1}{\mathcal{Z}} \exp \left[ -\beta \text{Re}[E[q]] \right] \tag{267}
\]

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:

\[
\hat{\rho} = \begin{pmatrix}
P_{\text{time-like-1}} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & P_{\text{time-like-n}}
\end{pmatrix} \tag{268}
\]

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.

### 7.3 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:

\[
Z[P] = \int_{q \in P} \exp \left[ 2\pi mc \frac{\hbar}{i} \int q \left( \sqrt{ds^2 - dt^2} \right) \right] Dq
\]  
(269)

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

\[
= \int_{q \in P} \exp \left[ \pm 2\pi mc \frac{\hbar}{i} \int q \sqrt{dx^2 - dt^2} \right] Dq
\]  
(270)

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

\[
= \int_{q \in P} \exp \left[ \pm 2\pi mc \frac{\hbar}{i} \int q dt \sqrt{\dot{x}^2 - 1} \right] Dq
\]  
(271)

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

\[
= \int_{q \in P} \exp \left[ \pm 2\pi mc \frac{\hbar}{i} \int q dt \left( i - \frac{i\dot{x}^2}{2} + O[\dot{x}]^4 \right) \right] Dq
\]  
(272)

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

\[
= \int_{q \in P} \exp \left[ \pm 2\pi mc \frac{i}{\hbar} \int q dt \left( -\frac{\dot{x}^2}{2} + 1 \right) \right] Dq
\]  
(273)

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

\[
= \int_{q \in P} \exp \left[ \pm 2\pi mc \frac{i}{\hbar} \int q dt \left( -\frac{\dot{x}^2}{2} + V[q] \right) \right] Dq
\]  
(274)

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:
\[ Z[P] = \int_{q \in P} \exp \left[ \pm 2\pi \frac{mc}{\hbar} \int_q dt \sqrt{1 - \dot{x}^2} \right] Dq \quad (275) \]

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

\[ = \int_{q \in P} \exp \left[ \pm 2\pi \frac{mc}{\hbar} \int_q dt \left( 1 - \frac{\dot{x}^2}{2} + O[\dot{x}]^4 \right) \right] Dq \quad (276) \]

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

\[ = \int_{q \in P} \exp \left[ - \frac{2\pi c}{ha} \int_q dt \left( \frac{-\dot{x}^2}{2} + V[q] \right) \right] Dq \quad (277) \]

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

### 7.4 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:

\[ \hat{\rho} = \begin{pmatrix} \alpha \alpha^* & \alpha \beta^* \\ \alpha^* \beta & \beta \beta^* \end{pmatrix} \quad (278) \]

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

\[ \hat{\rho} = \begin{pmatrix} P_{\text{time-like-1}} & 0 \\ 0 & P_{\text{time-like-2}} \end{pmatrix} \quad (279) \]

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 $O$ "measured it" as they "interacted/came into causal contact" with each other. More precisely, the probability amplitudes became decoherent as the geometric path integral was continued into the time-like region).

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 lost by crossing the boundary (although since the information is thermal, it cannot be used to do work).

The final step; how the system comes to occupy a specific state randomly selected from the mixture, is discussed in section 11.3 regarding the interpretation of quantum mechanics.

8 Results (Cosmology)

8.1 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 $(m, X_0, X_1, X_2)$:

$$S[m, X_0, X_1, X_2] = 2\pi k_B \frac{mc}{\hbar} \left( \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 \right) + \ln Z$$ \hspace{1cm} (280)

Then the total derivative of $S$ is:

$$dS = 2\pi k_B \frac{c}{\hbar} \left( \sigma_1 (m dX_1 + X_1 dm) + \sigma_2 (m dX_2 + X_2 dm) + \sigma_3 (m dX_3 + X_3 dm) \right)$$ \hspace{1cm} (281)

Re-arranging, we get:

$$dS = 2\pi k_B \frac{c}{\hbar} \left( \sigma_1 (m dX_1 + X_1 dm) + \sigma_2 (m dX_2 + X_2 dm) + \sigma_3 (m dX_3 + X_3 dm) \right)$$ \hspace{1cm} (282)

The poly-metric equation of state is:

$$||dS|| = \pm 2\pi k_B \frac{c}{\hbar} \sqrt{ (m dX_1 + X_1 dm)^2 + (m dX_2 + X_2 dm)^2 + (m dX_3 + X_3 dm)^2}$$ \hspace{1cm} (283)
We now assign the Schwarzschild radius to the metric: $X_1 = \frac{2Gm}{c^2}, X_2 = 0, X_3 = 0$, and $dX_1 = 2Gc^{-2} dm, dX_2 = 0, dX_3 = 0$:

$$dS = \pm 2\pi k_B \frac{c}{\hbar} \left( m \frac{2G}{c^2} \, dm + \frac{2Gm}{c^2} \, dm \right)$$  \hspace{1cm} (284)

$$= \pm 2\pi k_B \frac{c}{\hbar} \frac{4Gm}{c^2} \, dm$$  \hspace{1cm} (285)

$$= \pm k_B 8\pi G \frac{m}{\hbar c} \, dm$$  \hspace{1cm} (286)

We now integrate:

$$\int dS = \pm k_B 8\pi G \frac{m}{\hbar c} \int m \, dm$$  \hspace{1cm} (287)

$$S = \pm k_B 4\pi G \frac{m^2}{\hbar c} + C$$  \hspace{1cm} (288)

For a black hole, the mass relates to the area as:

$$A = 4\pi r^2 = 4\pi \left( \frac{2Gm}{c^2} \right)^2 = 4\pi \frac{4G^2 m^2}{c^4} \iff m^2 = A \frac{c^4}{16\pi G^2}$$  \hspace{1cm} (289)

Replacing $m^2$ in our integral result $S$, we get:

$$S = \pm k_B 4\pi G \frac{A}{\hbar c} \frac{c^4}{G^216\pi} + C$$  \hspace{1cm} (290)

$$= \pm k_B \frac{c^3 A}{\hbar G^4} + C$$  \hspace{1cm} (291)

Finally, taking the sign to be positive and $C = 0$, we get:

$$S = k_B \frac{c^3 A}{\hbar G^4}$$  \hspace{1cm} (292)

which is the Bekenstein-Hawking entropy.

