The road from reality: A purely mathematical proof of the existence of the observable universe

Alexandre Harvey-Tremblay¹,²

May 21, 2020

¹ independent researcher
² ahtremblay@gmail.com

Figure 1: This chart describes the most fundamental description of reality I believe to be possible. Specifically, my proposal is sufficiently powerful to inherit the coveted indubitable property of the Cartesian universal doubt method then to carry it forward such that the structure of reality, in the form of a model of the observable universe including the laws of physics, is inherited indubitably from pure reason, whilst avoiding the pitfall at the mind-body problem. The strategy is to construct a purely mathematical model of science, then to solve it for the laws of physics. In the model, reality is described as a set of realized experiments (the set of what "I" can prove), and the laws of physics are those that remain universally valid for all permutations or rearrangements of experiments under the only constraints of the set of all possible transformations, which we call nature. Essentially, experiments constrained by nature equals physics, in the most general mathematical sense imaginable.
Contents

1 Notation 3

2 Towards a mathematical theory of reality 4
  2.1 What, exactly, is wrong with postulated laws? 5
  2.2 What, exactly, is missing from mathematics? 11
  2.3 What is reality contingent upon? 15
  2.4 Hint: John A. Wheeler 19
  2.5 Hint: Gregory Chaitin 23

I Science 24

3 The Axioms of Formal Science 25
  3.1 Recap: measurement problem 28
  3.2 The true cause of measurements 29

II Nature 31

4 Technical Introduction 32
  4.1 Recap: Statistical Physics 32
  4.2 Recap: Algorithmic Thermodynamics 35
  4.3 Attempt 1: Algorithmic Equation of State 38
  4.4 Hint: Entropy, Information and Space-time 43
  4.5 Attempt 2: Designer Ensemble 45

5 Universal Thermodynamics 46
  5.1 Recap: Geometric Algebra 51
  5.2 Universal Representation 54
  5.3 Universal Metric 54
  5.4 Universal Ensemble 67

III Physics 71

6 Results (Spacetime Entropy) 71
  6.1 Law of Inertia 71
  6.2 Beekenstein-Hawking Entropy 72
  6.3 Action 73
  6.4 Fermi-Dirac statistics of events 74
  6.5 Quantum Mechanics 76
  6.6 Decoherence at the Time-like/Space-like Boundary 78
  6.7 Quantum Field Gravity 80
7 Results (Cosmology) .................................................. 81
  7.1 de Sitter Space ................................................. 81
  7.2 Cosmological Pressure ....................................... 82
  7.3 Cosmological Inertia ......................................... 83
  7.4 Entropic Derivation of $\Lambda$CDM ......................... 84

8 Conclusion ............................................................. 85

IV Annex ................................................................. 85

A Universal Interval and Norm of a universal event in four dimensions ............................................ 85

1 Notation

Parentheses will be used to denote the order of operations, and square brackets will be used 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 $S$ the action. Sets, unless a prior convention assigns it another symbol, will be written using the blackboard bold typography (ex: $L$, $W$, $Q$, etc.). Matrices will have a hat (ex: $\hat{A}$), vectors will be in bold (ex: $a, A$) and most other constructions (ex.: scalars) will have normal typography (ex. $a, A$). Finally, the identity matrix is $\hat{1}$, the null matrix is $\hat{0}$ and the unit pseudoscalar (of geometric algebra) is $I$. 
2 Towards a mathematical theory of reality

Figure 2 summarizes the current practice of physics. In it, both theoretical and empirical physics work in "tandem" to eventually, and hopefully, converge towards a correct model via an iterative falsification/refinements process. In the one hand, postulated laws are compared to empirical laws, and measured experimental states to predicted experimental states. Any discrepancy then ought to trigger a modification of the postulated laws, and the process begins again.

Figure 3 summarizes a reorganization which I believe is more representative of reality, and mathematically much more powerful: it is in fact able to derive the laws of physics as a theorem, thereby providing an account for their origins. In my proposal we find Axiomatic Science in lieu of theoretical physics, as a "mathematical reproduction" of empirical physics. Axiomatic science, as a formal system, is able to leverage the precision, "idealization" and clarity that mathematical formalisation often provides and applies it to the practice of science. It can best be visualized as a reversal of the relationship...
between laws and states typical of theories of physics. Its primary strength is that its starting point is anchored within reality in the form of postulated experimental states. Furthermore, the laws derived from axiomatic science are mathematically derived in a manner conceptually identical to their empirical counterparts. An empirical law is derived by repeating an experiment over a wide range of (similar) conditions then a general pattern is identified, and a universal equation of state in axiomatic science is derived as the universal pattern found by permuting over all possible arrangements of postulated experimental states. Because of its parallels with the real experimental case, a very strong case can then be made that this is the philosophically safest procedure possible to derive/define the law of physics.

I propose, and hope that this manuscript will make abundantly clear, that having neglected to fully formalize the practice of science within mathematics such that the corpus of physics is derivable as its theorem, is one of the biggest missed opportunity of theoretical physics.

2.1 What, exactly, is wrong with postulated laws?

As Exhibit A, consider a recent book, titled "The Road to Reality: A complete guide to the laws of the universe" by Roger Penrose, spanning over 1123 pages organized in 32 chapters from natural numbers to complex numbers to manifolds to quantum field theory (and beyond!). At each step of the way, the author introduces a few more mathematical concepts (in the forms of postulates or definitions) with the goal to bring us ever closer to "reality". If one’s goal is to postulate one’s way to reality, then Mr. Penrose is definitely the man to speak to. His book embodies, in my opinion, the most complete work in line with this methodology. But full stop, near the end on page 1033, Penrose ends with the following conclusion:

"I hope that it is clear, [...] our road to the understanding the nature of the real world is still a long way from its goal."

then continues with:

"If the ‘road to reality’ eventually reaches its goal, then in my view there would have to be a profoundly deep underlying simplicity about that end point. I do not see this in any of the existing proposals."

If even Mr. Penrose does not see a road to reality in any of the existing proposals, then what hope is there for the rest of us. Generally speaking, is erecting a massive "tower of postulates" for all to see, truly the best way to get to reality?
My earliest memory of thinking about this was on my second day of school (I promise it is relevant), but before I can explain what happened and why it happened I need to lay out the context leading up to it. My father’s strategy of choice to prepare me for the world was, I would summarize, to reinforce my understanding and manage my expectations of the world by constantly trying to transform the environment against my expectations (often while I wasn’t looking). I believe that he intuitively felt that by permuting over all possible (reasonable) states of the environment was the best and possibly only way to make sure I never develop an idea that is disconnected from reality. A specific example that comes to mind was one Easter when he brought a large chocolate bunny home and two smaller ones for myself, himself and my mother, respectively. I ate a tiny little bit around the ear of my chocolate bunny, then safely placed it in the cupboard. On the next day I woke up to find that half of my bunny was eaten, and my father is insisting that I am the one who ate it yesterday. Upon my objection, he insisted that I am just confused about the quantity I ate because he remembers it clearly, causing me to ponder on whose memories can be trusted more; his or mine, and how would I know. To cope with these random transformations and constant requisitioning of the assumptions, I came to the conclusion that I had to train myself not to inject any of my biases into my expectations of the world and just accept that the present state of the world is the foundation to reality, and my expectations of its future states (and in the extreme case even my memory of past states could be questioned) can be no less than any possible permutation of the environment. I can perhaps assign a likelihood to each scenario and perhaps have a backup plan for any undesirable scenario, so as to manage the risk — and that’s likely the best case scenario for surviving in the wild. I entered school with that mindset.

On my first day of school, the teacher taught us that one plus one equals two (and showed us how to work the symbols out as an equality). I remember being so flabbergasted by the genius of this equation that I barely slept during the night. Then, on the second day, the teacher extended this concept to all the numbers: “We learned yesterday that one plus one equals two, but it also works with two plus three equals five, and with three plus one equals four, and so on”. Then at some point she said, “and this is why if you take a rock from outside and then grab another rock, you will have two rocks in your hand”. As soon as she say that, my face changed completely to absolute distrust. I could not understand why she seemingly conceived of the relationship between ‘rules on a blackboard’ and ‘reality’ in the opposite logical direction of its true entailment. Of course, at that age I wasn’t able to articulate that thought using the language that I
use in the present text — I just had the intense intuition that she misunderstood something fundamental about reality and therefore her statements had to be verified before they could be trusted. So during the lunch break (for about 1.5 hours) I set out to do just that. I picked up rocks from the schoolyard and added them all out, permuting over the different arrangements I could construct and by so doing, verified a (tiny) subset of arithmetic. Okay, so I have established that it works with rocks, but does it work with... branches? So I went to get branches, and verified it again, and sure it worked for branches too. One of the other kids asked me what I was doing and I told him that I was trying to verify that what the teacher had said was true. He asked, surprised, "oh. You don’t believe her?". I responded along the lines of: "I am almost certain that she is right, but I cannot take the risk to take it on faith". Eventually, the bell rang and I ran out of time. Back in class began a long process of ruminating over what had transpired. Before I could continue with the program, I had to somehow grind away at the claim that all arithmetical permutations of numbers holds in reality. Clearly, it works with small numbers; I in fact just recently verified it in the schoolyard. For numbers larger than what I could personally verify, I convinced myself of the somewhat reasonable argument that possibly millions of other human beings where taught these equations before me and themselves have surely verified very exhaustively the claims. However, I reckoned that there was still a limit (a very large one indeed) beyond which arithmetical statements remain unverified by anybody. And an even a larger limit beyond which nobody ever could (presuming that the resources of our universe are finite). I could not rule out the fact that, outside some verifiable boundary, arithmetic expresses statements that are outside the scope of reality; thus I held as suspect the inclusion of ‘unbounded’ arithmetic within the “tower of postulates”.

Consequently, during the following school years, I developed and nurtured a healthy existential angst regarding our willingness to use an unscoped axiomatic basis (first with arithmetic, but eventually with any mathematical or physical theory) which I know does not connect exactly to reality. I tried to express my concerns a number of times with my teachers, but I do not think I made myself sufficiently clear as I recall one of the responses to be that I would learn all about rocks in the third year of high school. So I waited to let the program unfold expecting an eventual deconstruction of the disconnected tower of postulates in some upcoming more advanced classes, but the deconstruction never came; instead the complexity simply piled up at an exponential rate; from classical mechanics to eventually quantum field theory and everything in between. In my mind each additional layer would take us further away from reality. It would
take me decades to merely acquire the technical language sufficient to understand the problem clearly, then to pinpoint exactly what causes it, and finally how to cure it.

Returning to the question at hand, I have absolutely zero belief whatsoever that building a tower of postulates will ever bring us closer to reality. It is quite the historical/psychological curiosity why theoretical physics ended up being constructed in terms of a reversed logical entailment:

\[ \text{postulated-laws} \rightarrow \text{derived-states} \]  \hspace{1cm} (1)

Instead of its true logical entailment:

\[ \text{postulated-states} \rightarrow \text{derived-laws} \]  \hspace{1cm} (2)

And consequently that it deprived itself (and us) from answers to the most fundamental questions. This inverted construction is present in all formal physical theories; from Newton to quantum field theory. As an example, let us consider Newton’s second law of motion, mathematically expressed as follows:

\[ \sum_{F \in F} F = ma \]  \hspace{1cm} (3)

To come up with such a law, one presumes that Newton at least reviewed the published experimental data of his time, in addition to have conducted numerous experiments of his own. So clearly the law is logically entailed by ‘something’, and that ‘something’ is empirical evidence. How shaky then is the logical foundation of a theory that claims that a law (known to be derived from ‘something’) is an axiom (derived from nothing)? In fact, the consequence of writing \( F = ma \) as an axiom instead of as a theorem is that we end up inadvertently erasing the most important principles. Specifically, we create what amounts to a cargo cult, and allow me to explain.

A cargo cult is characterized as a belief, by a primitive or under-developed culture, that building an airstrip or a tower out of bamboo sticks will trigger the arrival of modern re-supply transport planes to deliver highly desirable cargo, based on the observation that a technologically advanced society has previously build a functional cargo-receiving airstrip nearby. These cults were first reported in Melanesia in the late 19th century following contact with western societies. According to one theory, the belief is held due to a lack of proper understanding of supply chain logistics essential to the delivery mission, as well as to an unawareness of the necessity of
building the airplanes in some (out of sight) assembly plant. What is the parallel with modern theoretical physics? In theoretical physics, we construct the largest “tower of postulate” that we can, whilst erasing from the formalism all of the logistics that brings us the laws of physics from reality (by writing them as axioms instead of as theorems), yet we somehow expect reality to be delivered to us on a silver platter by merely having constructed the tower. Well, it turns out that the structure of reality is all in the supply chain, and not in the tower.

I also place a not insignificant part of the blame in the human bias to wish to set the foundation of a mathematical theory to its simplest expression. It appears that since $F = ma$ is simpler than “100 pages of experimental data”, then it gets to be the axiom and not the data, even though reality is logically entailed in the reverse. Extending the argument to something as complex as the observable universe may appear as another problem and that may also have something to do with it. Let us consider a theory which takes the present experimental arrangement of the entire observable universe as its axiomatic basis. Since it may require upwards $10^{122}$ bits of information\(^3\) to write down its axiom, it could therefore be qualified as intractable. Even if such a theory could logically imply no wrong (by virtue of its axiom being an exact description of reality), one might nonetheless want something simpler (hence the bias part). My proposal however is able to derive the laws of physics without having to individually interrogate all $10^{122}$ bits of reality, by instead using algorithmic information theory to produce a concise representation of any and all possible experimental states, and to study the universal equation of state resulting from all arrangements and re-arrangements of experimental states. This preserves the universality of the problem yet makes it tractable.

I feel I must apologize to Roger Penrose for singling out his book and for the blunt introduction; In fact, I do have the utmost respect for the quality of his book as a reference tool of the mathematical concepts important to physics, and I have relied upon his work to formalize my own work in this very paper (we do after-all recover the laws of physics here, therefore a good portion of the tools remain usable). The clarity, utility and completeness of his book is in my opinion unmatched by any other modern works. Consequently, my intent is of course not to be attribute fault to Mr. Roger Penrose or to his book, but more to use his book as an illustration of the current state of affairs and (incorrect) expectations of the field as a whole, of which Penrose’s book just happens to be the single most complete embodiment of such.

I note that various other more technical problems can be associated, this time, to the practice of science itself. Many of these have

---

been discussed ad nauseam, so let us just reiterate them quickly for the sake of completeness. For instance, we can wonder if the iterative falsification/refinements process eventually converges to the truth. In principle, it is strictly possible to construct a theory whose domain is immune to the falsification/refinement algorithm. Russell warned us about the dangers of the teapot, but it was believed (or hoped) by many that we would be sufficiently alert to avoid the trap. However, the largesse of string theory caused many to pause and reconsider (this will be Exhibit B). String theory has avoided confirmation/falsification for some 35-plus years. Although thousands have participated in the attempt, thus far, the falsification/refinements process has not been able to restrict the "landscape" of string theory, which, without those checks and balances, has grown largely unimpeded. It may seem that out of concerns for the scientific process we ought to at some point stop investigating it, but beware this could also be a mistake; what if 'the truth' was just a few more years away? To stop or not to stop, that is the dilemma of all non-presently falsifiable, but possible, leads.

To accommodate long-tail processes such as string theory or others, it is more convenient to think of the problem of science as a halting problem instead of a convergence problem. I would in fact mark the late 20th century onwards practice of theoretical physics by the passage from an immediate falsification requirement\(^4\) (practicing science as a convergent algorithm) to a long-tail non-convergent approach (practicing science as a halting problem). In this contest, the problem adopts the least constraining form: will the falsification/refinement algorithm ever halt on the truth? As a halting problem, we are then free to investigate a given theory for as long as it remains fashionable, safely tucked away from the "dangers" of falsification and still collect dues from the science faculty, because the process could, in principle, eventually just halt on it (without actually converging towards it) and thus its still science.

But with this now less restrictive definition, new dilemmas emerges: are we to investigate every possible non-presently falsifiable, but strictly possible, leads (of which there are infinitely many) with the same intensity as we do for string theory? Without convergence, any possible lead could be the one that happens to halt, and could do so without any prior warnings. Clearly we don’t have the financial or physical resources to investigate everything. To avoid this dilemma, maybe we need to tweak the algorithm some more; but then we are in danger of entering a perpetual tweaking-hell; where tweaking the algorithm is on par with the difficulty of finding the actual solution.

In comparison, my proposal only requires an equivalence-thesis between measured states and postulated states, and thus has none of

\(^4\) We recall the 1920-1970 period in which the theoretical research on quantum field theory/standard-model was experimental tested in particle accelerators oftentimes within just a few years of their theoretical publications.
these problems. If the thesis holds, the laws derived from the process cannot diverge from reality.

2.2 What, exactly, is missing from mathematics?

Proving the existence of the observable universe purely mathematically is counter-intuitive and therefore it is often assumed to be an impossibility. However, in contradiction with expectations, a number of years ago I was able to lay out a precise path able to do so. Here, I will provide a simplified example of my technique, which illustrates the key concepts, then further in the paper we will thoroughly investigate the technique to produce a mathematical theory of reality.

Specifically, as our starting example, I will create a formal mathematical theory that has a shelf-life. Wait, "a shelf-life", in a mathematical theory... a shelf-life like with milk, or eggs? Yes, a shelf-life; meaning, the mathematical theory is perfectly usable today, but in some amount of "time" it will eventually rot. To the best my knowledge, rotting mathematical theories are a novel invention.