8.2 de Sitter Space

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

$$\alpha^2 = -c^2 X_0^2 + X_1^2 + X_2^2 + X_3^2$$  \hspace{1cm} (293)
with a cosmological horizon at \( r = \alpha \). One may also write \( \alpha \) as:

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

(294)

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

\[
S = k_B \ln Z + \lambda (c\gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3)
\]  

(295)

Then, we diagonalize the matrix representation:

\[
P(\hat{S} - k_B \ln \hat{Z})P^{-1} = \lambda \sqrt{-(cX_0)^2 + (X_1)^2 + (X_2)^2 + (X_3)^2}
\]

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\]

(296)

Using the \( \pm \) notation we group the unique eigenvalues into the poly-metric:

\[
\|S - k_B \ln Z\| = \pm \lambda \sqrt{-(cX_0)^2 + (X_1)^2 + (X_2)^2 + (X_3)^2}
\]  

(297)

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:

\[
\|S - k_B \ln Z\| = \pm \frac{2\pi k_B mc}{\hbar} \alpha
\]  

(298)

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 of de Sitter space is therefore quantified by the relation:

\[
\|S - k_B \ln Z\| = \pm 2\pi k_B \frac{mc}{\hbar} \alpha
\]  

(299)

Noting that in de Sitter \( \alpha \) is the Hubble radius at \( c/H \), we will derive, using entropy, the cosmological pressure and the cosmological law of inertia, then we will repeat the derivation but in a more general setting, with the addition of arbitrary thermodynamic quantities, and we will show that it is remarkably equivalent to \( \Lambda \)CDM.
8.3 Cosmological Pressure

We consider that the cosmological horizon bears an entropy for the same reason that the black hole apparent horizon bears an entropy (it delimits a boundary of information inaccessible to the observer). As we here consider de Sitter space, the cosmological horizon and the Hubble radius will be 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\[38, 25\]. With these replacements, the Unruh temperature becomes the cosmological horizon temperature\[38\]:

\[
T_{\text{de-Sitter-horizon}} := \frac{\hbar H}{2\pi k_B} \tag{300}
\]

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:

\[
T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{\hbar G^4} \, dA \tag{301}
\]

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:

\[
T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{\hbar G^4} 2r^{-1} \, dV \tag{302}
\]

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

\[
T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{\hbar G^4} \frac{H}{c} \, dV \tag{303}
\]

\[
= -\frac{H^2}{4\pi G} c^2 \, dV \tag{304}
\]

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\[38\]:

\[
T \, dS = -\frac{2}{3} \rho c^2 \, dV \tag{305}
\]

This result was obtained by Easson in \[38\], where it was suggested as a candidate explanation of the accelerated expansion of the universe.
8.4 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:

\[
T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{h G^4} (8\pi r \, dr) \quad (306)
\]

Then, with \( r \to c/H \), we get:

\[
T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{h G^4} 8\pi \frac{c}{H} \, dr
\]

\[
= -\frac{c^4}{G} \, dr \quad (307)
\]

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:

\[
T \, dS = -4Madr \quad (308)
\]

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 inertial mass of the cosmos is weighted at one fourth its energy content).

8.5 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_1 N_1, \ldots, \mu_n N_n \}\}. The equation becomes:

\[
||S - k_B \ln Z|| = \pm (2\pi k_B m c/\hbar) \alpha -\mu_1 N_1 - \cdots - \mu_n N_n \quad (310)
\]

\( ||S - k_B \ln Z|| \) denoting de Sitter entropy, \( \pm (2\pi k_B m c/\hbar) \alpha \) scalar thermodynamic quantities, and \(-\mu_1 N_1 - \cdots - \mu_n N_n \) deformed de Sitter entropy.
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, $\ln Z$ 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 $\{\mu_1 N_1, \ldots, \mu_n N_n\}$. 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 $\lim_{t \to \infty} \alpha \to (2\pi k_B mc/\hbar)^{-1} S$ and for all $i$, $\lim_{t \to \infty} N_i \to 0$.

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:

$$dE = (\rho c^2 + p) dV$$

(311)

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 $a \to c^2/r$. We use the relations derived by Easson[38], first for the radius:

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

(312)

where $a$ is a scaling factor. Then for the temperature:

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

(313)

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

$$T_{\text{cosmological-horizon}} dS = (\rho c^2 + p) dV$$

(314)

or, more specifically, as:

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

(315)

This is the sufficient starting point used by Easson[38] (in Annex A of his paper) to recover the Friedman equations of cosmology.
\[
H - \frac{k}{a^2} = -4\pi G \left( \rho + \frac{p}{c^2} \right) \\
H^2 = \frac{8\pi G}{3} \rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}
\]  
(316, 317)

as an equivalent representation of (315).

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.

It would thus appears that \( \Lambda CDM \) is to geometric statistical physics what the ideal gas law is to statistical physics.

9 Results (Standard model)

9.1 The Length of Poly-vectors

So far, we have used the poly-metric 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 poly-metric 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:

\[
\|v\|^2 := v \cdot v
\]

(318)

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

\[
v \cdot v = x^2 + y^2
\]

(319)

Investigating these definitions we found that, in some circles, 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 \( u := a + \sigma_x x + \sigma_y y + \sigma_x \sigma_y b \) (of the \( Cl_{0,3}(\mathbb{R}) \) algebra), then its inner product would be defined as:

\[
\|u\|^2 = a^2 + x^2 + y^2 + b^2
\]

(320)

The definition might be a valid mathematical norm, but it is physically incorrect. To help us understand what goes wrong with it, let us compare these kinds of mathematical norms to our definition (Poly-metric, Definition 17):
1. 0-vector: \( \mathbf{v} := a \) of the \( Cl_{0,3}(\mathbb{R}) \) algebra. Diagonalizing its matrix representation (in this case its already diagonal), we obtain:

\[
\| \mathbf{v} \| = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}
\]  

(321)

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

\[
\sqrt{\mathbf{v} \cdot \mathbf{v}} = a \quad \text{inner-product} \tag{322}
\]

\[
\| \mathbf{v} \| = a \quad \text{poly-metric} \tag{323}
\]

In this case, the eigenvalues of the poly-metric are the same as the inner-product.

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

\[
\hat{\mathbf{v}} = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} \implies \| \mathbf{v} \| = \begin{pmatrix} -\sqrt{x^2 + y^2 + z^2} & 0 \\ 0 & \sqrt{x^2 + y^2 + z^2} \end{pmatrix}
\]  

(324)

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

\[
\sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x^2 + y^2 + z^2} \quad \text{inner-product} \tag{325}
\]

\[
\| \mathbf{v} \| = \pm \sqrt{x^2 + y^2 + z^2} \quad \text{poly-metric} \tag{326}
\]

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

In this case, the absolute value of the eigenvalues of the poly-metric is the same as the inner-product, but one of our solutions has a minus sign. Let us keep this difference in mind, as we list more examples. Then we will explain the discrepancies.

3. 1-vector (#2): \( \mathbf{v} := \gamma_x x + \gamma_y y + \gamma_z z + \gamma_t t \) (of the \( Cl_{1,3}(\mathbb{R}) \) algebra). Diagonalizing its matrix representation:

\[
\hat{\mathbf{v}} = \begin{pmatrix} t & 0 & z & x - iy \\ 0 & t & x + iy & -z \\ -z & -x + iy & -t & 0 \\ -x - iy & z & 0 & -t \end{pmatrix}
\]  

\[
\implies \| \mathbf{v} \| = \sqrt{-x^2 - y^2 - z^2 + t^2} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]  

(328)
Its inner-product and poly-metric are, respectively:

\[ \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x^2 + y^2 + z^2 + t^2} \quad \text{"naive" inner-product} \quad (329) \]

\[ \| \mathbf{v} \| = \pm \sqrt{-x^2 - y^2 - z^2 + t^2} \quad \text{poly-metric} \quad (330) \]

where here we have used the ± notation to group the two unique 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 \( \mathbf{u} \cdot \mathbf{v} = \eta(\mathbf{u}, \mathbf{v}) \), then the length is:

\[ \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{-x^2 - y^2 - z^2 + t^2} \quad (331) \]

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

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

\[ \hat{\mathbf{u}} = \begin{pmatrix} z + a & x - iy \\ x + iy & -z + a \end{pmatrix} \quad (332) \]

we get:

\[ \| \mathbf{u} \| = \begin{pmatrix} a - \sqrt{x^2 + y^2 + z^2} & 0 \\ 0 & a + \sqrt{x^2 + y^2 + z^2} \end{pmatrix} \quad (333) \]

Let us now compare the poly-metric to the inner-product:

\[ \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{a^2 + x^2 + y^2} \quad \text{"naive" inner-product} \quad (334) \]

\[ \| \mathbf{u} \| = a \pm \sqrt{x^2 + y^2} \quad \text{poly-metric} \quad (335) \]

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 poly-metric by the element \( a \).

Let us investigate the contribution of \( a \) to the poly-metric 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-cab metric. Explicitly, the distance —perhaps 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 $Cl_{0,3}(\mathbb{R})$ algebra). Its poly-metric would be:

$$\|\hat{\mathbf{u}}\| = \frac{a + b + c}{\text{taxi-cab element}} \pm \frac{\sqrt{x^2 + y^2 + z^2}}{\text{euclidean element}}$$ (336)

In this example, the poly-metric 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.

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

$$\hat{z} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$ (337)

and the diagonal matrix is:

$$\|z\| = \begin{pmatrix} \alpha - i\beta & 0 \\ 0 & \alpha + i\beta \end{pmatrix}$$ (338)
The poly-metric is:

\[ \|z\| = \alpha \pm i\beta \]  
\[(339)\]

In comparison, in complex analysis, the norm of a complex number is usually taken to be \( |z| = \sqrt{\alpha^2 + \beta^2} \). How can we justify such a discrepancy between the established norm of a complex number and our definition?

Could it really have been the case all along that the norm of a complex number instead be a taxi-cab term paired to an imaginary Euclidean term? As we intend to show, the poly-geometry of the familiar complex numbers do not produce a Euclidean plane, but instead its own peculiar geometric object.

Consider the 2D case of a Euclidean plane. In this case, the properties of the algebraic basis are orthogonal:

\[ \sigma_x^2 = 1 \]  
\[(340)\]
\[ \sigma_y^2 = 1 \]  
\[(341)\]
\[ \sigma_x \sigma_y + \sigma_y \sigma_x = 0 \]  
\[(342)\]

whereas for the complex "plane", the relations are instead:

\[ 1^2 = 1 \]  
\[(343)\]
\[ i^2 = -1 \]  
\[(344)\]
\[ 1i + i1 = 2i \]  
\[(345)\]

For the complex "plane", these relations are much closer to the relation between a taxi-cab term and a Euclidean term, whose relations would be of the following type:

\[ 1^2 = 1 \]  
\[(346)\]
\[ \sigma_x^2 = 1 \]  
\[(347)\]
\[ 1\sigma_x + \sigma_x 1 = 2\sigma_x \]  
\[(348)\]

whose poly-metric (of \( u = a + x\sigma_x \)) is \( |u| = a \pm x \), but, as they are not identical (\( i^2 = -1 \) is not the same as \( \sigma_x^2 = 1 \)), the complex "plane" thus forms its own poly-geometric object.

To better understand the implication, let us compare two examples:
- Say someone asks your help to find a lost friend. If they say, "we walked 3km east, then 4km north, then I lost him", then clearly the last known location is 5km away.

- Now, say that someone knocks on your door with the following story: "My friend and I walked 3 km to the east in a straight line. Remarkably, there was a portal to an imaginary dimension at this point. We entered the portal and we walked 4 km inside of it also in a straight line, then I lost my friend. Can you help me find him?" Now before you go rushing to the portal, you suddenly remember your complex analysis class in which the norm on the complex "plane", for say $\alpha + i\beta$ is given by $\sqrt{\alpha^2 + \beta^2}$. Blindly believing this equation, you realize that you do not need to enter the portal at all. You can simply walk 5 km in a straight line in some suitable real direction to get to the person’s last known position, saving yourself a total of 2 km by skipping the portal.

Will you reach the destination faster than your friend who rushed to the portal? Well, if one believes the norm of a complex number to be $|z| = \sqrt{\alpha^2 + \beta^2}$ then one obtains the invalid conclusion that one can skip the portal to get there faster, and if one believes the norm to be a taxi-cab/imaginary-euclidean hybrid such as $||z|| = \alpha \pm i\beta$, then one concludes that one must enter the portal to get to the correct location.