The construction is surprisingly simple, yet its philosophical implications are incredibly powerful. To construct such a theory, I simply obfuscate a statement behind a computationally-intensive algorithm that I then add as an axiom. For instance, consider the contradictory statement of arithmetic $1 + 1 = 1$ that we assume I have encrypted using a secure\(^5\) perfect\(^6\) hash function.

For example, suppose my hash produces the following result:

$$\text{hash}[1 + 1 = 1] = \text{fa1869db4bf8f1767a5446b6a9290243}$$

Specifically, the hash function takes as input an element of $\mathbb{L}_{PA}$, the set of all valid sentences of arithmetic, and outputs an element of $\mathbb{L}_{hex}$ the set of all hexadecimal sentences:

$$\text{hash} : \mathbb{L}_{PA} \rightarrow \mathbb{L}_{hex}$$

$$\text{statement} \mapsto \text{hash}$$

I also define the inverse function:

$$\text{bruteforce} : \mathbb{L}_{hex} \rightarrow \mathbb{L}_{PA}$$

$$\text{hash} \mapsto \text{statement}$$

The bruteforce function finds the solution by brute force: it hashes all statements of $\mathbb{L}_{PA}$ in shortlex in a loop then halts once it finds the statement that matches the hash, then it outputs said statement. Reversing the map of a hash function is, by design, computationally intensive.

\(^5\) A secure hash function can only be inverted by brute force.
\(^6\) A perfect hash function is an injective function that maps each input to an hash, with no collisions.
Now, let me define a new axiom as follows:

**Definition (Axiom of rot).**

\[
\text{rot} := \text{bruteforce}[\text{fa1869db4bf6f176fa5446b6a9290243}] \tag{7}
\]

Finally, using the axiom of rot, I define a new formal theory as the union between the axiom of rot and the Peano’s axioms of arithmetic (PA):

**Definition (Rotting arithmetic).**

\[
\text{rot} \cup \text{PA} \tag{8}
\]

In the present case since I already revealed the rot statement to you, it follows that you know that Rotting arithmetic is ultimately inconsistent without having to execute the brute force function. But consider instead the following axiom:

**Definition (Axiom-X).**

\[
\text{Axiom-X} := \text{bruteforce}[0\text{cf}e383362bc6d7ac429a5755f605] \tag{9}
\]

Now I ask you, knowing the hash but not the statement, is the formal theory comprised of PA \(\cup\) Axiom-X, consistent or inconsistent? Maybe the original statement I chose was \(1 = 0\) (inconsistent), or maybe it was \(1 + 1 = 2\) (consistent). It may not be so obvious now whether Axiom-X causes the theory to rot or not, is it? If you are willing to work at it, you will eventually find the non-obfuscated form of the axiom by brute force. In this context, I find it rather illustrative to employ the terms fresh/rotten (as opposed to consistent/inconsistent) to accentuate the timely connection between finitely axiomatic systems and some notion of work. A finitely axiomatic system is either fresh (if no contradictions are known) or rotten (if contradictions are known). We note that one who randomly proves theorems in rotten arithmetic will almost certainly map out a very large portion of standard arithmetic before the axiom of rot becomes a problem. We also note that no finitely axiomatic system can rot without work.

Consider the case where Axiom-X may have been hashed such that more work would be required to brute force the solution than what is available in the universe. Seth Lloyd\(^7\) estimates that there are approximately \(10^{122}\) bits (and approximately the same amount of operations) available for computations in the universe. What if our brute force function requires, say, \(10^{122} + 1\) bits or higher to halt? Such a finitely axiomatic system, although rotten in principle, could actually never rot in our present day universe. Its shelf-life would exceed the age and size of the universe. Rotting arithmetic, with a

> 10^{122} bits bruteforce function would be mathematically rotten, but "physically" fresh.

As per the Gödel incompleteness theorem, we recall that a (sufficiently expressive) finitely axiomatic system cannot prove its own consistency. It could be the case, hypothetically, that some finitely axiomatic system, perhaps believed to be consistent, contain a deeply hidden contradiction. In fact, since the dept of mathematical proof complexity knows no bound, then a contradiction could be injected, accidentally or on purpose, at any level of complexity within a theory. My specific example with a bruteforce function shows how to purposefully inject a contradiction at a tunable level of complexity, but nonetheless, in principle, all (sufficiently expressive) finitely axiomatic systems have the potential to rot. In the general case, mathematics offers us no tool to rat out rot, other than pure computing power.

For the present example, I have used a hashing function in order to make my point obvious; inverting a hashing function is known to be computationally intensive, thus we immediately notice a connection between work and our knowledge of the non-obfuscated form of the axiom. But do not let the presence of an hashing function distract you; in fact, all mathematical theorems require the consumption of some, always non-zero, quantity of computing resources to be proven. In essence, all mathematical theorems are hidden behind a "computing pay-wall" which must be paid to unlock the proof. In many day-to-day cases the price is negligible and thus goes unnoticed. Ex: prove that 1 + 1 + 1 = 3 is a theorem of PA — the truth of this statement is immediately obvious and so we do not easily notice the computing cost, but it is there nonetheless. All proofs have a computing cost, whether the proofs are verified by computer or by any other devices.

Taking a practical example, if two mathematicians are stuck on a desert island, one might ask the other "Is ZFC consistent?". It is quite likely that they may produce an answer along the following lines: "Although according to Gödel’s incompleteness’ theorem ZFC cannot prove its own consistency, it has been studied for over 80-90 years by hundreds of thousands of people. Certainly if there was any (obvious) contradictions it would have been found by now. The fact that no one has so far found any is very convincing evidence (but of course will never be not proof of) that no contradictions will be found". So why do I believe very strongly that ZFC is consistent, but I would have must less faith in whatever some random guy posted online yesterday, even if I may not be immediately aware of any contradictions in either theories? If the consistency of a theory is entirely implied by its axioms, why does work (or time spent studying it)

---

have anything to do with our belief in its consistency? There are no frameworks which quantifies the intuition. We aim do understand why it appears to be the case that a person who spends, say, $10^{20}$ bits and operations of computation on a theory is seemingly closer to identify rot (if rot is present) than a person who spends only, say, $10^5$ bits and operations of computation, and therefore to quantify why one may logically have more faith in the former than the later.

So now comes the problem of actually constructing such a framework. To do so, we will rely on algorithmic information theory, initially on the works of Gregory Chaitin, but also on the more recent works of Baez and Stay regarding algorithmic thermodynamics which imports the tools of statistical physics into algorithmic information theory. Using these tools, we will create a statistical ensemble of programs comprised of a manifest, a domain and a ground state. The elements of the trio are related to each other as shown on Figure 4.

On the left, we have the set of axioms of the finitely axiomatic system. In the middle, we have the manifest. The manifest is the instantaneous state of the system; specifically, it is the set of all statements that are proven and verified to be true by the system. In this sense, the manifest is the mathematical description of the proven reality of the system. Then, on the far right we have the domain spanned by the axioms. In the general case, the domain is a non-computable infinite set comprising all statements provable from the axioms. The manifest is always sandwiched between the axioms and the domain, and is related to them as follows:

$$\text{Axioms} \subseteq \text{Manifest} \subset \text{Dom}[\text{Axioms}]$$

In the ground state, the manifest is equal to the axioms. To leave the ground state (and thereby prove any theorems), the system must consume mathematical work. It is the algorithmic equation of state that quantifies the amount of mathematical work required to excite the system as the manifest is moved further along the axis.

---


Figure 4: The algorithmic equation of state quantifies the amount of mathematical work required to excite the system to a given manifest.
2.3 What is reality contingent upon?

Philosophy allows us to understand why the this framework produces a purely mathematical description of reality. Specifically, its boils down to the logical relationship between manifest and mathematical work:

**Definition 1** (The fundamental structure of reality). The existence of the manifest is contingent on the existence of mathematical work. Consequently, since the manifest exist indubitably (as we will argue in a moment), then so does mathematical work.

To fundamentally understand the power of this statement, we have to start at the very beginning of the rational inquiry. Allow me first to lay out the groundwork using classical philosophy, then we will use our tools to 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 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 dubious 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) because a “demon” could trick his senses without him knowing. He also rejects the theorems of mathematics, because 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.

How did Descartes addressed the mind-body problem? 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 parts would be handled by the ‘infamous’ pineal gland, whose inner working he covers in great
detail in almost a 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"\textsuperscript{11}. We note that bio-electrical signals were characterized in 1791 by Luigi Galvani, as bioelectromagnetics some \approx 150 years later. Of course, the Cartesian pineal gland theory was incorrect, but the problem on how to connect the mind and the body remains.

The proof of the ‘cogito ergo sum’ by Descartes is the only proof I have encountered that exceeds the absolute genius that one plus one equals two is (that is quite an achievement). Furthermore, it is the only proof of the existence (of something/anything) that I find to be completely satisfactory within the body of philosophy; all other attempts fall short in one way or another. However, as good as the first part is, appealing to the biological structures of the brain to attempt a connection from mind to body violates the standards of the proof for the simple and obvious reason that the existence of these biological structures do not survive universal doubt and thus ought not to be used in the proof. How can this pitfall be avoided?

It may be subtle, but one may nonetheless have already noticed that the construction of the statistical ensemble of programs I have proposed, such that proving statements from a finitely axiomatic system is contingent on the consumption of mathematical work is a repeat, in disguise, of the universal doubt method of Descartes (including its conclusion), but applied to the whole structure of reality. Essentially, Descartes may have described the very first universal proof checker, and applied it to the set of universal mathematical statements (although in an informal manner, and consequently missed out on a massive opportunity).

Let us now do an exercise that will reveal this missed opportunity. We will attempt to produce a description of reality constructed entirely and exclusively out of indubitable statements, of the kind that Descartes would not rule out by his universal doubt method, justified on the simple idea that we will eventually want to claim that reality exist indubitably and that those statements will be key.

Let us start by defining what a language is:

\textbf{Definition 2 (Language).} A language \( L \), with alphabet \( \Sigma \), is the set of all sentences \((s_1, s_2, \ldots)\) that can be constructed from the elements of \( \Sigma \) and it includes the empty sentence \( \emptyset \):

\[
L := \{ \emptyset, s_1, s_2, \ldots \} \tag{11}
\]
For instance, the sentences of the binary language are:

\[ L_b := \{ \emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots \} \quad (12) \]

and its alphabet is:

\[ \Sigma_b = \{ 0, 1 \} \quad (13) \]

**Definition 3 (Shortlex ordering).** A shortlex ordering is a list of the sentences of \( L \), first ordered by length from shortest to longest, then alphabetically.

For instance, the shortlex ordering of \( L_b \) is:

\[ (\emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots ) \quad (14) \]

Let us now define a "Cartesian-Turing machine" which works as follows:

**Definition 4 (Minimal proof checker).**

\[ \text{MPC} : \quad L \longrightarrow \{ 1, \# \} \]

\[ \text{sentence} \quad \longmapsto \quad \text{result} \quad (15) \]

Under the hood, the machine works as a brute force automatic theorem prover. Specifically, the machine contains a set of internal rules of inference (logical axioms) which it uses to attempt to formulate a proof that the input sentence is true within the rules. If a proof is found, then \( \text{MPC}[\text{sentence}] \) outputs 1 otherwise it never halts. The machine may take a very long time to find a proof, but time is not our concern as we only require that should a proof exists, it eventually finds it. For example, once given a sentence as input, the machine could analyse every sentences of \( L \) in shortlex one by one and by scheduling its work according to a dovetailing algorithm, until one is found to be the proof, then outputs 1 and halts.

Let us now investigate how MPC behaves using an example. Say Descartes feeds the sentence \((1 + 1 = 2)\) to the MPC. Will the machine find a proof for it? Well lets see. To prove \((1 + 1 = 2)\), one requires PA (or an equivalent). However, since the machine only contains logical axioms, it will never halt because no proof will ever be found. On the surface, it may seem that the conclusion of Descartes regarding the idea that one can doubt of all mathematical theorems because the axioms they rely upon need not be true, is sound.

However, once in a while something quite interesting happens. Let’s say we feed all sentences of \( L \) in shortlex to MPC using dovetailing scheduling. Eventually this statement will be feed to MPC:
The statement states: PA proves that one plus one equals two. As we did with the previous example, we also ask here will MPC[PA ⊢ (1 + 1 = 2)] eventually halt? In this case, the answer is yes. Indeed, PA just so happens to be the missing part required for the machine to prove the statement. The statement supplied the missing set of axioms that is required to prove itself.

The second step of the exercise will be to construct a manifest exclusively using such statements. There are of course infinitely many statements of the type $A \vdash B$, and such statements include all possible mathematical proofs for all possible mathematical theories. Their significance rely on the fact that they are a means to describe reality to any level of complexity or expressivity desired or required without having to construct a preliminary “tower of postulates” to do so.

Let us define a set $M$, which we call a manifest, of $n$ statements provable by MPC:

$$M := \{A_1 \vdash B_1, A_2 \vdash B_2, \ldots, A_n \vdash B_n\}$$

The $B$s are the theorems, the $A$s are their axiomatic basis, and each pair forms an individual fact of the system. In the case of a real system such as that of the observable universe, any associated manifests would of course be extremely large and as such $n$ is also very big\textsuperscript{12}. The union of all $A$s forms the axiomatic basis of a “computational theory of everything” the $B$s. Although it is not the typical theory of everything referenced in physics, it does provide the means to prove all facts of the system, but it does so in a patchy manner and it is susceptible to initial graining effects. In the case where two or more facts have the same axiomatic basis, say $A_4 = A_9 = A_3$, the theory forms a “logical grain” such that the theorem $B_4, B_9, B_3$ are entailed from the same axiomatic basis. In the ground state, all facts simply imply themselves $B_1 \vdash B_1, B_2 \vdash B_2, \ldots$ (e.x “there is milk in the fridge because there is milk in the fridge (no insight)”), but to leave the ground state, mathematical work must be produced. Outside the ground state, there exists a “spread” between the axioms and the facts, whose magnitude is quantified in terms of mathematical work. The larger the spread, the more “algorithmically-rich” the system is and the more insightful the system has the potential to be.

The set of all scientific theories for the system are the various patches or grains that are formed as the mathematical work is expanded to increase the spread. Scientific theories are subject to re-

\textsuperscript{12} Seth Lloyd. Computational capacity of the universe. Physical Review Letters, 88(23):237901, 2002
finements as grains are fused or reorganized. Describing reality as 
a manifest allows one to produce a plurality of logically independent 
scientific theories, each valid within their own domain of 
applicability and each corresponding to a grain. In this sense, we may 
suspect that all scientific theories are subject to possible falsification-refinements as more mathematical work is produced. One may 
also expect knowledge of the facts themselves to be essential. 

But one formulation, qualified as universal, rises above all others; 
it is a universal equation of state which remains unfalsified with 
respect to all possible configurations of facts, grains and state of 
mathematical work. This universal equation of state is the origin of 
the laws of physics in the system; it is the theory of everystate. 

We are almost there, there is just a small thing missing. In fact, 
mathematical algorithms admit a slightly more general formulation 
than what we have used so far and it just so happens to be re-
quired in other to derive the universal equation of state as the laws 
of physics. The language $A \vdash B$ is not entirely universal as, amongst 
other things, it leaves out all theories that may not admit the $\vdash$ sym-
bol. Consequently, to be truly universal we will think of statements 
as arbitrary programs that halt, instead of as sentences. Let us in-
vestigate how mathematical algorithms truly connect to the physical 
world, as we enter the next hint.

2.4 Hint: John A. Wheeler

So, what is the right way to think of reality in terms of information? 

Information, physics and entropy have, of course, a long and rich 
history. As an example, let us recall the Landauer\textsuperscript{13} limit, an expres-
sion for the minimum amount of energy required to erase one bit of 
information for a system at thermodynamic equilibrium:

$$E \geq Tk_B \ln 2$$ \hspace{1cm} (18)

where $E$ is the energy (in Joules), $T$ is the temperature (in Kelvin) 
and $k_B$ is Boltzmann’s constant. Such relation, although considered 
extremely fundamental, cannot in our case be used as the starting 
equivalence. Indeed, since we are not yet at the stage where energy 
or temperature are defined, we thus cannot use a relation which 
refers to them in order to connect mathematical information to physi-
ical reality.

What about other connections found in the literature? A strong 
contender relying upon modern notions such as black-hole entropy 
(or more generally entropy-bearing horizons) and the holographic 
principle suggests a connection between information (on the surface

\textsuperscript{13} Rolf Landauer. Irreversibility and heat 
generation in the computing process. 
IBM journal of research and development, 
5(3):183–191, 1961; and Rolf Landauer 
et al. Information is physical. Physics 
of an horizon) and physical states (in the bulk of the enclosed volume). Let's hypothesise how and if we could use this contender in our case. Perhaps we are to map our manifest to the surface of an information-bearing horizon then use the holographic principle to recover the bulk? Alas, no - the same problem as before also occurs here but instead of energy/temperature, we have geometry/surface-gravity as the presumed pre-existing physical concepts. Unfortunately, any pre-existing physical quantities needed to formulate a relation between information and some element of physical reality, precludes those quantities from being given an origin within this information.

So we ask, is such a connection possible - can all physical concepts be reduced to information, or are there irreducible physical concepts? As we work our way to the proposed solution, let us review the best contender I have thus far identified in the literature. We will now investigate two hints; the first by John A. Wheeler regarding the 'participatory-universe' hypothesis, the second by Gregory Chaitin regarding the undecidability of mathematical formalism and the link between mathematics and science. Together these two hints will allow us to identify a universal relation entirely free of physical baggage.