Specifically, with the tax-cab/imaginary-euclidean hybrid norm, one will have to walk 3 km on the real line and 4 km on the imaginary line for a total of 7 km (not 5) to get to the correct position.

So why bother using the euclidean norm for a complex number? Essentially, in complex analysis one invents a new ‘universe’ where the complex numbers are re-written as $\mathbb{R} \times \mathbb{R}$ tuples, such as $(\alpha, \beta)$, and then are placed as points on said plane. Furthermore, on said plane, the imaginary term has been absorbed into one of the axis to make the points behave as if they are in euclidean space. This re-arrangement is mathematically valid, however, as it will soon become apparent once we apply these techniques to derive electromagnetism, physics requires that all geometric objects be expressed in the same space and compared to each other within the same ‘universe’, and by using the same techniques.

Remark: the presence of the $\pm$ sign is simply because one has freedom to define the distance on the imaginary line as either positive or negative with respect to the real line.

6. 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:

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

and its diagonalization is:

78
\[ \|q\| = \begin{pmatrix} a - i \sqrt{b^2 + c^2 + d^2} & 0 \\ 0 & a + i \sqrt{b^2 + c^2 + d^2} \end{pmatrix} \] \quad (350)

The poly-metric is:

\[ \|q\| = a \pm i \sqrt{b^2 + c^2 + d^2} \] \quad (351)

- With the quaternions, one can shortcut its way in 3-d space via the Pythagorean theorem, but only after one enters 'the portal'.

7. Consider a general poly-vector of \( Cl_{0,3}(\mathbb{R}) \):

\[ u := U + \sigma_x x + \sigma_y y + \sigma_z z + \sigma_x \sigma_y a_{xy} + \sigma_x \sigma_z a_{xz} + \sigma_y \sigma_z a_{yz} + \sigma_x \sigma_y \sigma_z V \] \quad (352)

The matrix representation of this vector is:

\[ \hat{u} = \begin{pmatrix} i a_{xy} + U + i V + z & i a_{yz} + a_{xz} + x - iy \\ i a_{yz} - a_{xz} + x + iy & -i a_{xy} + U + i V - z \end{pmatrix} \] \quad (353)

We diagonalize the matrix to find the eigenvalues:

\[ \|u\| = (U + iV) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sqrt{(x + i a_{yz})^2 + (y + i a_{xz})^2 + (z + i a_{xy})^2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \] \quad (354)

and we find the poly-metric:

\[ ||u|| = (U + iV) \pm \sqrt{(x + i a_{yz})^2 + (y + i a_{xz})^2 + (z + i a_{xy})^2} \] \quad (355)

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

Here, we describe a euclidean metric applicable to the real space, and another euclidean metric applicable to the imaginary space. In this poly-geometry, one can use the Pythagorean theorem to shortcut its way to the portal, then use the Pythagorean theorem again inside the portal to shortcut its way to the friend. This is the result one would expect from combining two separate spaces using the complex numbers.
8. (sketch) Consider a 1-vector \( \mathbf{u} \) of an arbitrary basis \( \{ \mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \} \) of \( Cl_{3,1}(\mathbb{R}) \). Let the basis be defined as:

\[
\begin{align*}
\mathbf{e}_0 &= t_0 \gamma_0 + x_0 \gamma_1 + y_0 \gamma_2 + z_0 \gamma_3 \\
\mathbf{e}_1 &= t_1 \gamma_0 + x_1 \gamma_1 + y_1 \gamma_2 + z_1 \gamma_3 \\
\mathbf{e}_2 &= t_2 \gamma_0 + x_2 \gamma_1 + y_2 \gamma_2 + z_2 \gamma_3 \\
\mathbf{e}_3 &= t_3 \gamma_0 + x_3 \gamma_1 + y_3 \gamma_2 + z_3 \gamma_3
\end{align*}
\]

(356) (357) (358) (359)

then

\[
\mathbf{u} = T \mathbf{e}_0 + X \mathbf{e}_1 + Y \mathbf{e}_2 + Z \mathbf{e}_3
\]

(360)

The matrix representation of \( \mathbf{u} \) is:

\[
\hat{\mathbf{u}} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
t_0 \gamma + t_1 \gamma_1 + t_2 \gamma_2 + t_3 \gamma_3 & 0 & 0 & 0 \\
T(-x_0 - y_0) + X(-x_1 - y_1) + Y(-x_2 - y_2) + Z(-x_3 - y_3) & 0 & 0 & 0 \\
T(x_0 + y_0) + X(x_1 + y_1) + Y(x_2 + y_2) + Z(x_3 + y_3) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

(361)

Finally, diagonalizing the matrix representation (steps omitted), we obtain the interval of general relativity.

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

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

(362)

The matrix representation of \( \mathbf{v} \) is:

\[
\hat{\mathbf{v}} = \begin{pmatrix}
G + t + iA_{12} - iV_{012} & A_{13} - iA_{23} + V_{013} - iV_{023} & -iU + z + A_{01} - iV_{023} & x - iy + A_{01} - iA_{02} \\
A_{12} + iV_{012} & G + t + iA_{12} + iV_{012} & -iU + z + A_{01} + iA_{02} & -iU - z + A_{03} - iV_{023} \\
-x + iy + A_{01} + iA_{02} & -x + iy + A_{01} - iA_{02} & G - t - iA_{12} + iV_{012} & A_{13} - iA_{23} - V_{013} + iV_{023} \\
x + iy + A_{01} - iA_{02} & x + iy + A_{01} + iA_{02} & -A_{03} + A_{04} + iA_{02} & G - t + iA_{12} - iV_{012}
\end{pmatrix}
\]

(363)

Diagonalizing this matrix, if it is possible, is left as an exercise.

Let us now use these results to investigate physical systems.
9.2 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:

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

(364)

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 poly-metric 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:

$$dS = \lambda(e_0 dX_0 + e_1 dX_1 + e_2 dX_2 + e_3 dX_3) + \lambda_a dU_a + \lambda_b dU_b + \lambda_c dU_c$$

(365)

The poly-metric is:

$$dS = \pm \lambda \sqrt{g_{\mu\nu} dX^\mu dX^\nu} + \lambda_a dU_a + \lambda_b dU_b + \lambda_c dU_c$$

(366)

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

$$\int_a^b \frac{dS}{d\tau} d\tau = \pm \int_a^b \left( \lambda \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} + \lambda_a \frac{\partial U_a}{\partial \tau} + \lambda_b \frac{\partial U_b}{\partial \tau} + \lambda_c \frac{\partial U_c}{\partial \tau} \right) d\tau$$

(367)

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 can then be incorporated into a path integral between two geometric events described by both euclidean and taxi-cab terms.