We summarize John A. Wheeler's participatory universe hypothesis as follows. First, for any experiments, regardless of their simplicity or complexity, the registration of counts (in the form of 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" 14) that what we consider to be the "it" is simply one out of many possible mixtures 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."

The bit is the anchor to reality. The bit would come into being in the final act, so to speak, and then constrains 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 alternatives 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 apriori (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? Thirty-nine years later, the results of the Planck Collaboration\(^{15}\) 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 such that the future time reversal, if any, ought to be derived 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 an example of what might be done. Some 11 years later he corrects his approach to the participatory-universe hypothesis.

In “Information, physics, quantum: The Search For Links”, he pro-

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

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

\[ 111010110001001110101010101 \] (19)

How valuable would this person find 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 stay the same, but their meaning is now completely different. Thus, some form of a 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 for a meaningful description.

But how do we describe the very complex world of experimental equipment without invoking physical baggage? I have the impression that this may have been a primary roadblock encountered by John A. Wheeler: formalizing equipment-evoked response seems to require some physical description of said equipment, and as this would contain physical baggage, then the fundamentality of the theory would be compromised.

The solution that I 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. This is how the connection to the algorithmic formulation is made. As shown on Figure 1, instead of connecting information to some complex pre-existing physi-
tical quantity, we here connect the algorithm to an experiment. The 'it' of Wheeler is a consequence of protocol-evoked responses, not equipment-evoked responses — a very important but subtle difference. As we will see with the next hint, shifting the description from equipment to protocol is the key to make the endeavor mathematically precise.

2.5 Hint: Gregory Chaitin

If we are to construct physical reality as a statistical ensemble of experiments (Figure 1), then there is likely one better way to recover the laws of physics; one must simulate the practice of science within the ensemble. But before we can formalize science within mathematics, we first need to identify a mathematical structure that behaves as science does.

Gregory Chaitin summarizes his work on the halting probability\(^{17}\), the \(\Omega\) construction, in the book "Meta Math!"\(^{18}\). Let \(U\) be the set of all universal Turing machines, then:

\[
\Omega : U \rightarrow [0, 1] \\
UTM \rightarrow \sum_{p \in \text{Dom}[UTM]} 2^{-|p|} \quad (20)
\]

The image of \(\Omega\) is a set of 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 \(\Omega\), and then come back to this section.

In the book "Meta Math!" Gregory Chaitin states that the following is his 'strongest' incompleteness theorem:

"A finitely axiomatic system (FAS) can only determine as many bits of \(\Omega\) as its complexity.

As we showed in Chapter V, there is (another) constant \(c\) such that a formal axiomatic system FAS with program-size complexity \(H[FAS]\) can never determine more than \(H[FAS] + c\) 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 regarding 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 axiomatized 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, 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, "Proving Darwin: making biology mathematical", which references many of these concepts.
Part I

Science

3 The Axioms of Formal Science

Axiomatic Science is likely to be both the least-biased and the most universal framework of reality possible. It can be applied to any and all physical systems. In the case were formal science is applied to the sum-total of all reality, both the microscopic description and the equation of state are universal; the first in the mathematical sense (it spawns the domain of a universal Turing machine) and the second in a physical sense (it accounts for all possible transformations of the system).

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

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

(21)

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 \( \text{TM} \) 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 that hold are the programs that halt. The set includes all provable mathematical statements and it is universal in the computer theoretic sense.

Definition 6 (Domain of science). Let \( \mathbb{D} \) be the domain (Dom) of formal
science. We can define $\mathcal{D}$ in reference to a universal Turing machine $UTM$ as:

$$\mathcal{D} := \text{Dom}[UTM] \quad (22)$$

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

**Definition 7 (Manifest).** A manifest $M$ is a subset of $\mathcal{D}$:

$$M \subset \mathcal{D} \quad (23)$$

**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.

**Definition 8 (States — set of all manifests).** The set of all manifests comprises all possible state of affairs of the world, and is defined as:

$$S := \mathcal{P}[\mathcal{D}] \quad (24)$$

where $\mathcal{P}[\mathcal{D}]$ is the powerset of $\mathcal{D}$. Of course, $M \in S$.

**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 $\tilde{M}$ which corresponds to its state:

$$\exists! \tilde{M} \quad (25)$$

- $\tilde{M}$ is called the ‘reference manifest’.
- The symbol $M$ will denote any manifest in $S$, whereas $\tilde{M}$ 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 $\tilde{M}$ exists ontologically.
- Unique to $\tilde{M}$, its elements are verified.
Intuition: The reference manifest is how the world presents itself to us in the most direct, unmodelled, uninterpreted and uncompressed manner. Brutely knowing the manifest is how one perceives the world without understanding any patterns and without knowing any laws of physics.

As stated in the previous philosophical section, the fact that the experiments of the reference manifest are verified implies, implicitly, the existence of mathematical work in quantities exactly sufficient to verify them, which in the present context, we call nature.

**Definition 9 (Nature).** Nature is a system of mathematical work used to verify experiments. Thus, experiments are verified in nature. Precisely, nature is the set of all functions mapping manifests in $S$ to computing transformations in $T$ (usually taken to be a complete set of matrices) — $g : S \rightarrow T$, and is noted as $\mathcal{N}$ such that:

$$\mathcal{N} := S^T$$

(26)

Note: We will investigate nature in more detail (including many examples) in the next section.

As infinitely many manifests $M$ can be constructed from the elements of $D$, one may wonder why it is the reference manifest $\hat{M}$ 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 $S$ according to a probability measure $\rho[M]$.

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 $\hat{M}$ and not $M$ is actual. Essentially, this is a side-effect of the notion that the state of affairs represents the axiomatic basis of 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 $\hat{M}$ to information, and it is precisely because $\hat{M}$ is randomly selected from a larger set that this is possible. We briefly recall the mathematical theory of information of Claude Shannon: Specifically, $\hat{M}$ will be interpreted as a message randomly selected from the set $S$. Using $\rho[M]$, we will be able to quantify the amount of natural information in the message $\hat{M}$.

**Definition 10 (Natural Information).** We define natural information as the information one gains by knowing which manifest is randomly selected from $S$, according to the probability distribution $\rho[M]$ that maximizes the entropy under the constraint of nature $\mathcal{N}$.
Those familiar with statistical physics will no doubt have identified that the setup of axiomatic science is fundamentally that of a statistical ensemble. Indeed, in the present case, the laws of physics will be derived as a universal equation of state of nature valid over all possible permutations or rearrangements of manifests or of experiments, and resulting from maximizing the entropy in the usual sense of statistical physics. Nature, defined as an uncountable set of functions, here acts as the universal constraints of the system. In this context, we refer to the laws of physics as the ‘structure’ of reality so as to contrast them to typical theories of physics in which the laws of physics are postulated, in which case we call them the ‘model’ of reality. Axiomatic science through its connection to entropy reveals that the laws of physics is the structure that maximizes the amount of information associated to the message $M$ under the constraint of nature: intuitively, they are the laws that make reality maximally informative. In the next section, we will produce an exact mathematical expression for natural information, but before we do that, let us now apply the full power of axiomatic science to resolve the quantum measurement problem, and to truly understand what a law of physics is.

3.1 Recap: measurement problem

We are all familiar with the measurement problem. In quantum physics the unitary evolution of the wave function $\psi(r,t)$ is deterministic, but the notion breaks down if measurements are made by the observer.

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:

$$|\psi\rangle|\phi\rangle \rightarrow \sum_n c_n |\chi_n\rangle|\phi_n\rangle$$

(27)

After the measurement the system is in one of eigenstates $|\chi_n\rangle$ with probability $|c_n|^2$. Why it is that the deterministic unitary evolving system adopts the “un-deterministic” initiative to collapse itself randomly in one of multiple states after measurement is quite the mystery. Nothing in theoretical quantum physics predicts that such a thing would occur. Consequently, the notion of the measurement is introduced into quantum mechanics 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. Decoherence from the environment $|e\rangle$ is
introduced as follows:

\[ |\psi\rangle|\phi\rangle|e\rangle \rightarrow \sum_n c_n |\chi_n\rangle|\phi_n\rangle|e_n\rangle \tag{28} \]

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. Decoherence is a possible mechanism to account for why a quantum superposition of eigenstates is unlikely to be observed macroscopically because interaction with the environment very quickly causes the system to evolve towards a classical probability distribution. However, decoherence is ultimately of no help in regards to explaining why the system is in a definite state after the measurement.

The problem remains: how does the system go from a mixture of states to a definite state, 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 considered non-falsifiable.

### 3.2 The true cause of measurements

Rather than taking some arbitrary set of laws as postulates, axiomatic science addresses the problem from the other direction by taking as its sole axiom the ontological existence of the state of affairs referenced by \(\mathbb{M}\). An interpretation of quantum mechanics is available within this setup as a direct consequence of applying the concepts of statistical physics in terms of microstate/macrostate. The terms microstate and macrostate are somewhat misleading in the present context because we are of course not dealing with difference in sizes
as we typically do in usual statistical physics. Here, microscopic refers to a description of the system in terms of a set of experiments in the form of a reference manifest $\hat{M}$, whereas macroscopic refers to a description in terms of the constraints of nature $\mathcal{N}$ neutral to any specific manifests and invariant with respect to any rearrangement of manifest or experiments within a manifest. Both the microstate and the macrostate description have the ability to spawn the whole observable universe.

This is essentially an inversion on the typical place of the observer in ordinary statistical physics, where it is associated to a large (macroscopic) device measuring things like the volume or the energy, etc. In axiomatic science, and with the inversion, the observer is aware of the measurement result provided by it’s knowledge of the microscopic description, whereas the macroscopic description is a logico-deductive model formulated by the observer under the principle of insufficient reason to bound the states the system could involve into. 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, which accounts for the quantum measurement problem when projected outside the scope of the reference manifest. Simply put, the laws of physics cannot tell us the result of a measurement, for the same reason that $pV = nRT$ cannot tell us where in the box the molecules of air are.

The quantum measurement problem appears as unexplained only when one postulates the laws of physics because such models erases the science part, and thus can no longer deliver the laws of physics at the end of the road. When one postulates the laws, then solves them, one obtains a plurality of manifests as possible solutions, only one of which is the reference manifest. Since one intuitively expected that the laws of physics ought to produce the reference manifest, one may then become baffled as to why the laws of physics produce a plurality of manifests as their solutions instead of just the reference manifest.

The fundamental motivation in maximizing the entropy of natural information is to release the constraints imposed by the microscopic description so as to facilitate formulating the broadest possible pattern about nature, such that the pattern survives all possible rearrangements of experiments or permutations of manifests. 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, one can say "I am a physicist, but I could have been a doctor instead", or one 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 one would have instead said: ‘I am a physicist, but I could have been superman’.

How credible is that claim? Supposedly, we may admit that 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, wilts ruling out only the ‘truly’ impossible.

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 possible (universal in the mathematical sense), and the entropy of natural information is maximized to generate the broadest possible rules using nature as the constraints, comprising the set of all possible transformations (universal in the physical sense), thereby producing the laws of physics as a universal equation of state.

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 associated to the recursive enumeration of the domain of science and thus is attributed to the role of constructing the micro-state of the statistical physics system.
Part II

Nature

4 Technical Introduction

To precisely quantify the relationship between natural information, entropy, mathematical work and how this implies the laws of physics as the bulk description of system of verified experiments, we will eventually introduce universal 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\textsuperscript{21}.

1. The average of all experimental measurements of a given observable in a macroscopic system converges to a well defined value, called a constraint.

2. "Any macroscopic system at equilibrium is described by the maximum entropy ensemble, subject to constraints that define the macroscopic system."\textsuperscript{22}

The first assumption is responsible for implying a number of fixed macroscopic quantities, known as the constraints. 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:

$$P := \left\{ \rho: Q \rightarrow [0, 1] \mid \sum_{q \in Q} \rho[q] = 1 \right\}$$

(29)

The observables, in general, are $n$ functions defined as:

$$O_i: P \rightarrow \mathbb{R} \quad \rho \mapsto \sum_{q \in Q} \rho[q] O_i[q]$$

(30)

where $O_i: Q \rightarrow \mathbb{R}$. All constraints are observables, but not all observables are constraints. Typical thermodynamic quantities are shown in Table 1.

The second assumption is responsible for implying the probability measure which maximizes the entropy:
Under said constraints. This probability measure, which can be obtained from the method of the Lagrange multipliers by maximizing the entropy under the constraints, is the Gibbs ensemble:

\[
S : \mathbb{P} \rightarrow [0, \infty[ \\
\rho \mapsto -k_B \sum_{q \in Q} \rho[q] \ln \rho[q]
\]

(31)

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

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

(33)

The observables form a set of \(n\) constraints that we call the set of observables, or the thermodynamic bulk state:

\[
O_i = Z^{-1} \sum_{q \in Q} O_i[q] \exp \left( -\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q] \right)
\]

(34)

The thermodynamic bulk quantities are also given by the following \(n\) relations:

\[
\frac{\partial \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i} = O_i
\]

(35)

And the variance by the following \(n\) relations:
\[
\frac{\partial^2 \ln Z[a_1, \ldots, a_n]}{\partial a_i^2} = (\Delta O_i)^2
\]  
(36)

The entropy for this ensemble is:

\[
S[a_1, \ldots, a_n] = k_B (\ln Z + a_1 \bar{O}_1 + \cdots + a_n \bar{O}_n)
\]  
(37)

Taking the total derivative of the entropy, we obtain:

\[
dS[a_1, \ldots, a_n] = k_B (a_1 d\bar{O}_1 + \cdots + a_n d\bar{O}_n)
\]  
(38)

which is called the equation of the state of the system.

Thermodynamics is derived from statistical physics which is concerned primarily by the equation of state (38). Thermodynamic changes (and cycles) can be realized by changing the quantities \(\{a_1, \ldots, a_n\}\) and/or by modifications of \(Q\). Under modification of \(Q\), usually by cross product: \(Q \times Q_1 = Q_2\), or by set complement \(Q \setminus Q_3 = Q_4\), quantities which are invariant \(\{a_1, \ldots, a_n\}\) 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:

\[
a_1 := \beta
\]  
(39)

\[
a_2 := \gamma
\]  
(40)

\[
a_3 := \delta
\]  
(41)

\[
A_1[q] := E[q]
\]  
(42)

\[
A_2[q] := V[q]
\]  
(43)

\[
A_3[q] := N[q]
\]  
(44)

the partition function would be:

\[
Z[Q, \beta, \gamma, \delta] = \sum_{q \in Q} \exp \left( -\beta E[q] + \gamma V[q] + \delta N[q] \right)
\]  
(45)

The Gibbs measure is:

\[
\rho(q, \beta, \gamma, \delta) = \frac{1}{Z} \exp \left( -\beta E[q] - \gamma V[q] - \delta N[q] \right)
\]  
(46)

The observables are:
\[ E = \frac{1}{Z} \sum_{q \in Q} E[q] \exp \left( -\beta E[q] - \gamma V[q] - \delta N[q] \right) \]  
\[ V = \frac{1}{Z} \sum_{q \in Q} V[q] \exp \left( -\beta E[q] - \gamma V[q] - \delta N[q] \right) \]  
\[ N = \frac{1}{Z} \sum_{q \in Q} N[q] \exp \left( -\beta E[q] - \gamma V[q] - \delta N[q] \right) \]

The entropy is:

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

and the equation of state is:

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

### 4.2 Recap: Algorithmic Thermodynamics

Many authors 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. First let us introduce \( \Omega \). Let \( U \) be the set of all universal Turing machines, then:

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

\[ \text{UTM} \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \]

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. Chaitin’s construction (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). 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 calculated the first 64 bits of \( \Omega \) for some well-defined

\[ C. \ H. \ Bennett, \ P. \ Gacs, \ Ming \ Li, \ P. \ M. \ B. \ Vitanyi, \ and \ W. \ H. \ Zurek. \ Information \ distance. \ IEEE \ Transactions \ on \ Information \ Theory, \ 44(4):1407–1423, \ July \ 1998; \ Gregory \ J. \ Chaitin. \ A \ theory \ of \ program \ size \ formally \ identical \ to \ information \ theory. \ J. \ ACM, \ 22(3):329–340, \ July \ 1975; \ Edward \ Fredkin \ and \ Tommaso \ Toffoli. \ Conservative \ logic. \ International \ Journal \ of \ Theoretical \ Physics, \ 21(3):219–253, \ Apr \ 1982; \ Andrei \ Nikolaevich \ Kolmogorov. \ Three \ approaches \ to \ the \ definition \ of \ the \ concept \ “quantity \ of \ information”. \ Problemy \ peredachi \ informatsii, \ 1(1):3–11, \ 1965; \ A \ K \ Zvonkin \ and \ L. \ A \ Levin. \ The \ complexity \ of \ finite \ objects \ and \ the \ development \ of \ the \ concepts \ of \ information \ and \ randomness \ by \ means \ of \ the \ theory \ of \ algorithms. \ Russian \ Mathematical \ Surveys, \ 25(6):83, \ 1970; \ R.J. \ Solomonoff. \ A \ formal \ theory \ of \ inductive \ inference. \ part \ i. \ Information \ and \ Control, \ 7(1):1 – 22, \ 1964; \ Leo \ Szilard. \ On \ the \ decrease \ of \ entropy \ in \ a \ thermodynamic \ system \ by \ the \ intervention \ of \ intelligent \ beings. \ Behavioral \ Science, \ 9(4):301–310, \ 1964; \ Kohtaro \ Tadaki. \ A \ generalization \ of \ chaitin’s \ halting \ probability \ omega \ and \ halting \ self-similar \ sets. \ Hokkaido \ Math. \ J., \ 31(1):219–253; \ 02 \ 2002; \ and \ Kohtaro \ Tadaki. \ A \ statistical \ mechanical \ interpretation \ of \ algorithmic \ information \ theory. \ In \ Local \ Proceedings \ of \ the \ Computability \ in \ Europe \ 2008 \ (CiE \ 2008), \ pages \ 425–434. \ University \ of \ Athens, \ Greece, \ Jun \ 2008 \]

universal Turing machine \( \text{utm}1 \in U \) as:

\[
\Omega[\text{utm}1] = 0.0000001000000100000110..._2
\]  

(53)

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\textsuperscript{28} suggested augmenting \( \Omega \) with a multiplication constant \( D \), which acts as an ‘algorithmic decompression’ term on \( \Omega \).