9.3 Area terms and electromagnetism

The bi-vector basis elements of $Cl_{1,3}(\mathbb{R})$ are:

$$\gamma_0\gamma_1, \gamma_0\gamma_2, \gamma_0\gamma_3, \gamma_1\gamma_2, \gamma_1\gamma_3, \gamma_2\gamma_3$$

(368)

One can construct an ensemble using all 2-basis generators. The equation of state would be:
\[ \text{d}S = \lambda_{E_x} \gamma_0 \gamma_1 \text{d}E_x + \lambda_{E_x} \gamma_0 \gamma_2 \text{d}E_y + \lambda_{E_x} \gamma_0 \gamma_3 \text{d}E_z + \lambda_{B_x} \gamma_1 \gamma_3 \text{d}B_x + \lambda_{B_y} \gamma_1 \gamma_2 \text{d}B_y + \lambda_{B_z} \gamma_1 \gamma_2 \text{d}B_z \]

(369)

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

\[
\text{d} \hat{S} = \begin{pmatrix}
-\i B_z \lambda_B & (\i B_x - B_y) \lambda_B & E_x \lambda_E & (E_x - \i E_y) \lambda_E \\
(\i B_x + B_y) \lambda_B & i B_z \lambda_B & (E_x + \i E_y) \lambda_E & -E_x \lambda_E \\
E_x \lambda_E & (E_x - \i E_y) \lambda_E & -i B_z \lambda_B & (-i B_x - B_y) \lambda_B \\
(E_x + \i E_y) \lambda_E & -E_x \lambda_E & (\i B_x + B_y) \lambda_B & i B_z \lambda_B
\end{pmatrix}
\]

(370)

Using diagonalization, the equation of state becomes the poly-metric of electromagnetism:

**Definition 23** (Poly-metric of electromagnetism).

\[ (\text{d}S)^2 = \lambda_{E_x}^2 || \text{d}E ||^2 - \lambda_{B_x}^2 || \text{d}B ||^2 \pm 2 \i \lambda_{E_x} \lambda_{B_x} (\text{d}E \cdot \text{d}B) \]

(371)

We have used the \(\pm\) sign to group the unique eigenvalues.

We can rewrite the metric as:

\[ (\text{d}S)^2 = \lambda_{E_x}^2 || \text{d}E ||^2 - \lambda_{B_x}^2 || \text{d}B ||^2 \pm 2 \i \lambda_{E_x} \lambda_{B_x} (\text{d}E \cdot \text{d}B) \]

(372)

These are simply the Lorentz invariants of electromagnetism.

We recall the familiar equality between the tensor representation and the invariant representation, as follows:

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

(373)

\[ \frac{1}{4} \epsilon^{abcd} F_{ab} F_{cd} = 2 \frac{1}{c} \text{B} \cdot \text{E} \]

(374)

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

Let the geometric events \(p_1\) and \(p_2\) be bi-vectors. As we have previously done in Section 7.1, we can here construct a path integral as an interpolation of the dynamics between \(p_1\) and \(p_2\). Remarkably, since the distance of the paths connecting the two events is the 2-vector poly-metric and since said metric is
the invariants of electromagnetism, then the construction of a path integral over this metric will eventually produce a quantized version of electromagnetism. The photon, as intuited by John A. Wheeler[2], is thus emergent as the object which connects the clicks of $p_1$ to $p_2$ in space-time via 'interpolation', for events expressed using the 2-basis elements of the geometric algebra.

10 Results (Sketches)

10.1 Normalization

A geometric 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 (the pseudo-probability associated with infinite energy states is exactly 0% according to the Gibbs-geometric measure). Consequently, the high energy spectrum would be dominated by a specific temperature (Unruh/Hawking/Hubble), which becomes the normalization condition for the energy levels of the system.

10.2 Fields (physics)

We have expressed our equations using paths and points. Consequently, our expression described the movement of a test particles within the space defined by the poly-metric. However, nothing prevents us from extending these definition to include fields. In this case, the geometric path integral becomes a geometric field integral, and instead of describing the dynamics of a test particle within a background poly-metric, we now describe the dynamics of the field itself within the poly-geometry:

**Definition 24** (Geometric field integral).

$$ Z[\Phi] = \int_{\phi \in \Phi} \exp \left[ -\frac{\lambda}{k_B} \int_{\mu} \frac{\partial}{\partial \mu} \|X[\phi]\| \, d\mu \right] D\phi $$

where $\phi$ are 4-tuples, and where $\mu$ is a 4-measure, and where $\Phi$ is the set of all field configurations.

In the case of a field in 3+1 space-time, one can define the notion of distance $\|X[\phi]\|$ with respect to the metric tensor as follows: $\sqrt{-|g|}$, where $|g|$ is the determinant of $g_{\mu\nu}$. Then, the geometric field integral has the same mathematical form as the Einstein-Hilbert action.

11 Discussion

We will now apply formal science to unresolved problems within the foundation of physics. As stated in the introduction, unlike artificial models containing only a physics part, formal science contains both a science part and a physics part, and
is therefore inherently more fundamental. Formal 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 formal science. This discussion will culminate in a novel interpretation of
quantum mechanics, necessitated by formal science which, we believe, resolves
the quantum measurement problem.

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

11.1 Mathematical description of the observer

A hyperwebster is a construction by Ian Stewart of an infinite dictionary contain-
ing all finite sentences of a given language. The dictionary is split into chapters;
one for each symbol. For instance, in the binary language, the chapter 0 contains
all sentences starting with 0, sorted in shortlex, and the chapter 1 contains all
sentences starting with 1, also sorted in shortlex:

\[
\begin{align*}
\text{Chapter 0:} & = (0, 00, 000, 0000, 00000, \ldots) \\
\text{Chapter 1:} & = (1, 10, 110, 1100, 11000, \ldots)
\end{align*}
\]

Then, by a finitely describable transformation, part of the (infinite) data
can be erased whilst still preserving the complete content of the dictionary.
Specifically, if one erases chapter 1 and the first symbol of the sentences of
chapter 0, one gets:

\[
\begin{align*}
\text{Transformed Chapter 0:} & = (\emptyset, \emptyset0, \emptyset01, \emptyset000, \emptyset0000, \ldots) \\
& \approx \{(0, 00, 000, 0000, 00000, \ldots), (1, 10, 110, 1100, 11000, \ldots)\}
\end{align*}
\]

The transformed chapter 0 contains the words of both chapters 0 and 1, and
thus it contains the same words as those of the full dictionary (even thought
more than a full chapter was erased!).