Chaitin construction \( \Omega[\text{UTM}] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \) \( \rightarrow \) Tadaki ensemble \( \Omega[\text{UTM}, D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|} \)

(54)

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

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

(55)

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 \( |\bar{p}| \):

\[
|\bar{p}| = \sum_{p \in \text{Dom}[\text{UTM}]} |p| 2^{-D|p|}
\]  

(56)

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

\[
S[\text{UTM}, D] = \ln \Omega + D|\bar{p}| \ln 2
\]  

(57)

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\textsuperscript{29} 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 constraints. 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 the Baez-Stay ensemble:

$$
\Omega : \mathbb{U} \times \mathbb{R}^3 \rightarrow \mathbb{R} \\
(\text{UTM}, \beta, \gamma, \delta) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-\beta E[p] - \gamma V[p] - \delta N[p]} \quad (58)
$$

Noting its similarities to the Gibbs ensemble of statistical physics, 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 \textit{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. $\rho = \gamma/\beta$ is the \textit{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 \textit{algorithmic potential} (analogous to chemical potential). Roughly speaking, this counts how much the mean log runtime increases when you increase the mean output while holding the number of programs in the ensemble and their mean length fixed.

"...

–John C. Baez and Mike Stay

From (Equation 58), 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 \textit{algorithmic heat} and \textit{algorithmic work}. Finally, we note that other authors have suggested other alternative mappings in other but related contexts\textsuperscript{30}. 


4.3 Attempt 1: Algorithmic Equation of State

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.

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 \( O_k[p] \) the program length, and \( o_f \) the computing frequency conjugated with \( O_t[p] \) the program time. The partition function is:

\[
Z : \mathbb{U} \times \mathbb{R}^2 \rightarrow \mathbb{R} \\
(\text{UTM}, o_k, o_f) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-o_k O_k[p] - o_f O_t[p]} \quad (59)
\]

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 \( P[D] \). Executing a manifest \( M \in P[D] \) 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 are:

\[
Z : \mathbb{R}^3 \rightarrow \mathbb{R} \\
(o_k, o_f, o_\mu) \mapsto \sum_{M \in P[D]} 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \tag{60}
\]

The corresponding probability measure is:

\[
\rho : W \times \mathbb{R}^3 \rightarrow \mathbb{R} \\
(M, o_k, o_f, o_\mu) \mapsto Z^{-1} 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \tag{61}
\]

The probability measure maximizes the entropy subject to the following bulk constraints:
\[ \mathcal{O}_x = \sum_{M \in \mathcal{P}[D]} O_x[M] 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \quad (62) \]

\[ \mathcal{O}_t = \sum_{M \in \mathcal{P}[D]} O_t[M] 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \quad (63) \]

\[ \mathcal{O}_n = \sum_{M \in \mathcal{P}[D]} O_n[M] 2^{-o_k O_x[M] - o_f O_t[M] - o_\mu O_n[M]} \quad (64) \]

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

- The computing repentry \(o_k\) counts how many times the average tape usage \(O_x\) must be doubled to double the entropy of the ensemble while holding the average clock time \(O_t\) and the average concurrency \(O_n\) fixed.

- The computing frequency \(o_f\) counts how many times the average clock time \(O_t\) must be doubled to double the entropy of the ensemble while holding the average tape usage \(O_x\) and the average concurrency \(O_n\) fixed.

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

\[ Z' : \quad \mathcal{U} \times \mathbb{R}^2 \quad \rightarrow \quad \mathbb{R} \]

\[ (\text{UTM}, o_k, o_t) \quad \rightarrow \quad \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-o_k O_x[p] - o_t O_f[p]} \quad (65) \]

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

- The Computing time \(o_t\) counts how many times the average clock frequency \(O_f\) must be doubled to double the entropy of the ensemble while holding the average tape usage \(O_x\) and the average concurrency \(O_n\) fixed.
2. Size-cutoff formulation:

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

(66)

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:

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

(67)

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:

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

(68)

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

Interpretation:

1. Feasible computing complexity:

The use of the letter \( O \) to identify thermodynamic observables is coincidentally very convenient here. Recall the Big O notation used in computational complexity theory to denote program complexity. 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 computational complexity theory considers this intractable problem to be an easier problem than one requiring \( n^2 \) steps. Consequently, computational complexity theory based on Big O notation does not quite connect to the physical reality of computation with limited available resources.

However, using an ensemble of algorithmic thermodynamics, a cost-to-compute, measured in entropy, can be attributed to carrying out a computation using finite resources.

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 61), has the following equation of state:

\[
\frac{dS}{\beta} = \frac{dO}{T} + \frac{dO}{t} + \frac{dO}{n} \tag{69}
\]

Using this equation of state, we can quantify the computing ‘distance’ between two states 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 and represents a typical equation of state of thermodynamics.

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

   \[
   Z(\beta) = \sum_{q \in Q} \exp(-\beta E[q])
   \]

   The average energy is given by

   \[
   \overline{E} = -\partial \ln Z / \partial \beta
   \]

   and its fluctuations are

   \[
   \overline{(\Delta E)^2} = \partial^2 \ln Z / \partial \beta^2
   \]

   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 act 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 is 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 a ‘supercomputer’.

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

4.4  **Hint: Entropy, Information and Space-time**

In 2002, Lloyd\(^\text{32}\) calculated the total number of bits available for computation in the universe, as well as the total number of operations that could have occurred since the universe’s beginning.

For both quantities (the quantity of bits stored in the universe and the quantity of operations made on those bits), Lloyd obtains the number \(\approx 10^{122}k_B[^\text{bit}]\). This number is consistent with other approaches; for instance, the Bekenstein-Hawking entropy\(^\text{33}\) of a ‘holographic surface’ at the cosmological horizon\(^\text{34}\) (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 of his paper. 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:

\[
\#\text{ops} \approx \frac{\rho_c c^5 t^4}{\hbar} \approx \frac{t_o^2 c^5}{G\hbar} = \frac{1}{t_p^2} t^2 \tag{70}
\]

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 \(\approx 10^{122}k_B[^\text{bit}]\). He states:

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

---


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.

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\textsuperscript{35} 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\textsuperscript{36} with the discovery of the Hawking temperature derived from quantum field theory on curved space-time. We note the discovery of the Bekenstein-Hawking entropy, connecting the area of the surface of a horizon to be proportional to one fourth the number of elements with Planck area that can be fitted on the surface:

$$S = k_B a^3 / (4hG) A.$$  

We mention Ted Jacobson\textsuperscript{37} 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\textsuperscript{38} 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\textsuperscript{39} 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 (\textsuperscript{40}), the Coulomb force (\textsuperscript{41}), Yang-Mills gauge fields (\textsuperscript{42}), and cosmology (\textsuperscript{43}). Some criticism has, however, been voiced\textsuperscript{44}, including by Visser\textsuperscript{45}.

Even more recently, a connection between entanglement entropy and general relativity has been supported by multiple publications\textsuperscript{46}.

Finally, we mention the body of work of George Ellis regarding


the evolving block universe hypothesis detailed in \(^47\) and the connection between space-time events, general relativity and quantum mechanics.

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

### 4.5 Attempt 2: Designer 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.

Finding a specific system of statistical physics means attributing an implementation to the thermodynamic observable functions \((O_x|p_i\rangle, O_t|p_i\rangle, \text{etc.})\) used in the partition function. A similar approach was 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. A "tower of postulates" would be sneaked into the definition of a postulated ensemble.

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:

\[
\text{d}E = T \text{d}S + p \text{d}V - \mu \text{d}N
\]  

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 geometric algebra applied to the thermodynamic observables such that the equation of state forms a complete basis of the \(n \times n\) matrix representation of all possible transformations of the system. In this

generalization, which we call universal thermodynamics, the general conservation relation above becomes a special case of an even more general conservation relation that, remarkably, has the suitable properties, and the equation of state is able to carry the system from any state to any state.

5 Universal Thermodynamics

I initially identified the potential to generalize statistical physics as I attempted to create thermodynamic cycles that are consistent with the symmetries of space-time. By doing so, I realized that such cycles could be produced if the relevant thermodynamic constraints obeyed a non-commutative algebra. By comparison, in normal thermodynamics, the constraints are scalars.

My initial goal was 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, observables, conjugate variables, cycles/transformations, etc.). To achieve that, it is helpful 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\(^4\). Furthermore, from G. W. Gibbons and S. W. Hawking’s 1977 article\(^49\), 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, I 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 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 $\mathcal{O}$. Indeed, to prevent $\mathcal{O}$ from knowing its future, future events must also be void of information from $\mathcal{O}$’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, simply by attributing an entropy to the microscopic states which comprises the random events spread in space-time. We will now describe the physical quantity relevant to universal thermodynamics.

**Definition 11** (Specific repetency). *As we construct an equation of state, a physical quantity will be introduced as the Lagrange multiplier: the specific repetency $\tilde{k}$ with units $m^{-1}$. The specific repetency is a proportionally factor that quantifies the entropy associated to the interval between two events in space-time. It replaces the role normally assumed by the temperature in thermodynamics.*

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1[q], X_2[q], X_3[q]$</td>
<td>space</td>
<td>[meter]</td>
<td>microscopic</td>
</tr>
<tr>
<td>$\tilde{k}$</td>
<td>specific repetency</td>
<td>[1/meter]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\overline{X}_1, \overline{X}_2, \overline{X}_3$</td>
<td>entropic space</td>
<td>[meter]</td>
<td>extensive</td>
</tr>
<tr>
<td>$X_0[q]$</td>
<td>time</td>
<td>[second]</td>
<td>microscopic</td>
</tr>
<tr>
<td>$c\tilde{k}$</td>
<td>specific frequency</td>
<td>[1/second]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\overline{X}_0$</td>
<td>entropic time</td>
<td>[second]</td>
<td>extensive</td>
</tr>
</tbody>
</table>

Table 4: The physical quantities of the geometric ensemble

This specific repetency is the conjugated variable to a distance $\overline{X}_1, \overline{X}_2, \overline{X}_3$ and/or time $c\overline{X}_0$. These quantities are summarized in Table 4. By convention, we will prefix the Lagrange multiplier with the word "specific" and its averaged conjugated quantity will be prefixed with the word "entropic". The specific repetency is an intensive quantity, whereas entropic space $\overline{X}_1, \overline{X}_2, \overline{X}_3$ and entropic time $c\overline{X}_0$ are extensive quantities. Indeed, a process taking 1 min followed by a process taking 2 min takes a total of 3 min (extensive). For the $\overline{X}_1, \overline{X}_2, \overline{X}_3$ quantities: 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 $c\tilde{k}$ (say once per second) has no impact on the frequency of the other elements of the group (intensive). The same argument applies to the specific repetency (intensive). The units of $\tilde{k}$ are $m^{-1}$, the units of $\overline{X}_1, \overline{X}_2, \overline{X}_3$ are the meters, the units of $\overline{X}_0$ are the seconds, and the units of $c\tilde{k}$ are $s^{-1}$.

We note that the temperature ($k_B T = 1/\beta$) has no central role in universal 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
A purely mathematical proof of the existence of the observable universe

system (using temperature, energy, and entropy) would be inapprop-
riate as the system would be outside equilibrium. However, with
our strategy, it is precisely because faster-than-light communication is
impossible that a quantity, the specific repetency — closely connected
to the speed of light, can take the role normally assumed by the tem-
perature as a Lagrange multiplier of the ensemble. As we will see, by
using this quantity as a Lagrange multiplier instead of the temper-
ature, the ensemble applies an entropy to all of space-time, with or
without horizons, and thus determines its complete structure.

To understand in more detail, let us investigate a hypothetical
thermodynamic transformation involving the observables \( \bar{X}_1, \bar{X}_2 \) and
\( \bar{X}_3 \). The equation of state of such a system would be:

\[
k_B^{-1} dS = \bar{k}(d\bar{X}_1 + d\bar{X}_2 + d\bar{X}_3)
\]

(72)

For a change over the quantities \( \bar{X}_1, \bar{X}_2 \) and \( \bar{X}_3 \) 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 go-
ing 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 Eu-
clidean 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 thermody-
namic 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:

\[
k_B^{-1} dS = \bar{k}(\sigma_1 d\bar{X}_1 + \sigma_2 d\bar{X}_2 + \sigma_3 d\bar{X}_3)
\]

(73)

We note that in this expression, \( S \) becomes an entropy vector, and
this will be addressed rigorously in the next section. But for now, we
will see that this entropy will become a real number by squaring it
and we will see that the entropy conforms to the Euclidean distance.
By squaring, we obtain:

\[
k_B^{-2} \bar{k}^{-2}(dS)^2 = \sigma_1^2 (d\bar{X}_1)^2 + \sigma_2^2 (d\bar{X}_2)^2 + \sigma_3^2 (d\bar{X}_3)^2
\]

\[
+ (\sigma_1 \sigma_2 + \sigma_2 \sigma_1) d\bar{X}_1 d\bar{X}_2 + (\sigma_1 \sigma_3 + \sigma_3 \sigma_1) d\bar{X}_1 d\bar{X}_3
\]

\[
+ (\sigma_2 \sigma_3 + \sigma_3 \sigma_2) d\bar{X}_2 d\bar{X}_3
\]

(74)

In the case where \( \sigma_1, \sigma_2 \) and \( \sigma_3 \) are commutative, the cross terms
\( \sigma_1 \sigma_2 + \sigma_2 \sigma_1, \sigma_1 \sigma_3 + \sigma_3 \sigma_1 \) and \( \sigma_2 \sigma_3 + \sigma_3 \sigma_2 \) do not cancel, but if they are,
say matrices, that obey the following relations:
\[
\sigma_1^2 = 1 \quad (75)
\]
\[
\sigma_2^2 = 1 \quad (76)
\]
\[
\sigma_3^2 = 1 \quad (77)
\]
\[
\sigma_1\sigma_2 + \sigma_2\sigma_1 = 0 \quad (78)
\]
\[
\sigma_1\sigma_3 + \sigma_3\sigma_1 = 0 \quad (79)
\]
\[
\sigma_2\sigma_3 + \sigma_3\sigma_2 = 0 \quad (80)
\]

Then, the cross-terms cancel and we obtain:

\[
k^{-2}\bar{k}^{-2}(dS)^2 = (d\bar{X}_1)^2 + (d\bar{X}_2)^2 + (d\bar{X}_3)^2 \quad (81)
\]

The entropy, here, is a real number again.

The resulting equation of state has the mathematical form of the Euclidean distance. 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 \(\bar{X}_1, \bar{X}_2, \bar{X}_3\) and \(c\bar{X}_0\) as its thermodynamic quantities. The equation of state, using the generators \(\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}\) is:

\[
k^{-1}\bar{k}^{-1}dS = \gamma_0 c d\bar{X}_0 + \gamma_1 d\bar{X}_1 + \gamma_2 d\bar{X}_2 + \gamma_3 d\bar{X}_3 \quad (82)
\]

Here, \(\bar{X}_0\) is an extensive quantity with units \(s\) and it is conjugated with \(c\bar{k}\) having units \(s^{-1}\). Squaring the equation of state gives:

\[
k^{-2}\bar{k}^{-2}(dS)^2 = \gamma_0^2 c^2 (d\bar{X}_0)^2 + \gamma_1^2 (d\bar{X}_1)^2 + \gamma_2^2 (d\bar{X}_2)^2 + \gamma_3^2 (d\bar{X}_3)^2
\]
\[
+ c(\gamma_0\gamma_1 + \gamma_1\gamma_0) \ d\bar{X}_0 \ d\bar{X}_1 + c(\gamma_0\gamma_2 + \gamma_2\gamma_0) \ d\bar{X}_0 \ d\bar{X}_2 + c(\gamma_0\gamma_3 + \gamma_3\gamma_0) \ d\bar{X}_0 \ d\bar{X}_3
\]
\[
+ (\gamma_1\gamma_2 + \gamma_2\gamma_1) \ d\bar{X}_1 \ d\bar{X}_2 + (\gamma_1\gamma_3 + \gamma_3\gamma_1) \ d\bar{X}_1 \ d\bar{X}_3
\]
\[
+ (\gamma_2\gamma_3 + \gamma_3\gamma_2) \ d\bar{X}_2 \ d\bar{X}_3 \quad (83)
\]

The cross-terms cancel, provided that the generators obey the
Then, the equation of state is:

\[
k_B^{-1}k^{-1}(dS)^2 = c^2(d\vec{X}_0)^2 - (d\vec{X}_1)^2 - (d\vec{X}_2)^2 - (d\vec{X}_3)^2
\]  

(95)

The square of the entropy is a real number again.