Let us now apply this terminology to formal science. We will call the programs
that halt for UTM the \textit{words} of the \textit{dictionary of science}. The dictionary of
science is equal to the domain of science:

\[
\text{Dictionary-of-science} := \mathbb{D} = \text{Dom}[\text{UTM}]
\]
Using this terminology we will organize the words into equivalent and complete chapters, and we will attempt to find a 'best' chapter to use as a model of reality.

We note that we will work using an ideal case; for instance, the life expectancy of a modern human being is roughly 75 years, and therefore one can only do so much in that time, but in the following discussion we will consider an idealized observer $O$ able to, in principle, recursively enumerate the domain of science on a universal Turing machine. The domain of $O$ is consequently the domain of science:

$$\text{Dom}[O] = \text{Dom}[\text{UTM}] = D$$  \hspace{1cm} (381)

In the case of a universe which embeds observers able to practice science, the 'dictionary of science' (i.e. the domain of science) used to describe the World, can be partially enumerated by the observer. A partial enumeration of the domain of science by an observer using a universal Turing machine (say, by running $n$ programs in dovetail according to their shortlex order), could be represented as follows:

$$\{p_1, p_2, \ldots, p_n\} \subset \text{Dom}[O]$$  \hspace{1cm} (382)

This partial enumeration, if it happened in reality, must be part of the reference manifest (lest it didn’t happen). Thus:

$$\{p_1, p_2, \ldots, p_n\} \subset \tilde{M}$$  \hspace{1cm} (383)

Since the inclusion of $\{p_1, p_2, \ldots, p_n\}$ into $\tilde{M}$ holds independently of the content of the partial enumeration it must therefore hold for all partial enumerations of the domain of science by $O$, including the limiting case. Therefore, all manifests (including the reference manifest) are subsets of $\text{Dom}[O]$:

$$\{p_1, p_2, \ldots, p_n\} \subset \tilde{M} \subset \text{Dom}[O]$$  \hspace{1cm} (384)

Remarkably, in the limiting case (let us use the symbol $\tilde{M}_{\infty}$ for the limiting case), the observer with access to a universal Turing machine will eventually recursively enumerate the domain of science in its entirety. The set $\tilde{M}_{\infty}$ is, like $\Omega$, also not computable, but nonetheless it is mathematically definable. Consequently, $\{p_1, p_2, \ldots\} = \text{Dom}[O]$, and therefore:

$$\text{Dom}[O] \subset \tilde{M}_{\infty} \subset \text{Dom}[O] \implies \tilde{M}_{\infty} = \text{Dom}[O] = \text{Dom}[\text{UTM}]$$  \hspace{1cm} (385)

Furthermore, as $\tilde{M}_{\infty}$ is recursively enumerable there would exist a specific enumeration strategy (i.e. an exact Turing machine $TM$) for which the partial
enumeration of Dom[O] by O is and remains equal to the reference manifest throughout all additions\(^3\) of programs to the manifest, such that the following is an identity throughout the evolution of the system:

\[
\{p_1, p_2, \ldots, p_n\} \equiv \hat{M}
\]  

(386)

With this equivalence, we have essentially constructed two computationally equivalent chapters. The first chapter tells the usual story; let’s call it chapter-\(U\) for universe. In this story, the universe verifies the vast majority of all experiments, and the observer simply ‘lives’ inside the reference manifest; its participation, if any, corresponds at most to a partial enumeration of a subset of \(\hat{M}\). Roughly, this story describes the observer primarily as a spectator. The second chapter, lets call it chapter-\(O\) for observer, tells another story; the observer takes the primary role. In fact, the reference manifest is constructed entirely by the observer as it practices science in nature. The universe is a participatory-universe.

The two stories although computationally equivalent, do produce different interpretations. Indeed, in the first story both the existence of the universe and of the observer remains a mystery, whereas in the second story the universe can be considered, remarkably, entirely emergent as a result of observer participation. The second story thus explains more than the first. This second story is the one that we will explore in this discussion.

Because there exists an enumeration strategy \(\text{TM}\) in which the partial enumeration by \(O\) is and remains equal to \(\hat{M}\) throughout the evolution of the system, we can now formulate a mathematically precise definition of the observer:

**Definition 25 (Observer).** An observer \(O\) is a program \(\text{TM}\) which recursively enumerates the domain of science. Specifically, \(\text{TM}\) outputs a sequence of numbers (where each number is mapped to an element of the domain of science).

**Definition 26 (Reference Observer).** If the partial enumeration of \(O\) is the reference manifest, then we use the over-circle symbol and call \(\hat{O}\) a reference observer\(^4\).

### 11.2 An observer-centric model of reality

Before the heliocentric model of the solar system, there was of course, the geocentric model. The latter, despite being ontologically incorrect, could nonetheless

---

\(^3\)Allowing removal of programs from \(\hat{M}\) under macroscopic changes do not break the equivalence, but do introduce some subtleties. Each time a program is removed, a different \(\text{TM}\) must be picked to define the new enumeration strategy, which defines another observer (Definition 25).

\(^4\)The reference observer is not unique for a given reference manifest. Although all reference observers for a given reference manifest do uniquely produce the reference manifest from their partial enumeration, they may start to diverge from the reference manifest as more experiments are enumerated. Observers who begin to diverge from the reference manifest as more programs are enumerated would cease to be qualified as reference observers.
account for the movement of the planets with remarkable precision (Claudius Ptolemy (90-168) - Almagest), but instead of simple elliptical motion, one had to utilize a plurality of circles within circles (epicycles). It is in fact a mathematical theorem that if one uses sufficiently-many epicycles, then one can approach any smooth curve to any degree of precision. Interestingly, it was long recognized that one could simplify the epicycle description to a ellipsoid description by placing the Sun, not the Earth, at center of the solar system (the first recorded heliocentric model is attributed to Aristarchus of Samos (c. 270 BC)). Although significantly simpler, the replacement at the time was perceived to be philosophically preposterous, and to reduce controversy the heliocentric model was publicly advertised as a clever mathematical trick (De Revolutionibus (1543) by Nicolaus Copernicus, published posthumously, contained an unsigned preface by Osiander arguing that the system proposed by Copernicus is useful for computation but need not be true).

Here, to attenuate controversy we may be tempted to limit our claim to stating that the re-shuffling of formal science to an observer-centric model of reality is, also, a mathematically useful change of perspective, such that numerous philosophical problem dissipate away, and such that the laws of physics naturally emerge from the description. However, not unlike the passage from the geocentric model to the heliocentric model, there is here a more powerful truth at play by taking the re-shuffle literally.

To fundamentally understand the power of this re-shuffling, we have to start at the very beginning of the rational inquiry. Allow us first to explain the situation using classical philosophy then we will modernize the argument. We will recall the philosophy of René Descartes (1596–1650), the famous 17th century french philosopher most directly responsible for the mind-body dualism ever so present in western philosophy. Descartes’ main idea was to come up with a test that every statement must pass before it will be accepted as true. The test will be the strictest test imaginable. Any reason to doubt a statement will be a sufficient reason to reject it. Then, any statement which survives the test will be considered irrefutable. Using this test, and for a few years, Descartes rejected every statement he considered. The laws and customs of society, as they have ambiguous logical justifications, are obviously amongst the first to be rejected. Then, he rejects any information that he collects with his senses; vision, taste, hearing, etc, on the grounds that a “demon” (think LSD) could trick his senses without him knowing. He also rejects the theorems of mathematics on the grounds that axioms are required to derive them, and such axioms could be false. For a while, his efforts were fruitless and he doubted if he would ever find an irrefutable statement. But, eureka! He finally found one which he published in 1641. He doubts of things! The logic goes that if he doubts of everything, then it must be true that he doubts. Furthermore, to doubt he must think and to think, he must exist (at least as a thinking being). Hence, ‘cogito ergo sum’, or ‘I think, therefore I am’. This quite remarkable argument is, almost by itself, responsible for the mind-body dualism of western philosophy.

In the Treatise of man (written before 1637, but published posthumously) and later in the The passions of the soul (1649), Descartes presents his picture
of man as composed of both a body and a soul. The body is a 'machine made of earth' and the soul is the center of thoughts. Communication between the two part would be handled by the 'infamous' pineal gland, whose inner working he covers in great detail in few hundred pages. Specifically, he states "[The] mechanism of our body is so constructed that simply by this gland's being moved in any way by the soul or by any other cause, it drives the surrounding spirits towards the pores of the brain, which direct them through the nerves to the muscles; and in this way the gland makes the spirits move the limbs"[68].

Now, if we give him the benefit of the doubt and we were to replace the word "spirits" with the word "electrical signals" (characterized in 1791 by Luigi, as bioelectromagnetics) and the word "pineal gland" with "cortex", then it's actually not too far off...

We believe that there exists a way to extend Descartes' universal doubt argument to a much larger body than merely the human body, but to achieve the extension, surprisingly, we will have to think about the problem purely mathematically (in terms of programs) rather than biologically. The key is to realize that the statement 'I think, therefore I am', as powerful as it is, is nonetheless the very small tip of a very deep logical iceberg. As we will show, there exists an infinite number of mathematical statements which would survive Cartesian doubt. Consider a statement of the kind:

\[ 1 + 1 = 2 \] (387)

As per his methodology, Descartes is correct to doubt of such a statement. Indeed, Peano's axiom (PA) are assumed to be true without proof and thus do not survive Cartesian doubt; consequently, neither does the statement. However, if we rewrite the statement as:

\[ \text{PA} \vdash (1 + 1 = 2) \] (388)

(to be read as: PA proves that one plus one equals two)

then the statement embeds within its description all assumptions necessary to prove it. Remarkably, as one cannot doubt the formal proof that verifies the statement, then all such statement survives Cartesian doubt. In this formulation, the truth of a statement is conditional upon the inclusion of all required assumptions within its description. In logic, such statements are known as necessary truths and they form a rich mathematical space. They are not to be confused with tautologies which are only defined for propositional logic, a much simpler decidable theory.

In fact, we can show that the set of all necessary truths is as complex as the whole of mathematics (it is Turing complete). Indeed, consider all statements of the form \( A \) proves \( B \). Asking for the verification that \( A \) proves \( B \) can be equated to asking for a proof that TM halts for \( h \). A theory of necessary truths has a domain equatable to the domain of the observer, of science and of a universal Turing machine.

88
This final equivalence produces yet another chapter containing the dictionary of science, giving formal science its strongest possible philosophical footing. We now have the opportunity to construct an argument, using formal science as a mathematical description of the observer, such that the laws of physics, including the description and existence of the cosmos, are entirely implied as a necessary truth merely by the existence of the observer.

Indeed, Axiom 2 is merely a consequence of demanding that the reference manifest $\mathcal{M}$ enumerated by $\mathcal{O}$ be formally verified. In algorithmic thermodynamics the resource used for the verification are a set of computing resources:

$$\mathcal{N} = \{$$
$$\mathcal{O}_f = \sum_{p \in \mathcal{D}} O_f[p] \rho[p],$$
$$\mathcal{O}_k = \sum_{p \in \mathcal{D}} O_k[p] \rho[p],$$
$$\cdots$$
$$\}$$

(389)

and in geometric statistical physics, these resources are the geometric bulk resources (allowing verification of the reference manifest by means of quantum computing / geometric path integral):

$$\hat{\mathcal{N}} = \{$$
$$\mathcal{E}_1 \mathcal{X}_1 = \mathcal{E}_1 \sum_{p \in \mathcal{D}} X_1[p] \rho[p],$$
$$\mathcal{E}_2 \mathcal{X}_2 = \mathcal{E}_2 \sum_{p \in \mathcal{D}} X_2[p] \rho[p],$$
$$\cdots$$
$$\}$$

(390)

With Axiom 1 and Axiom 2, the world is an experimentally-verified system. Finally, adopting the principle of insufficient reason to the observer’s enumeration of the domain of science, turns out to produce a model, known as physics, as the undefeatable placeholder model.