In the general case, one begins by defining an arbitrary non-commutative basis as follows:

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

(96)

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:

\[
\mathbf{e}_i \vec{X}_i = \sum_{q \in Q} e_i \vec{X}_i[q] \rho[q]
\]  

(97)

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 (38):

\[
k_B^{-1}k^{-1} dS = e_0 \, d\vec{X}_0 + \cdots + e_{n-1} \, d\vec{X}_{n-1}
\]  

(98)

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

$$e_i^2 = 1 \quad (99)$$
$$e_i e_j = e_j e_i \quad (100)$$

one obtains the relation $k_B^{-1} dS = \tilde{k} dX_0 + \cdots + \tilde{k} dX_{n-1}$, which is of the same mathematical form as equation (38) and describes a taxi-cab metric. Therefore, geometric thermodynamics is indeed a generalization of thermodynamics; a fact quite important to what we are trying to achieve. Indeed, statistical physics has 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.

5.1 Recap: Geometric Algebra

A geometric algebra $G$ is spawned by a generator relation:

$$e_\mu \cdot e_\nu = \frac{1}{2} (e_\mu e_\nu + e_\nu e_\mu) = g_{\mu\nu} \quad (101)$$

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

<table>
<thead>
<tr>
<th>basis elements</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>${1,}$</td>
<td>grade-0</td>
</tr>
<tr>
<td>$e_0, e_1, e_2, e_3,$</td>
<td>grade-1</td>
</tr>
<tr>
<td>$e_0 e_1, e_0 e_2, e_0 e_3, e_1 e_3, e_2 e_3,$</td>
<td>grade-2</td>
</tr>
<tr>
<td>$e_0 e_1 e_2, e_0 e_1 e_3, e_0 e_2 e_3, e_1 e_2 e_3,$</td>
<td>grade-3</td>
</tr>
<tr>
<td>$e_0 e_1 e_2 e_3$</td>
<td>grade-4</td>
</tr>
</tbody>
</table>

Multivectors of $G$ can be constructed as a linear combination of these basis elements. For instance:
Vectors and multivectors:

<table>
<thead>
<tr>
<th>example</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v := 1$</td>
<td>0-vector, or scalar (107)</td>
</tr>
<tr>
<td>$v := 3e_0 + 4e_1$</td>
<td>1-vector, or vector (108)</td>
</tr>
<tr>
<td>$v := 3e_0e_3 + 2e_2e_1$</td>
<td>2-vector, or bivector (109)</td>
</tr>
<tr>
<td>$v := 5e_0e_1e_2$</td>
<td>3-vector, or trivector (110)</td>
</tr>
<tr>
<td>$v := 2e_0e_1 \ldots e_k$</td>
<td>k-vector (111)</td>
</tr>
<tr>
<td>$v := 1 + 2e_0 + 5e_2e_1$</td>
<td>multivector (112)</td>
</tr>
</tbody>
</table>

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

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

$$v := r + r_0e_0 + r_1e_1 + r_2e_2 + r_{01}e_0e_1 + \ldots \quad \text{where } r, r_0, r_1, r_{01}, \ldots \in \mathbb{R}$$

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

$$v := z + z_0e_0 + \ldots \quad \text{where } z, z_0, \ldots \in \mathbb{C}$$

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 $\{e_0, e_1, e_2, e_3\}$ we denote the algebra by $\mathcal{G}_4$ generally, or $\mathcal{G}_4(\mathbb{C})$ if the algebra is complex with four generators, or $\mathcal{G}_4(\mathbb{R})$ if the algebra is real with four generators.

Furthermore, if the generator relation is orthogonal:

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

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

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

then,
\[ \gamma_0 \gamma_0 = 1 \quad (117) \]
\[ \gamma_1 \gamma_1 = -1 \quad (118) \]
\[ \gamma_2 \gamma_2 = -1 \quad (119) \]
\[ \gamma_3 \gamma_3 = -1 \quad (120) \]
\[ \gamma_0 \gamma_1 + \gamma_1 \gamma_0 = 0 \quad (121) \]
\[ \gamma_0 \gamma_2 + \gamma_2 \gamma_0 = 0 \quad (122) \]
\[ \gamma_0 \gamma_3 + \gamma_3 \gamma_0 = 0 \quad (123) \]
\[ \gamma_1 \gamma_2 + \gamma_2 \gamma_1 = 0 \quad (124) \]
\[ \gamma_1 \gamma_3 + \gamma_3 \gamma_1 = 0 \quad (125) \]
\[ \gamma_2 \gamma_3 + \gamma_3 \gamma_2 = 0 \quad (126) \]

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 multivectors \( v \) and \( u \) is:

\[ vu = v \cdot u + v \wedge u \quad (127) \]

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

\[ v := a e_0 + b e_1 \quad (128) \]
\[ u := c e_0 + d e_1 \quad (129) \]

Then, the geometric product is

\[ vu = (ae_0 + be_1)(ce_0 + de_1) \quad (130) \]
\[ = ae_0 c e_0 + ae_0 d e_1 + be_1 c e_0 + be_1 d e_1 \quad (131) \]
\[ = ac + ade_0 e_1 - db e_0 e_1 + bd \quad (132) \]
\[ = ac + bd + (ad - db)e_0 e_1 \quad (133) \]
\[ = v \cdot u + v \wedge u \quad (134) \]

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 $\sigma^{\mu\nu\rho} = \frac{1}{6} [\gamma^\rho, \gamma^\nu, \gamma^\mu]$

5. 1 matrix $\sigma^{\mu\nu\rho\delta} = \frac{1}{24} [\gamma^\rho, \gamma^\nu, \gamma^\mu, \gamma^\delta]$

The most important fact for us is that a complex linear combination using all 16 elements of $Cl_4(C)$ forms a complete basis of $4 \times 4$ complex matrices.

5.2 Universal Representation

In mathematics it is common to think of matrices as a representation of the geometric algebra. But if one has an affinity with matrices, can one not instead think of geometric algebra as a representation of such matrices? This may not be possible for all matrices, but for at least the complex $4 \times 4$ matrices and $Cl_4(C)$, each is a complete basis of the other. Since our starting point is algorithms/computation/ transformation, then as the first and foremost description, we can think of a ensemble of programs such that its possible transformations are described by a complete set of matrices. Then, changes in the system are governed by an equation of state able to cycle over all states attainable by these transformations. In this case, we will then think of the geometric algebra, forming a complete basis of such matrices, as a representation of these transformations. Consequently, our interpretation herein of space-time will be that it is a representation (a very visualizable one indeed) of such universal transformations.

5.3 Universal Metric

Definition 12 (Spacetime Event). A spacetime event $s$ is a $1$-vector of the space generated by $Cl_{3,1}(R)$:

$$s = X_0 e_0 + X_1 e_1 + X_2 e_2 + X_3 e_3$$  (135)

Definition 13 (Universal Event). A universal event $s$ is a multivector of the space generated by $Cl_4(C)$ that uses its complete basis:

$$s := a + X_0 e_0 + X_1 e_1 + X_2 e_2 + X_3 e_3 + E_1 e_0 e_1 + E_2 e_0 e_2 + E_3 e_0 e_3 + B_1 e_2 e_3 + B_2 e_1 e_3 + B_3 e_1 e_2 + V_0 e_1 e_2 e_3 + V_1 e_0 e_2 e_3 + V_2 e_0 e_1 e_3 + V_3 e_0 e_1 e_2 + b e_0 e_1 e_2 e_3$$  (136)

This definition generalizes to $n$ dimensions and to arbitrary geometric algebras $G_n$. 
It is well-known that the metric of space-time events is (in the orthonormal case / special relativity):

\[(ds)^2 = -(c \, dX_0)^2 + (dX_1)^2 + (dX_2)^2 + (dX_3)^2 \quad (137)\]

or (in the general case / general relativity):

\[(ds)^2 = g_{\mu \nu} \, dX^\mu \, dX^\nu \quad (139)\]

But how does a metric generalize to universal events? Another way of asking the question is: how can we define the distance between two arbitrary multivectors of the same geometric algebra? First, let me state my solution, then we will investigate it:

**Definition 14 (Universal Interval).** Let \( \Delta \hat{s} \) be the matrix representation of \( \Delta s \in \mathcal{G}_n \). Then, the universal interval \( \Delta s \in \mathcal{C} \) for a given matrix representation of \( s \) is given by this relation:

\[
\det \left[ \Delta s \hat{1} - \Delta \hat{s} \right] = 0 \quad (140)
\]

Solving it for \( \Delta s \) yields the universal norm of \( \hat{s} \).

We will now investigate how this definition compares to the usual definition of the interval. Let us recall that for a 1-vector of Euclidean space, its norm is defined as the inner product:

\[
\|v\|^2 := v \cdot v = x^2 + y^2 + z^2 \quad (141)
\]

It may be tempting to define the interval of a multivector as a simple extension of the usual inner product, but alterations are often required on the usual inner product to create an interval that is relevant for physics, and specifically which alterations to use or not to use is not always easy to guess right. My definition assigns a physically relevant notion of distance automatically to all multivectors even in non-obvious cases. In fact, the universal interval is the physically-relevant quantity describing the ‘length’ of vectors and multivectors in all geometric spaces. To help us understand how it works, let us first compare various kinds of established norms and intervals to our definition, and then we will look at novel cases. Unless otherwise stated, we will be working with \( \text{Cl}_4(\mathbb{C}) \) represented by the gamma matrices \( \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\} \) (which we note as \( \text{Cl}_4^0(\mathbb{C}) \)).
1. **Scalar**: The event $\mathbf{v} := a$ has the following matrix representation:

\[
\hat{v} := \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix}
\] (142)

The universal interval is obtained as follows:

\[
0 = \det \begin{pmatrix} \Delta s & 0 & 0 & 0 \\ 0 & \Delta a & 0 & 0 \\ 0 & 0 & \Delta a & 0 \\ 0 & 0 & 0 & \Delta a \end{pmatrix} \Rightarrow 0 = (\Delta s - \Delta a)^4 \Rightarrow \Delta s = \Delta a
\] (143)

The universal interval compares to the conventional definition as follows:

\begin{align*}
\Delta s &= \Delta a \quad \text{Conventional} \quad (144) \\
\Delta s &= \Delta a \quad \text{Universal Interval} \quad (145)
\end{align*}

In this case, both intervals are $a$.

2. **Pythagorean norm (2D)**: In $\text{Cl}_2(\mathbb{R})$, the event $\mathbf{v} := xe_1 + ye_2$ is:

\[
\hat{v} = x \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + y \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} x & y \\ y & -x \end{pmatrix}
\] (146)

Evaluating the universal interval produces this relation:

\[
0 = (\Delta s)^2 - (\Delta x)^2 - (\Delta y)^2 \Rightarrow (\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2
\] (147)

Which is the Pythagorean norm.

3. **Euclidean norm (3D)**: In $\text{Cl}_3(\mathbb{R})$, the event $\mathbf{v} := x\sigma_x + y\sigma_y + z\sigma_z$ is:

\[
\hat{v} = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix}
\] (148)

Evaluating the universal interval produces this relation:
\[ 0 = (\Delta s)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \implies (\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \]

which is the Euclidean norm.

4. **spacetime event:** The event \( v : t \gamma_0 + x \gamma_1 + y \gamma_2 + z \gamma_3 \) has the following matrix representation:

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

Evaluating the universal interval produces this relation:

\[ 0 = ((\Delta s)^2 - (\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2)^2 \]

Compared to the conventional approach, the intervals are:

\[ \eta_{\mu\nu} \Delta X^\mu \Delta X^\nu = -(\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 + (\Delta t)^2 \quad \text{conventional} \]

\[ (\Delta s)^2 = -(\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 + (\Delta t)^2 \quad \text{our definition} \]

We note that the conventional definition and our definition give the same interval.

5. **bi-vectors:** Let us now apply my definition in a geometric context where no intervals are currently known. The event \( b := E_1 \gamma_0 \gamma_1 + E_2 \gamma_0 \gamma_2 + E_3 \gamma_0 \gamma_3 + B_1 \gamma_1 \gamma_3 + B_2 \gamma_1 \gamma_2 + B_3 \gamma_1 \gamma_2 \), a bi-vector spanning all six 2-basis elements, has the following matrix representation:

\[
b := \begin{pmatrix}
-iB_3 & -iB_1 + B_2 & E_3 & E_1 - iE_2 \\
-iB_1 - B_2 & iB_3 & E_1 + iE_2 & -E_3 \\
E_3 & E_1 - iE_2 & -iB_3 & -iB_1 + B_2 \\
E_1 + iE_2 & -E_3 & -iB_1 - B_2 & iB_3 \\
\end{pmatrix}
\]
identified with physical relevance. The universal interval for this bivector is verbose:

\[
\begin{align*}
0 &= B_1^4 + 2B_1^2 B_2^2 + B_2^4 + 2B_1^2 B_3^2 + 2B_2^2 B_3^2 + B_3^4 \\
&+ 2B_2^2 E_1^2 - 2B_2^2 E_1^2 - 2B_3^2 E_1^2 + E_1^4 + 8B_1 B_2 E_1 E_2 \\
&- 2B_1^2 E_2^2 + 2B_2^2 E_2^2 - 2B_3^2 E_2^2 + 2E_1^2 E_2^2 + E_2^4 \\
&+ 8B_1 B_3 E_1 E_3 + 8B_2 B_3 E_2 E_3 - 2B_1^2 E_3^2 - 2B_2^2 E_3^2 \\
&+ 2B_2^2 E_3^2 + 2E_1^2 E_3^2 + 2E_2^2 E_3^2 + E_3^4 + 2B_1^2 ||s||^2 \\
&+ 2B_2^2 ||s||^2 + 2B_3^2 ||s||^2 - 2E_1^2 + ||s||^2 - 2E_2^2 ||s||^2 \\
&- 2E_3^2 ||s||^2 + ||s||^4 
\end{align*}
\] (155)

but its four roots are more tractable:

\[
\begin{align*}
- \sqrt{-B_1^2 - B_2^2 - B_3^2 + E_1^2 + E_2^2 + E_3^2 - 2i(B_1 E_1 + B_2 E_2 + B_3 E_3)} \\
\end{align*}
\] (156)

\[
\begin{align*}
\sqrt{-B_1^2 - B_2^2 - B_3^2 + E_1^2 + E_2^2 + E_3^2 - 2i(B_1 E_1 + B_2 E_2 + B_3 E_3)} \\
\end{align*}
\] (157)

\[
\begin{align*}
- \sqrt{-B_1^2 - B_2^2 - B_3^2 + E_1^2 + E_2^2 + E_3^2 + 2i(B_1 E_1 + B_2 E_2 + B_3 E_3)} \\
\end{align*}
\] (158)

\[
\begin{align*}
\sqrt{-B_1^2 - B_2^2 - B_3^2 + E_1^2 + E_2^2 + E_3^2 + 2i(B_1 E_1 + B_2 E_2 + B_3 E_3)} \\
\end{align*}
\] (159)

These terms are not unknown to physics. Let us rewrite them using dot product notation, to make it pop out more:

\[
\begin{align*}
- \sqrt{-||B||^2 + ||E||^2 - 2iB \cdot E} \\
\sqrt{-||B||^2 + ||E||^2 - 2iB \cdot E} \\
- \sqrt{-||B||^2 + ||E||^2 + 2iB \cdot E} \\
\sqrt{-||B||^2 + ||E||^2 + 2iB \cdot E} 
\end{align*}
\] (160) (161) (162) (163)

The term \(-||B||^2 + ||E||^2\) and the term \(B \cdot E\) are simply the two fundamental Lorentz invariants of electromagnetism, here grouped as a complex number.

**Definition 15** (Electromagnetism Interval). The electromagnetism interval is given by applying the universal interval to a general bivector of the \(\mathbb{C}l_4(C)\) algebra, as given by Equation 155.
Using this norm, we now have the opportunity to treat both electromagnetism and general relativity as mere features of a single universal interval. Unlike other proposals in the literature, we note the benefit that here we do not need to add additional dimensions beyond 3+1 spacetime to achieve it.

So, what about the other forces — are they in the universal interval? Before we can answer that, let us first show the difficulty.

6. complete multivector: The matrix representation of a complete multivector such as:

\[
\hat{s} = \begin{pmatrix}
  a + X_0 e_0 + X_1 e_1 + X_2 e_2 + X_3 e_3 \\
  B_2 - i B_1 - V_2 - i V_1 \\
  -B_2 - i B_1 - V_2 - i V_1 \\
  -i B_2 + X_3 - i V_3
\end{pmatrix}
\]

is:

\[
\hat{s} = \begin{pmatrix}
  a + X_0 - i B_3 - i V_3 & B_2 - i B_1 + V_2 - i V_1 & -ib + X_3 + E_3 - i V_0 & X_1 - i X_2 + E_1 - i E_2 \\
  -B_2 - i B_1 - V_2 - i V_1 & a + X_0 + i B_3 + i V_3 & X_1 + i X_2 + E_1 + i E_2 & -ib - X_3 - E_3 - i V_0 \\
  -ib - X_3 + E_3 + i V_0 & -X_1 + i X_2 + E_1 - i E_2 & a - X_0 - i B_3 - i V_3 & B_2 - i B_1 - V_2 + i V_1 \\
  -X_1 - i X_2 + E_1 + i E_2 & -ib + X_3 - E_3 + i V_0 & -B_2 - i B_1 + V_2 + i V_1 & a - X_0 + i B_3 - i V_3
\end{pmatrix}
\]

We note that the characteristic polynomial required to solve this matrix for its eigenvalues is of degree 4. Although an arbitrary quartic polynomial contains an exact algebraic solution, finding a simplified form of this expression is challenging. To illustrate the difficulty, we attempted a symbolic computation using Mathematica:

\[
y_0 = \text{KroneckerProduct}[\text{PauliMatrix}[3], \text{IdentityMatrix}[2]]; \\
y_1 = \text{KroneckerProduct}[	ext{I PauliMatrix}[2], \text{PauliMatrix}[1]]; \\
y_2 = \text{KroneckerProduct}[	ext{I PauliMatrix}[2], \text{PauliMatrix}[2]]; \\
y_3 = \text{KroneckerProduct}[	ext{I PauliMatrix}[2], \text{PauliMatrix}[3]]; \\
o01 = 1/2 (y_0 \cdot y_1 \cdot y_3 \cdot y_0); \\
o02 = 1/2 (y_0 \cdot y_2 \cdot y_2 \cdot y_0); \\
o03 = 1/2 (y_0 \cdot y_3 \cdot y_3 \cdot y_0); \\
o12 = 1/2 (y_1 \cdot y_2 \cdot y_2 \cdot y_1); \\
o31 = 1/2 (y_3 \cdot y_1 \cdot y_1 \cdot y_3); \\
o23 = 1/2 (y_2 \cdot y_3 \cdot y_3 \cdot y_2); \\
y_5 = \text{I} \ y_0 \cdot y_1 \cdot y_2 \cdot y_3; \\
y_0 = y_5 \cdot y_0;
\]
\begin{verbatim}
v1 = y5.y1;
v2 = y5.y2;
v3 = y5.y3;
M = a \text{IdentityMatrix}[4] + X0 y0 + X1 y1 + X2 y2 + X3 y3 + E1 o01 + E2 o02 + E3 o03 + B3 o12 + B2 o31 + B1 o23 + V0 v0 + V1 v1 + V2 v2 + V3 v3 + b y5; M // MatrixForm

S1 = Part[Eigenvalues[M], 1]
S2 = Part[Eigenvalues[M], 2]
S3 = Part[Eigenvalues[M], 3]
S4 = Part[Eigenvalues[M], 4]
ToRadicals[S1]
Simplify[%]
\end{verbatim}

The difficulty is that \text{Part[Eigenvalues[M],1]} produces a \text{Root[\ldots]} expression (specifically, a general polynomial equation of degree 4) whose \text{LeafCount} is 1703, and \text{ToRadicals[S]} produces an algebraic expression with a \text{LeafCount} of 52007. Using \text{Simplify[\ldots]} on an expression with such a high \text{LeafCount} far exceeds the computing resources I have at my disposal. The solution contains thousands and thousands of terms. Even then, the simplified result produced by Mathematica is not guaranteed to yield an easily understandable result.

Due to its size, the universal interval of a universal event has been pushed to Annex A.

Consequently, the universal norm embeds quite a lot of complexity. Upon manual inspection, many of the terms appears as related to unitary groups up to \text{U}(3) and to interference between such terms, and therefore *may* embed a theory of particle physics. However, before we attack this complexity manually (using insight instead of raw computation), we must develop a better understanding the universal interval.

Let me here state that there is an alternative way to define the universal interval independently of its matrix representation, that is possibly even better. However I was not able to work out the \text{Cl}_{4}(\mathbb{C}) case generally and thus I elected to use the matrix representation, which is friendlier to symbolic computations (using the tools/software I am familiar with). To achieve this representation-free definition, it suffices to construct a polynomial using the geometric product and other algebraic and involutive operations to eliminate all basis elements. The resulting polynomial is a characteristic polynomial associated to the multivector and is analogous to the matrix characteristic polyno-
mials. Let me give a few simple examples on how to construct such a
characteristic polynomial from a multivector.

7. In the case of a 1-vector of $\text{Cl}_3(\mathbb{C})$ one can find the characteristic
polynomial easily by simply taking the geometric product once:

$$ v := x\sigma_x + y\sigma_y + z\sigma_z $$

$$ \implies v^2 = x^2 + y^2 + z^2 $$

where $v^2$ is the geometric product. The roots of $v^2$ are scalars and
corresponds to the eigenvalue of the matrix representation of $v$.

8. Sometimes the polynomial contains both elements of degree two
and of degree one:

$$ v := a + x\sigma_x $$

$$ \implies v^2 = a^2 + x^2 + 2ax\sigma_x $$

$$ = a^2 + x^2 + 2a(x\sigma_x + a) - 2a^2 $$

$$ = x^2 - a^2 + 2av $$

$$ \implies v^2 - 2av = x^2 - a^2 $$

The roots on this equation are scalars and are analogous to the
eigenvalues of its matrix representation.

9. Finally, in the case of a universal event of $\text{Cl}_4(\mathbb{C})$, a polynomial of
degree 4 would be required to produce 4 roots (since the matrix
representation has 4 eigenvalues).

Let us now investigate some intervals that are "outside the norm".

10. Euclid/taxicab norm: Let us consider the universal interval of the
following multivector: $u := a + b + c + \sigma_x x + \sigma_y y + \sigma_z z$ (of the
$\text{Cl}_3(\mathbb{R})$ algebra). It has the following matrix representation:

$$ \hat{u} = \begin{pmatrix} z + a + b + c & x - iy \\ x + iy & -z + a + b + c \end{pmatrix} $$

Evaluating the universal interval, we obtain this relation:

$$ 0 = (\Delta + \Delta b + \Delta c)^2 - 2(\Delta a + \Delta b + \Delta c)\Delta s + (\Delta s)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 $$

Solving for the roots, we get:
\[ \Delta s = \Delta a + \Delta b + \Delta c \pm \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \]  

Here the departure from familiar norms is quite remarkable: we have taxi-cab contributions to the interval by the scalar elements. To understand why the scalars find themselves out of the square root, 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 apples, and the y-axis denotes the number of oranges. Say one wishes to procure 3 apples and 4 oranges. Question: can one appeal to the Pythagorean theorem to negotiate down the price of 3 apples and 4 oranges for the low price of 5 fruits? 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 \(d = \sqrt{(\Delta x)^2 + (\Delta y)^2}\).

The universal interval is able to construct a norm which is a combination of taxi-cab terms and Euclidean terms, as appropriate for the situation.

11. Let us now apply a similar logic as the Euclid/taxicab norm, but for the complex numbers. Here we will use the representation-free procedure to eliminate the geometric element \(I\) and to find the interval:

\[ z : = a + bI \]  \hspace{1cm} (176)  
\[ z^2 = a^2 - b^2 + 2abI \]  \hspace{1cm} (177)  
\[ = a^2 - b^2 + 2a(a + bI) - 2a^2 \]  \hspace{1cm} (178)  
\[ = -a^2 - b^2 + 2az \]  \hspace{1cm} (179)
Consequently the universal interval of a complex number is:

$$(\Delta s)^2 - 2\Delta a\Delta s + (\Delta a)^2 + (\Delta b)^2 = 0$$  \hspace{1cm} (180)

The roots are:

$$\{\Delta a + \Delta b I, \Delta a - \Delta b I\}$$  \hspace{1cm} (181)

Why are we not getting the usual complex norm $\sqrt{(\Delta a)^2 + (\Delta b)^2}$ as the universal interval for the complex numbers? To understand the discrepancy, consider the following purely hypothetical scenarios:

- Say someone asks your help to find a lost friend. If they say, "we walked 3 km east, then 4 km north, then I lost him", then clearly the last known location is 5 km away (Euclidean norm).

- 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 complex norm, for say $a + ib$ is given by $\sqrt{a^2 + b^2}$. Then blindly believing this equation, you realize there is a faster way to the lost friend. You can simply walk 4 km in a straight line to get to the person’s last known position, saving yourself a total of 2 km by skipping the portal. Will you reach the lost friend’s last known location faster than the another person who rushed to the portal then walk towards the endpoint? Well, if one believes the norm of a complex number to be $|z| = \sqrt{a^2 + b^2}$ 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 complex taxi-cab norm such as $a \pm b I$, then one concludes that one must enter the portal to get there. Specifically, with the tax-cab norm for complex numbers, one will have to walk 3 km on the real line and 4 km on the imaginary line for a total of $(3+4i)$ km (not 5) to get to the correct position.

The take home message is that the universal interval does not give the norm of a complex number as measured within the complex plane (as $\sqrt{a^2 + b^2}$ does) but as measured within physical space. Indeed, since position/movement in real space presumably does not
influence position/movement in imaginary space, the norm retains
the imaginary component and relates it to the real component via
a complex taxi-cab norm.

12. **quaternions**: In $Cl_3(R)$, the event $q := a + b\sigma_{yz} + c\sigma_{xz} + d\sigma_{xy}$ is:

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

(182)

The universal interval of the quaternion is:

$$(\Delta a)^2 + (\Delta b)^2 + (\Delta c)^2 + (\Delta d)^2 - 2\Delta a\Delta s + (\Delta s)^2 = 0
$$

(183)

And the roots are:

$$
\Delta s = \Delta a \pm i\sqrt{(\Delta b)^2 + (\Delta b)^2 + (\Delta c)^2}
$$

(184)

Once one enters the quaternionic portal, one is free to use the
Pythagorean theorem to move around.

We may wonder, is there nonetheless a way to get $\sqrt{a^2 + b^2}$ for
complex numbers, and $\sqrt{a^2 + b^2 + c^2 + d^2}$ for quaternions? Interestingly, a universal norm can be constructed using similar techniques.
The universal norm of a multivector, unlike its universal interval,
maps to a real instead of a complex number, and even recovers and
extends the concept of the complex norm.

**Definition 16** (Universal Norm). *Let $\Delta s$ be the matrix representation of $\Delta s \in G_n$. Then, the universal norm $\|\hat{s}\| \in R$ for a given matrix representation of $s$ with dimension $k^2$ is given by this relation:*

$$
\|\hat{s}\| = \sqrt[\hat{s}]{\det \Delta s}
$$

(185)

Let us give a few examples.

13. **complex number**: In $Cl_2(R)$, the event $z := a + bI$ has the following matrix representation:

$$
z := \begin{pmatrix}
a & b \\
b & a
\end{pmatrix}
$$

(186)

The universal norm is:
14. **quaternions:** In $\text{Cl}_3(\mathbb{R})$, the event $q := a + b\sigma_{yz} + c\sigma_{xz} + d\sigma_{xy}$ is:

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

(188)

The universal norm of the quaternion is:

$$\|q\| = \sqrt{\det \begin{pmatrix} a + ib & ic + d \\ ic - d & a - ib \end{pmatrix}} = \sqrt{a^2 + b^2 + c^2 + d^2}$$

(189)

15. **Electromagnetism norm:** As shown earlier, the matrix representation of the event $b := E_1\gamma_0\gamma_1 + E_2\gamma_0\gamma_2 + E_3\gamma_0\gamma_3 + B_1\gamma_2\gamma_3 + B_2\gamma_1\gamma_3 + B_3\gamma_1\gamma_2$ has the following eigenvalues:

$$-\sqrt{-\|B\|^2 + \|E\|^2 - 2iB \cdot E}$$

(190)

$$\sqrt{-\|B\|^2 + \|E\|^2 - 2iB \cdot E}$$

(191)

$$-\sqrt{-\|B\|^2 + \|E\|^2 + 2iB \cdot E}$$

(192)

$$\sqrt{-\|B\|^2 + \|E\|^2 + 2iB \cdot E}$$

(193)

**Definition 17** (Electromagnetism norm). *The norm of electromagnetism is given by multiplying the eigenvalues (to obtain the determinant):*

$$\|s\|^2 = (\|B\|^2 + \|E\|^2)^2 + 4(B \cdot E)^2$$

(194)

It groups the imaginary and the real part of the eigenvalue of the interval of electromagnetism using the complex norm.

16. **complete multivector:** Due to its size, the universal norm of a universal event has been pushed to Annex A.

**Result 1** (Dimensionality of spacetime limited to 4). *We note that the characteristic polynomial of the matrix representation of a universal event of $\text{Cl}_4(\mathbb{C})$ is of degree 4. Algebraic solutions (in terms of additions, subtraction, multiplication, division and exponentiation by rational numbers) exists*
for all polynomials of degrees 4 or lower, however, as per the Abel–Ruffini theorem, an algebraic solution does not exist for polynomial of degrees 5 or higher. Consequently, an algebraic solution of a universal metric is only possible for the complete basis of a geometric algebras of dimensions 4 or lower. This places the upper bound for a universal metric to no more than 4. Beyond four dimensions, no general expression for the universal metric is possible, and the algebraic form of each metric is (potentially) specific to each numerical value of the entry of the matrix.

Let us now turn our attention towards expressing the universal norm as a metric. When I showed to a few mathematicians that I had a new ‘cool’ metric and that it is a polynomial of degree 4, the immediate reaction was skeptical. But it turns out that a polynomial metric of degree four can be made to follow all the axioms of norms (or pseudo-norms), metric (or pseudo-metric) and distance functions (or pseudo distance functions). There is an already known group of obscure norms, called p-norms, extending Euclidean norms to degrees p. I think the most evidently valid degree four metric is simply this one:

\[(d_s)^4 = (dx)^4 + (dy)^4 + 2(dx)^2(dy)^2\]  \hspace{1cm} (195)

What is this unusual degree 4 metric? Well, if one takes the square root on each sides, one gets \[(d_s)^2 = (dx)^2 + (dy)^2\] which is the plain old simple Euclidean metric. So at least one metric of degree four is perfectly fine. Let us now consider more interesting options.

17. Sticking with degree two but adding a bit of degree one spice, for example we could have:

\[(d_s)^2 = 2 da ds - (da)^2 + (db)^2\]  \hspace{1cm} (196)

Which, if we solve as a polynomial, we get two roots for the metric as:

\[ds = da \pm i db\]  \hspace{1cm} (197)

which assigns the norm of a complex number to either a complex taxi-cab norm or its conjugate.

18. Let us now state a metric of degree 4 that does not have a representation as degree two:
(\text{ds})^4 = (\text{dx})^4 + (\text{dy})^4 \quad (198)

And another one:

(\text{ds})^4 = (\text{dx})^4 + (\text{dx})^3 \, \text{dy} \quad (199)

19. A general metric of degree 4 is:

(\text{ds})^4 = g_{\mu\nu\alpha\beta} \, \text{d}X^\mu \, \text{d}X^\nu \, \text{d}X^\alpha \, \text{d}X^\beta \quad (200)

where $g_{\mu\nu\alpha\beta}$ is a $4 \times 4 \times 4 \times 4$ tensor with the usual symmetry requirement and is defined in the usual sense as a function which takes two multivectors from the tangent space of a manifold and outputs a scalar. Because it is of degree 4, the metric has extra degrees of freedom compared to a usual degree 2 metric.

In the following section, we will introduce universal thermodynamics as the framework able to tackle a physical system undergoing arbitrary changes described, and we will use the universal metric to do so.

5.4 Universal Ensemble

We are now ready to produce a microscopic description of universal thermodynamics by using the full flexibility of geometric algebra, including that of multivectors. To facilitate the discussion, let us adopt the following naming convention. Let $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ be thermodynamic constraints:

1. **Entropic space**, for the quantities $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ along with basis $\sigma_x, \sigma_y, \sigma_z$ or basis $\gamma_1, \gamma_2, \gamma_3$.

2. **Entropic time**, for the quantity $\mathbf{X}_0$ along with basis $\gamma_0$.

3. **Entropic space-time**, for the group comprising the quantities $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ along with basis $\gamma_0, \gamma_1, \gamma_2, \gamma_3$.

4. **Entropic geometry**, for any group comprising higher dimensional extensive quantities (such as areas, volumes, etc), as always expressed in geometric algebra.

A typical thermodynamic system is defined by a set of constraints $\mathbf{E}, \mathbf{V}, \ldots$. Changes in the state of the system are governed by an equation of state $\text{d}S = T \, \text{d}\mathbf{E} + p \, \text{d}\mathbf{V}$:
If interpreted as a metric, the equation of state of such a system is a taxi-cab metric because the constraints are scalar: \((dS)^2 = (T \, dE)^2 + (p \, dV)^2 + (2T_p \, dE \, dV)\).

In universal thermodynamics, we wish to define a thermodynamic system able to describe a cycle/transformation between states that are described by general matrices, just as:

\[
\text{state-a} \rightarrow \text{state-b}
\]

\[
(a_{00} \ldots a_{0n}) \rightarrow (b_{00} \ldots b_{0n})
\]

As stated earlier, for \(n \geq 4\) there are no algebraic universal metric, but specifically in the case \(n = 4\) the equation of state can be expressed equivalently as a universal metric, giving the system a purely geometric interpretation.