With these axioms and definitions, formal science becomes merely a proof that the property of ‘surviving Cartesian doubt’ tickles down from $\text{Dom}[\mathcal{O}]$ all the way to the cosmos (Section 8.5, on $\Lambda$CDM). A version of Descartes’ datum is alluded to with these replacements: ‘think’ becomes ‘participate’, ‘being’ becomes ‘observer’ and ‘body’ becomes ‘universe’ yielding the observer participates, therefore the universe emerges.

**Conclusion 1 (Emergence of the Universe).** The universe (cosmos) is automatically emergent as the bulk geometric description of $\mathcal{O}$’s verified recursive
11.3 Interpretation of quantum mechanics

What is generally considered the quantum measurement problem; that is, to explain why the system ends up in a given state, randomly selected from a post-measurement mixture, is not an add-on in formal science, but is instead a theorem resulting from the relationship between science and physics. In formal science, the quantum measurement "problem" is in fact not a problem, but instead a necessary injection in order to derive the laws of physics. The injection makes the reference manifest maximally informative; in fact, nature would be literally uninformative (in the Shannon sense) without it. 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:

$$|\psi\rangle|\phi\rangle \rightarrow \sum_n c_n |\chi_n\rangle|\phi_n\rangle$$  \hspace{1cm} (391)

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:

$$|\psi\rangle|\phi\rangle|e\rangle \rightarrow \sum_n c_n |\chi_n\rangle|\phi_n\rangle|e_n\rangle$$  \hspace{1cm} (392)

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 (Section 7.1), we have presented an alternative take on decoherence. The reference manifest enumerated by the observer, in geometric statistical physics, acquires the shape of the light cone, and its verified content is bounded to the time-like region. Whereas outside the time-like region of the light cone, the system is described using a geometric description neutral to the fact that these experiments are not yet verified, which is remarkably equivalent to quantum mechanics. Specifically, we have shown that decoherence occurs at the time-like/space-like boundary of the geometric path integral; 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.

However, the fundamental question remains; 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 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.

Formal science proposes to address the problem from the other direction by taking as its axiom the ontological existence of the state of affairs reference by $\mathcal{M}$. The primary connection is that the quantum measurement is quantified, in formal science, by natural information, and furthermore, that the entropy of natural information is the primary ingredient which allows us to derive the laws of physics in the first place. 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, formal science postulates that the reference manifest (Axiom 1), describing the state of affairs of the world, 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 8).

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, practicing science and thus having access to natural information, 'has knowledge of' the reference manifest, but the laws of physics, derived from formal 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 merely logically derivable by the observer (the set of all manifests). Augmented with Assumption 4, the gap is precisely the sum-total of all quantum measurements required to connect the solutions of the laws of physics to the reference manifest.

The quantum measurement problem appears unexplained 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, inspired by empirical data, simply 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 formal science: natural information must be erased to derive the laws of physics using science as the starting point.

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 pattern about nature, such that the pattern survives all future additions of experiments. 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 (even the one that didn’t happen) are compatible with the laws of physics' can be made. Unavoidably, 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 formal 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 (it is a universal mathematical structure), and the entropy of natural information is maximized to generate the broadest possible rules, yielding 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 27** (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 may 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 has no concept of "it could have happened, but it didn’t"; as far as its concerned, 'it did not happen' is identical to 'it cannot happen'. The manifest theory is invalided at the next event, as it is unable to reliably produce the correct output before said event occurs.

Picking the laws of physics as our explanatory tool of choice, rather than the manifest theory, is a choice we make because we prefer to understand the world by a pattern which holds for all present and future events that can be registered by the manifest, rather than by a brute computation immediately falsified at the next event. This 'preference' is formalized implicitly by Assumption 2, the fundamental assumption of 'nature'. Indeed, when one is presented with a world that exists brutally, 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. The "price to pay" to have laws of physics that are not falsified by the next random event, is for the pattern to hold for all possible events.

We note that it may seem intuitive to think that it should be the reference manifest which ought to obey the laws of physics, but it is not; it is in fact the world $W = (D, M, N)$, intentionally the bare minimum to define natural information, which is constrained by them. This may not seem too far off, but still, incorrectly thinking that $M$ does is a source of confusion.
In formal science, the reference manifest is the only manifest whose corresponding states of affairs exists ontologically. Listing the other manifests as hypothetical alternatives, a step necessary to formulate a pattern and to calculate the sum of the entropy of natural information 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 formal science and the others, occurs. For formal 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 ontological existence. Formal 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.

Our claims regarding the quantum measurement problem can be see as a direct consequence of taking the axioms of statistical physics literally, and understanding that the corpus of physics is derivable as a macroscopic description of the observer’s enumeration of the domain of science which is attributed as the microscopic description. This is essentially a switcheroo on the standard place of the observer in ordinary the statistical physics. In ordinary statistical physics, the framework is usually interpreted in the sense that the system (for instance an ideal classical gas) does occupy a specific micro-state at any given time, however, the observer does not know which of the possible states it is. Then under conditions of equilibrium, one can simplify the description using a handful of macroscopic variables (usually $E, V, N$ which are known to the observer). In our interpretation however, the observer (aware of the measurement result) is instead attributed as the micro-state of the statistical system via its enumeration of the domain of science. The macroscopic description is logico-deductive model formulated by the observer under the principle of insufficient reason due to the inherent inability to provide an account for the choice of partial enumeration. Finally, as they are derived from statistical physics, the laws of physics acquire the principle of maximum entropy with respect to the space of possible solutions.

**Conclusion 2** (QM interpretation proposed by formal science). *Formal 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). Formal 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. Formal 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. The observer is responsible of the recursive enumeration of the domain of science and thus is attributed the role of the micro-state of the system of statistical physics. The laws of physics are derived, under the principle*
of insufficient reason, as the macroscopic description of a system of statistical physics, and therefore acquire the principle of maximal entropy over the space of possible solutions.

12 Conclusion

Formal science produces an observer-centric description of the universe. Formal science is the minimal (no physical baggage) formal description of the observer sufficient to prove physics from first principles. Furthermore, via Axiom 1 and Axiom 2, the universe is automatically emergent as the basket of verification resources which holds the proof of all experiments realized by the observer(s).

References