Using the definition of nature \(\mathcal{N}\), as the set of a maps between manifests and transformations, as the constraints, we can define a partition function of the probability measure that maximizes the entropy subject to said constraints:

**Definition 18 (Universal partition function).** The partition function associated to the probability measure that maximizes the entropy of the space of all possible function \(g \in \mathcal{N}\) is a system of functional integrals that maximizes the entropy for each eigenvalues of the universal metric. Specifically in 4 dimensions the system contains 4 equations. Let \(\{g_1, g_2, g_3, g_4\}\) be the eigenvalues of the universal metric, then the universal partition function (in 4 dimensions) is the following system of equations:

\[
\hat{Z} := \left(\begin{array}{cccc}
Z_1 & 0 & 0 & 0 \\
0 & Z_2 & 0 & 0 \\
0 & 0 & Z_3 & 0 \\
0 & 0 & 0 & Z_4
\end{array}\right) = \left(\begin{array}{cccc}
\int D[g_1] \exp(-\hat{k}S[g_1]) & 0 & 0 & 0 \\
0 & \int D[g_1] \exp(-\hat{k}S[g_1]) & 0 & 0 \\
0 & 0 & \int D[g_1] \exp(-\hat{k}S[g_1]) & 0 \\
0 & 0 & 0 & \int D[g_1] \exp(-\hat{k}S[g_1])
\end{array}\right)
\]

**Definition 19 (Universal equation of state).** The entropy associated to the ensemble with partition function \(\hat{Z}\) is:
Let us now write this system as a norm by multiplying the eigenfunction over the eigenvalues:

\[
\begin{pmatrix}
S_1 & 0 & 0 & 0 \\
0 & S_2 & 0 & 0 \\
0 & 0 & S_3 & 0 \\
0 & 0 & 0 & S_4
\end{pmatrix}
= \begin{pmatrix}
(k_B \langle \bar{k} \rangle(S_{[g_1]} + \ln Z_1)) & 0 & 0 & 0 \\
0 & (k_B \langle \bar{k} \rangle(S_{[g_2]} + \ln Z_2)) & 0 & 0 \\
0 & 0 & (k_B \langle \bar{k} \rangle(S_{[g_3]} + \ln Z_3)) & 0 \\
0 & 0 & 0 & (k_B \langle \bar{k} \rangle(S_{[g_4]} + \ln Z_4))
\end{pmatrix}
\]

And the equation of state is its total derivative:

\[
\begin{pmatrix}
dS_1 & 0 & 0 & 0 \\
dS_2 & 0 & 0 & 0 \\
dS_3 & 0 & 0 & 0 \\
dS_4 & 0 & 0 & 0
\end{pmatrix}
= \begin{pmatrix}
k_B \langle \bar{k} \rangle \text{d}S_{[g_1]} \\
0 & k_B \langle \bar{k} \rangle \text{d}S_{[g_2]} \\
0 & 0 & k_B \langle \bar{k} \rangle \text{d}S_{[g_3]} \\
0 & 0 & 0 & k_B \langle \bar{k} \rangle \text{d}S_{[g_4]}
\end{pmatrix}
\]

Finally, one can structure the equation of state as a norm by multiplying the eigenvalues:

\[
dS_1 \, dS_2 \, dS_3 \, dS_4 = (k_B \langle \bar{k} \rangle)^4 \text{d}S_{[g_1]} \, \text{d}S_{[g_2]} \, \text{d}S_{[g_3]} \, \text{d}S_{[g_4]}
\]

Let us start with a simple example. Assume a 1-vector valued function over \( Cl_4(\mathbb{C}) \) such as \( g := X_0 \gamma_0 + X_1 \gamma_1 + X_2 \gamma_2 + X_3 \gamma_3 \). Then the system of equations of state is:

\[
dS = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = k_B \langle \bar{k} \rangle \text{d}S_{\sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2}} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

We will rename \( \sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2} \) to \( \sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2} \). Let us now write this system as a norm by multiplying the eigenvalues:

\[
(-dS)^2 (dS)^2 = (-1)(-1)(1) \left( k_B \langle \bar{k} \rangle \text{d}S_{\sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2}} \right)^4
\]

Then by simplifying:

\[
dS = \pm k_B \langle \bar{k} \rangle \text{d}S_{\sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2}} = \pm k_B \langle \bar{k} \rangle \frac{1}{2\sqrt{X_0^2 - X_1^2 - X_2^2 - X_3^2}} (2X_0 \, dX_0 - 2X_1 \, dX_1 - 2X_2 \, dX_2 - 2X_3 \, dX_3)
\]
By noticing that the square root is simply the entropy (within a constant) \( S \propto k_B \sqrt{\sum \left( X_i^2 - X_j^2 \right)} \), we can rewrite the equation as:

\[
S \, dS = \pm (k_B \tilde{k})^2 (X_0 \, dX_0 - X_1 \, dX_1 - X_2 \, dX_2 - X_3 \, dX_3)
\]  

(211)

In normal statistical physics, one simply needs to take the total derivative once and one gets the equation of state. However, in universal statistical physics one must take the total derivative over and over until all elements are infinitesimals. With 4 eigenvalues, one may need to take the total derivative up to 4 times. In the present example however, we will only need to take it once more. We note that we take the derivative of an infinitesimal to be 0.

\[
d(S \, dS) = \pm (k_B \tilde{k})^2 (dX_0 \, dX_0 - dX_1 \, dX_1 - X_2 \, dX_2 - X_3 \, dX_3)
\]  

(212)

\[
(dS)^2 + d(dS) = \pm (k_B \tilde{k})^2 \left( (dX_0)^2 + dX_0 \, d(dX_0) - (dX_1)^2 + dX_1 \, d(dX_1) - \ldots \right)
\]  

(213)

\[
\pm (k_B \tilde{k})^{-2} (dS)^2 = (dX_0)^2 + (dX_1)^2 + (dX_2)^2 + (dX_3)^2
\]  

(214)

which is the interval of special relativity, here expressed in terms of entropy. The above derivation can be repeated with the universal interval, and we will investigate a few more examples in the results section. We now generalize the example to the following results:

**Result 2** (Entropic interval). We are already familiar with the relationship between \( ds = c \, d\tau \) connecting the interval to the proper time. Now, we have identified that the interval of space-time connects to its entropy as follows:

\[
(ds)^4 = (k_B \tilde{k})^{-4} (dS)^4
\]  

(215)

We can further connect the entropy directly to the proper time as follows:

\[
(c \, d\tau)^4 = (k_B \tilde{k})^{-4} (dS)^4
\]  

(216)

Consequently, we identified that any change in proper time is accompanied by a change in entropy.

Then finally, this produces an arrow of time; the ’true’ arrow of time:

**Result 3** (Arrow of time). Any statistical system of physics will evolve towards the direction of increased entropy as fluctuations occurs. Space-time is no different. Since the interval and the entropy are related, then the entropy of future events is such that the system is powered to evolve towards its own future.
Part III

Physics

6 Results (Spacetime Entropy)

6.1 Law of Inertia

We will now attribute an exact expression for the specific repetency, in terms of existing physical constants. As one may recall, in usual statistical physics the Lagrange multiplier $\beta$ is eventually associated with 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):

$$\frac{\partial S}{\partial E} = \beta k_B$$

(217)

is equivalent to the following equation (of thermodynamics):

$$\frac{\partial S}{\partial E} = T^{-1}$$

(218)

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

To find an expression for $\tilde{k}$, 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 $\tilde{k}$ such that the two are equated. Let us begin with the 1-vector:

$$g = X_1 \gamma_1 + X_2 \gamma_2 + X_3 \gamma_3$$

(219)

The equation of state of this metric is:

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

(220)

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 metric. We recall the definition of an entropic force:

$$dS = \frac{F}{T} \, dx$$

(221)
As a specific example of an entropic force, we can think of the tension within the chain of a polymer in a warm bath. With our equation of state, we are not very far from the general definition of an entropic force. In fact, the units of $\tilde{k}$ are the same as those of $F/T$.

So which value of $F$ and $T$ to use? The natural choices, proposed by Erik Verlinde\(^5\) are to take $T$ as the Unruh temperature and $F$ as the law of inertia. Let us solve for $\tilde{k}$ using:

\[
T_{\text{Unruh}} = \frac{\hbar a}{2\pi ck_B} \quad (222)
\]

\[
F = ma \quad (223)
\]

Then:

\[
k_B \tilde{k} = FT^{-1} \quad (224)
\]

\[
= ma \frac{2\pi ck_B}{\hbar a} \quad (225)
\]

\[
= 2\pi k_B \frac{mc}{\hbar} \quad (226)
\]

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.

**Result 4** (Expression for the specific repetency).

\[
\tilde{k} = 2\pi \frac{mc}{\hbar} \quad (227)
\]

6.2 **Beckenstein-Hawking Entropy**

Let us now explore another means to derive/confirm the expression for $\tilde{k}$. Starting from the entropy of inertia, we now consider that the distance is fixed to the Schwarzschild radius $2Gm/c^2$.

\[
S = 2\pi k_B \frac{mc}{\hbar} \sqrt{(X_1)^2 + (X_2)^2 + (X_3)^2} \quad (228)
\]

\[
= 2\pi k_B \frac{mc}{\hbar} \left( \frac{2Gm}{c^2} \right) \quad (229)
\]

\[
= 4\pi k_B \frac{Gm^2}{hc} \quad (230)
\]

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

\[
A = 4\pi r^2 = 4\pi \left( \frac{2Gm}{c^2} \right)^2 \Rightarrow m^2 = A \frac{c^4}{16\pi G^2} \quad (231)
\]
Replacing \( m^2 \) in our result for \( S \), we get:

\[
S = 4\pi k_B^2 \frac{G}{\hbar c} A \frac{c^4}{G^2 16\pi} \\
= k_B \frac{1}{4} \frac{c^3}{\hbar G} A
\]

which is the Bekenstein-Hawking entropy.

6.3 Action

We have established in an earlier result that the arrow of time is powered by entropy and points towards the future of the observer, but we only worked out the example for a flat spacetime. In the case of generally curved space-times, the direction of motion in space-time is the geodesic. In this case, the observer experiences entropic forces along their paths in space-time which "tilt" the direction of maximal entropy.

The equation of state corresponding to the movement of a test particle in a gravitational field is:

\[
dS = \pm 2\pi k_B \frac{mc}{\hbar} \sqrt{g_{\mu\nu}} \, dX^\mu \, dX^\nu
\]

Let us parametrize the metric equation of state over a path \( \tau \) and then integrate:

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

We rearrange as follows:

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

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

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

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

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

and the corresponding equations of motion are:
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, system evolution along the geodesic 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).

**Result 5.** The action relates to a change of entropy as follows:

$$S \equiv \pm \frac{\hbar}{2\pi k_B} \int \frac{\partial}{\partial \tau} S[\tau] \, d\tau$$

(240)

### 6.4 Fermi-Dirac statistics of events

We now attack the microscopic description. 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.

We start with:

$$dS = \pm k_B \hbar \sqrt{(dX_0)^2 - (dX_1)^2 - (dX_2)^2 - (dX_3)^2}$$

(241)

But, for simplicity, we will consider the 1+1 space-time case:

$$dS = \pm k_B \hbar \sqrt{(dX_0)^2 - (dX_1)^2}$$

(242)

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. Consequently, the Fermi-Dirac distributions for this equation of state are:

$$\langle n \rangle_{\text{future}} = \frac{1}{\exp \left( \frac{k_B \hbar \sqrt{X_0^2 - X_1^2}}{k_B \hbar} \right) + 1}$$

(243)

$$\langle n \rangle_{\text{past}} = \frac{1}{\exp \left( \frac{-k_B \hbar \sqrt{X_0^2 - X_1^2}}{k_B \hbar} \right) + 1}$$

(244)

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:
\[
\langle n \rangle = \begin{cases} 
\frac{1}{\exp(-k_B\sqrt{\lambda_0^2-x_1^2})+1} & \lambda_0 < 0 \\
\frac{1}{2} & \lambda_0 = 0 \\
\frac{1}{\exp(k_B\sqrt{\lambda_0^2-x_1^2})+1} & \lambda_0 > 0 
\end{cases} 
\] (245)

\(\langle n \rangle\) is shown in Figure (5) using a 2-dimensional 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 reoccur). Interestingly, the occupancy probability of events nonetheless describes a probability space (the occupancy rate varies between 0% to 100% within the light-cone). 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 5a and 5b). 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 5c). The observer rides the tide. 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. We will now explore the significance in the next section.

6.5 Quantum Mechanics

The probability of occupancy of an event is obtained by applying the Fermi-Dirac statistics to the ensemble. 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.

The white region outside the light cone described in Figure 5a), 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:

$$\sqrt{c^2(\Delta t)^2 - (\Delta x)^2}_{\Delta t=0, \Delta x=5} = \sqrt{c^2(0)^2 - 5^2} = i5$$  \hspace{1cm} (246)

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

Consider the partition function corresponding to the metric $g = X_0 e_0 + X_1 e_1 + X_2 e_2 + X_3 e_3$. Using the universal metric method, we associate the following partition function to the eigenvalues:

$$Z = \int D[g] \exp\left(\pm k \int_{\tau} \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} \, d\tau\right)$$  \hspace{1cm} (247)

In the general case, a path in space-time may cross the light-cone boundary of the observer. For example look at of Figure 5a) then imagine a path $\tau$ that may start in the white region, enter the red zone then end at the observer. The part of this path that are in the white region will acquire the imaginary term $i$ in the action whereas
the parts that are in the red zone will not. Paths of this nature are shown on Figure 6. To account for the entire path, 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.

Figure 6: a) We investigate the path integral from P to O entirely in a space-like region with complex occupancy. As an approximation, we consider P and O to be so far away from the observer that we can consider all paths to be space-like separated from O (≈ we neglect the contribution of any paths which may cross into the time-like regions). In this case, we obtain the usual Feynman path integral. b) We investigate a path from R to O close to the boundary of the light cone of O. The path starts as space-like separated and enters within the light cone of O as time-like separated. In this case, the path begins as a quantum system capable of interference, and decoherence/loss-of-interference is experienced as it enters the light cone.

\[
Z = \int D[g] \exp \left( \pm \text{Re}\left[ \frac{1}{\sqrt{-g}} \int \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \, d\tau \right] \right) \pm i \text{Im}\left[ \frac{1}{\sqrt{-g}} \int \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \, d\tau \right]
\]

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:

\[
Z = \int D[g] \exp \left[ \pm i \text{Im}\left[ \frac{1}{\sqrt{-g}} \int \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \, d\tau \right] \right]
\]

Furthermore, when O describes purely time-like paths, it is the imaginary part that is eliminated:

\[
Z = \int D[g] \exp \left[ \pm \text{Re}\left[ \frac{1}{\sqrt{-g}} \int \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \, d\tau \right] \right]
\]
6.6 Decoherence at the Time-like/Space-like Boundary

Let us investigate the role of each part of the path in more details 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) part and a space-like (imaginary) part. 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 more details, it suffices to replace the Lagrange multipliers $\tilde{k}$ by the previously obtained coefficient $2\pi mc/\hbar$. Let $A_E[q]$ be the time-like part (the energy part) of the functional integral and let $A_S[q]$ be its space-like part (the action part):

$$Z = \int D[g] \exp (A_S[g]) \exp (A_E[g])$$

where the space-like part is:

$$A_S[g] = \frac{i}{\hbar} \text{Im} \left[ -2\pi mc \int \sqrt{g} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau} \frac{d\tau}{L} \text{Lagrangian} \right]$$

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 coeffi-
cient with $a/a = 1$ (and with $a \neq 0$), then we get:

$$A_{E}[q] = -\frac{2\pi c}{\hbar a} \text{Re} \left[ ma \int_{\mathcal{O}} \sqrt{g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \tau}} \text{Lagrangian} \right]$$

(253)

Energy

Thermal states: $-\beta E[q]$

where the term $2\pi c/\hbar a$ is the Unruh temperature.

A similar process would occur at any horizons, even in the case of curved spacetimes. From $\mathcal{O}$’s point of view, the statistical weight of paths inside the horizon is described by complex amplitudes in a 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. In curved spacetimes, this temperature is the Hawking temperature. Specifically, if we replace $a \rightarrow c^4/(4\pi G)$ we get:

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

(254)

Thus, a 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 7.4). 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. We call this an unprepared quantum system. However, this need not be the case for an unaccelerated observer advancing into the future and capturing a larger sector of the space-like region within its light cone over time. In flat space-time, the space-like region is not hidden by a horizon, thus information from the region can eventually enter the light cone of the observer.

A prepared quantum system has additional observables $O_i[q]$ each with expectation value $\langle O_i \rangle$ but unlike for its unprepared counterpart, the probability measure over the paths is maximized under the added constraints of the observables’ expectation values. In such case, information about the preparation will be available to the observer under repeated measurements of multiple copy of the quantum system. Outside the light cone, preparation information is given in the forms of expectation values by the usual quantum equation:
\[ \langle O_i \rangle = \frac{\int D[g] O_i[g] \exp \left[ \frac{i}{\hbar} S[g] \right]}{\int D[g] \exp \left[ \frac{i}{\hbar} S[g] \right]} \]  

(255)

and inside the light, the same information is available but as ‘information-bearing thermal’ radiation / decohered quantum states:

\[ \langle O_i \rangle = \frac{\int D[g] O_i[g] \exp \left[ -\beta E[g] \right]}{\int D[g] \exp \left[ -\beta E[g] \right]} \]  

(256)

In the first case, and using the Von Neumann formalism, preparation information is available in the form of pure states, but in the second decoherent case, it is available as a mixture, such as:

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

(257)

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

6.7 Quantum Field Gravity

Let us take a multivector comprised only of the pseudoscalar basis of \( \mathbb{C}l_4(\mathbb{C}) \):

\[ g = R[X_0, X_1, X_2, X_3] e_0 \wedge e_1 \wedge e_2 \wedge e_3 \]  

(258)

This represents the 4-volume element. We can rewrite \( g \) using the basis \( \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\} \) instead of \( \{e_0, e_1, e_2, e_3\} \), and we get:

\[ g = R[X_0, X_1, X_2, X_3] \sqrt{|g|} \gamma_0 \gamma_1 \gamma_2 \gamma_3 \]  

(259)

Using the universal metric, we obtain the eigenvalues as the roots of this metric polynomial:

\[ s^4 = (R[X_0, X_1, X_2, X_3] \sqrt{|g|})^4 \]  

(260)

\[ s = \pm R[X_0, X_1, X_2, X_3] \sqrt{|g|} \]  

(261)
When we insert the eigenvalues into the universal partition function as functionals, we get:

$$Z = \int D[g] \exp\left( \pm \hbar \int R[X_0, X_1, X_2, X_3] \sqrt{|g|} \, dX_0 \, dX_1 \, dX_2 \, dX_3 \right)$$  \hspace{1cm} (262)$$

Here, the exponent is the Hilbert-Einstein action. We recall that $R[X_0, X_1, X_2, X_3]$ is a complex number, and consequently its imaginary part is the well known path integral over the possible metrics of the gravitational fields. The presence of the real as well as the imaginary part are very important to the universal formulation; they are responsible for decoherence at the surface of horizons.

7 Results (Cosmology)

Having a system of equations for our partition function means that the geometry can be entirely emergent from the entropy of the system as an equation of state. Here, we will study this concept into greater detail by constructing an entropy model of $\Lambda$CDM.

7.1 de Sitter Space

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

$$a^2 = -c^2 X_0^2 + X_1^2 + X_2^2 + X_3^2$$  \hspace{1cm} (263)$$

with a cosmological horizon at $r = a$. Let us now compare it to the 1-vector metric entropy:

$$S = k_B \ln Z + k_B \tilde{k} \sqrt{-X_0^2 + X_1^2 + X_2^2 + X_3^2}$$  \hspace{1cm} (264)$$

First, let us consider the ground state of the system by posing $S = 0$:

$$k_B \ln Z = -k_B \tilde{k} \sqrt{-X_0^2 + X_1^2 + X_2^2 + X_3^2}$$  \hspace{1cm} (265)$$

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:

$$k_B \ln Z = \pm \frac{2\pi k_B mc}{\hbar} \sqrt{-X_0^2 + X_1^2 + X_2^2 + X_3^2}$$  \hspace{1cm} (266)$$
Since $Z$ is a constant, we recover the same mathematical form as the hyperboloid equation in 3+1 space-time characteristic of de Sitter space, except that now that the quantities on each side of the expression have the units of the entropy. The entropy of de Sitter space, where $\alpha$ is the radius to the horizon, is therefore, in the ground state, quantified by the relation:

$$k_B \ln Z = \pm 2\pi k_B \frac{mc}{\hbar} \alpha$$  \hspace{2cm} (267)

Noting that in de Sitter $\alpha$ is the Hubble radius at $c/H$, we will derive, using entropy, the cosmological pressure and the cosmological inertia.

### 7.2 Cosmological Pressure

We consider that the cosmological horizon (at $r = \alpha$) 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 empty 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\(^\text{51}\). With these replacements, the Unruh temperature becomes the cosmological horizon temperature\(^\text{52}\):

$$T_{\text{de-Sitter-horizon}} := \frac{\hbar H}{2\pi k_B}$$  \hspace{2cm} (268)

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:

$$T \, dS = -\frac{\hbar H \, k_B c^3}{2\pi k_B \, \hbar G^4} \, dA$$  \hspace{2cm} (269)

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:

$$T \, dS = -\frac{\hbar H \, k_B c^3}{2\pi k_B \, \hbar G^4} 2r^{-1} \, dV$$  \hspace{2cm} (270)


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

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG} \left( \frac{H}{c} \right)^2 dV$$

$$= -\frac{H^2}{4\pi G} c^2 dV$$ (272)

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

$$T \, dS = -\frac{2}{3} \rho c^2 dV$$ (273)

This result was obtained by Easson, where it was suggested as a candidate explanation of the accelerated expansion of the universe.

7.3 Cosmological Inertia

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

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG} \left(8\pi r \, dr\right)$$

$$= -\frac{c^4}{4G} dr$$ (276)

Then, with $r \rightarrow c/H$, we get:

$$T \, dS = -\frac{\hbar H}{2\pi k_B} \frac{k_B c^3}{hG} \frac{c}{H} \, dr$$

$$= -\frac{4M}{G} dr$$ (277)

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:

$$T \, dS = -4Ma \, dr$$

Using these results, we assign to the energy of 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).
7.4 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 ($\approx 5$ giga-parsec and $\approx 4.1$ giga-parsec, respectively).

To account for this difference, we increase the entropy of the system such that the position of the Hubble horizon with respect to the cosmological horizon ends up slightly further. For generality we can in fact include any number of scalar thermodynamic quantities $\{\mu_1 \bar{N}_1, \ldots, \mu_n \bar{N}_n\}$ in the equation, so as to offset the entropy. The equation becomes:

$$S - k_B \ln Z = \pm \left(\frac{2\pi k_B mc}{\bar{h}}\right) \alpha - \mu_1 \bar{N}_1 - \cdots - \mu_n \bar{N}_n$$

where $\alpha$ represents the distance to the Hubble horizon. The contribution from these terms works in the opposite direction to that of the scalar terms $\{\mu_1 \bar{N}_1, \ldots, \mu_n \bar{N}_n\}$. 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 $\lim_{t \to \infty} \alpha \to \left(\frac{2\pi k_B mc}{\bar{h}}\right)^{-1}S$ and for all $i$, $\lim_{t \to \infty} \bar{N}_i \to 0$.

By attributing the role of bookkeeper (of the matter and energy yet to leave the horizon) to $\{\mu_1 \bar{N}_1, \ldots, \mu_n \bar{N}_n\}$, 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:

$$dE = (pc^2 + p) dV$$

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$^{55}$, first for the radius:

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

---

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

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

(281)

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

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

(282)

or, more specifically, as:

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

(283)

This is the sufficient starting point used by Easson$^56$ (in Annex A of his paper) to recover the Friedman equations of cosmology:

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

(284)

$$H^2 = \frac{8\pi G}{3} \rho - \frac{k c^2}{a^2} + \frac{\Lambda c^2}{3}$$

(285)

as an equivalent representation of (283).

It would thus appear that $\Lambda$CDM is to universal statistical physics what the ideal gas law is to statistical physics.

8 Conclusion

As we always suspected, the laws of physics, at their most fundamental level, describes the states of all experiments as constrained by nature, but now with exact mathematical definitions for each term in italic. Furthermore, investigate of the universal metric/norm and how/if it ties to the standard model (beyond electromagnetism) will be investigate in future papers.

Part IV

Annex

A Universal Interval and Norm of a universal event in four dimensions

I have executed the following code in Mathematica 12:
y0 = KroneckerProduct[PauliMatrix[3], IdentityMatrix[2]];
y1 = KroneckerProduct[PauliMatrix[2], PauliMatrix[1]];
y2 = KroneckerProduct[PauliMatrix[2], PauliMatrix[2]];
y3 = KroneckerProduct[PauliMatrix[2], PauliMatrix[3]];
o01 = 1/2 (y0.y1 - y1.y0); o01 // MatrixForm;
o02 = 1/2 (y0.y2 - y2.y0); o02 // MatrixForm;
o03 = 1/2 (y0.y3 - y3.y0); o03 // MatrixForm;
o12 = 1/2 (y1.y2 - y2.y1); o12 // MatrixForm;
o13 = 1/2 (y1.y3 - y3.y1); o13 // MatrixForm;
o23 = 1/2 (y2.y3 - y3.y2); o23 // MatrixForm;
o31 = 1/2 (y3.y1 - y1.y3); o31 // MatrixForm;
y5 = y0.y1.y2.y3;
v0 = y5.y0; v0 // MatrixForm;
v1 = y5.y1; v1 // MatrixForm;
v2 = y5.y2; v2 // MatrixForm;
v3 = y5.y3; v3 // MatrixForm;
M = a IdentityMatrix[4] + X0 y0 + X1 y1 + X2 y2 + X3 y3 + E1 o01 +
   E2 o02 + E3 o03 + B1 o12 + B2 o31 + B1 o23 + V0 v0 + V1 v1 +
   V2 v2 + V3 v3 + b y5;
Expand[Det[M]]

The output for the universal interval is:

\[ a^4 + 2 a^2 b^2 + b^4 + 2 a^2 B1^2 - 2 b^2 B1^2 + B1^4 + 2 a^2 B2^2 -
  2 b^4 B2^2 + 2 B1^2 B2^2 + 2 B3^2 + 2 a^2 B3^2 - 2 b^2 B3^2 +
  2 B1^2 B3^4 + 2 B2^2 B3^4 + 2 B3^4 - 8 a b B1 E1 - 2 a^2 E1^2 +
  2 b^2 E1^2 + 2 B1^2 E1^2 - 2 B2^2 E1^2 - 2 B3^2 E1^2 + E1^4 -
  8 a b B2 E2 + 8 B1 B2 E1 E2 - 2 a^2 E2^2 + 2 b^2 E2^2 -
  2 B1^2 E2^2 + 2 B2^2 E2^2 - 2 B3^2 E2^2 + 2 E1^2 E2^2 + E2^4 -
  8 a b B3 E3 + 8 B1 B3 E1 E3 + 8 B2 B3 E2 E3 - 2 a^2 E3^2 +
  2 b^2 E3^2 - 2 B1^2 E3^2 - 2 B2^2 E3^2 + 2 B3^2 E3^2 + 2 E1^2 E3^2 +
  2 E2^2 E3^2 + E3^4 - 4 a^3 s - 4 a b^2 s - 4 a b^2 s - 4 a B1^2 s - 4 a B2^2 s -
  4 a B3^2 s + 8 b B1 E1 s + 4 a E1^2 s + 8 b B2 E2 s + 4 a E2^2 s +
  8 b B3 E3 s + 4 a E3^2 s + 6 a^2 s^2 + 2 b^2 s^2 + 2 B1^2 s^2 +
  2 B2^2 s^2 + 2 B3^2 s^2 - 2 E1^2 s^2 + 2 b^2 s^2 - 2 B1 E2 V0 V1 +
  2 B2 V0 V2 + 2 B3 V0 V2 + 2 E1^2 V0 V2 + 2 E2 V0 V2 +
  2 E3 V0 V2 + 4 a s V0 V2 - 2 s^2 V0 V2 + V0^4 + 8 B3 E2 V0 +
  8 B2 E3 V0 V1 + 2 a^2 V1 V2 + 2 b^2 V1 V2 - 2 B1^2 V1 V2 +
  2 B2 V1 V2 + 2 B3 V1 V2 - 2 E1^2 V1 V2 + 2 E2 V1 V2 +
  2 E3 V1 V2 + 2 a s V1 V2 + 2 s^2 V1 V2 - 2 V0 V2 V1 + V1^4 +
  8 B3 E1 V0 V2 - 8 B1 E3 V0 V2 - 8 B1 B2 V1 V2 - 8 E1 E2 V1 V2 +
  2 a^2 V2 V2 + 2 b^2 V2 V2 + 2 B1 V2 V2 - 2 B2 V2 V2 + 2 B3 V2 V2 +
  2 E1^2 V2 V2 - 2 E2 V2 V2 + 2 E3 V2 V2 - 4 a s V2 V2 + 2 s^2 V2 V2 -
\]
2 V0^2 V2^2 + 2 V1^2 V2^2 + V2^4 - 8 B2 E1 V0 V3 + 8 B1 E2 V0 V3 -
8 B1 B3 V1 V3 - 8 E1 E3 V1 V3 - 8 B2 B3 V2 V3 - 8 E2 E3 V2 V3 +
a^2 V3^2 + 2 b^2 V3^2 + 2 B1^2 V3^2 + 2 B2^2 V3^2 - 2 B3^2 V3^2 +
2 E1^2 V3^2 + 2 E2^2 V3^2 - 2 E3^2 V3^2 - 4 a s V3^2 + 2 s^2 V3^2 -
2 V0^2 V3^2 + 2 V1^2 V3^2 + 2 V2^2 V3^2 + V3^4 + 8 a B1 V1 X0 -
8 b E1 V1 X0 - 8 a B1 s V1 X0 + 8 a B2 V2 X0 - 8 b E2 V2 X0 -
2 a^2 X0^2 - 2 b^2 X0^2 + 2 B1^2 X0^2 + 2 B2^2 X0^2 + 2 B3^2 X0^2 +
2 E1^2 X0^2 + 2 E2^2 X0^2 + 2 E3^2 X0^2 + 4 a s X0^2 - 2 s^2 X0^2 +
2 V0^2 X0^2 + 2 V1^2 X0^2 + 2 V2^2 X0^2 + 2 V3^2 X0^2 + X0^4 -
8 b B3 V1 X2 - 8 a E3 V1 X2 + 8 E3 s V1 X2 + 8 b B3 V2 X1 +
8 a E3 V2 X1 - 8 E3 s V2 X1 - 8 b B2 V3 X1 - 8 a E2 V3 X1 +
8 E2 s V3 X1 - 8 B3 E2 X0 X1 + 8 B2 E3 X0 X1 - 8 V0 V1 X0 X1 +
2 a^2 X1^2 + 2 b^2 X1^2 - 2 B1^2 X1^2 + 2 B2^2 X1^2 + 2 B3^2 X1^2 -
2 E1^2 X1^2 + 2 E2^2 X1^2 + 2 E3^2 X1^2 - 4 a s X1^2 + 2 s^2 X1^2 +
2 V0^2 X1^2 + 2 V1^2 X1^2 - 2 V2^2 X1^2 - 2 V3^2 X1^2 -
2 X0^2 X1^2 + X1^4 - 8 a B2 V0 X2 + 8 b E2 V0 X2 + 8 B2 s V0 X2 -
8 b B3 V1 X2 - 8 a E3 V1 X2 + 8 E3 s V1 X2 + 8 b B3 V2 X1 +
8 a E1 V3 X2 - 8 E1 s V3 X2 + 8 B3 E1 X0 X2 - 8 B1 E3 X0 X2 -
8 V0 V2 X0 X2 - 8 B1 B2 X1 X2 - 8 E1 E2 X1 X2 + 8 V1 V2 X1 X2 +
2 a^2 X2^2 + 2 b^2 X2^2 + 2 B1^2 X2^2 + 2 B2^2 X2^2 + 2 B3^2 X2^2 +
2 E1^2 X2^2 + 2 E2^2 X2^2 + 2 E3^2 X2^2 - 4 a s X2^2 + 2 s^2 X2^2 +
2 V0^2 X2^2 + 2 V1^2 X2^2 + 2 V2^2 X2^2 - 2 V3^2 X2^2 -
2 X0^2 X2^2 + 2 X1^2 X2^2 + 2 X2^4 - 8 a B3 V0 X3 + 8 b E3 V0 X3 +
8 B3 s V0 X3 + 8 a E2 V1 X3 - 8 E2 s V1 X3 - 8 B1 E2 X3 +
8 a E1 V2 X3 - 8 a E1 V2 X3 + 8 E1 s V2 X3 - 8 B2 E1 X0 X3 +
8 B1 E2 X0 X3 - 8 V0 V3 X0 X3 - 8 B1 B3 X1 X3 - 8 E1 E3 X1 X3 +
8 V1 V3 X1 X3 - 8 B2 B3 X2 X3 - 8 E2 E3 X2 X3 + 8 V2 V3 X2 X3 +
2 a^2 X3^2 + 2 b^2 X3^2 + 2 B1^2 X3^2 + 2 B2^2 X3^2 + 2 B3^2 X3^2 +
2 E1^2 X3^2 + 2 E2^2 X3^2 - 2 E3^2 X3^2 - 4 a s X3^2 + 2 s^2 X3^2 +
2 V0^2 X3^2 - 2 V1^2 X3^2 - 2 V2^2 X3^2 + 2 V3^2 X3^2 -
2 X0^2 X3^2 + 2 X1^2 X3^2 + 2 X2^2 X3^2 + X3^4

and the universal norm is:

a^4 + 2 a^2 b^2 + b^4 + 2 a^2 B1^2 - 2 b^2 B1^2 + B1^4 + 2 a^2 B2^2 -
2 b^2 B2^2 + 2 B1^2 B2^2 + B2^4 + 2 a^2 B3^2 - 2 b^2 B3^2 +
2 B1^2 B3^2 + 2 B2^2 B3^2 + B3^4 - 8 a b B1 E1 - 2 a^2 E1^2 +
2 b^2 E1^2 + 2 B1^2 E1^2 - 2 B2^2 E1^2 - 2 B3^2 E1^2 + E1^4 -
8 a b B2 E2 + 8 B1 B2 E1 E2 - 2 a^2 E2^2 + 2 b^2 E2^2 -
2 B1^2 E2^2 - 2 B2^2 E2^2 - 2 B3^2 E2^2 + 2 E1^2 E2^2 + E2^4 -
8 a b B3 E3 + 8 B1 B3 E1 E3 - 2 a^2 E3^2 +
2 b^2 E3^2 - 2 B1^2 E3^2 - 2 B2^2 E3^2 + 2 B3^2 E3^2 + 2 E1^2 E3^2 +
2 E2^2 E3^2 + E3^4 - 2 a^2 V0^2 - 2 b^2 V0^2 + 2 B1^2 V0^2 +
2 B2^2 V0^2 + 2 B3^2 V0^2 + 2 E1^2 V0^2 + 2 E2^2 V0^2 +
References


