The road from reality: The universe exists as indubitable structured information

Alexandre Harvey-Tremblay

August 30, 2020

This manuscript 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 the laws of physics, is inherited indubitably from pure reason, whilst side-stepping the historical pitfall at the mind-body problem. To derive the laws of physics, the strategy is to construct a purely mathematical model of science whose domain is the set of all indubitable statements, then to solve it for the laws of physics by maximizing the entropy of said statements. In the model, reality is defined as the set of all realized experiments (the set of what 'I' has proven), and the laws of physics are those that remain universally valid for all permutations or rearrangements of experiments constrained only by nature, which we define as the group of all possible transformations of the set. At its most fundamental level, physics is—quite simply—the probability measure that makes reality maximally informative (to 'I'/the observer) within the constraint of nature. Specifically in regards to novel physics, the procedure yields the mathematical origin of the (extended) Born rule as the connection between physics and reality, while space-time itself is emergent as the maximally structured interference pattern producible by said extended Born rule; a process which, notably, is self-limited to precisely 4 space-time dimensions.

Contents

1 Notation

2 Towards a mathematical model of reality
   2.1 Anti-pattern: postulated laws
   2.2 Math offers no free lunches
   2.3 The fundamental structure of reality

3 Towards a mathematical model of science
   3.1 Hint: John A. Wheeler
   3.2 Hint: Gregory Chaitin

4 Foundation
   4.1 The Axioms of Science
   4.2 The Axioms of Reality
   4.3 The Axioms of Physics
Towards a mathematical proof of physics

Recap: Statistical Physics

Recap: Algorithmic Thermodynamics

Attempt 1: Literal system

Attempt 2: Designer Ensemble

Anti-pattern: axiomatic fixing

Physics as the ultimate probability measure of reality

Intuition: Universal Thermodynamics

The mathematical origin of the Born rule

The connection between science and reality

Space-time as an interference pattern

The Universal Norm

Quartic Hilbert spaces

Octic Hilbert spaces

Quantum computation

Discussion

What is space-time?

The measurement problem (recap)

The measurement problem (solution)

A Sketch/Figure

Notation

Parentheses will be used to denote the order of operations and square brackets will be used exclusively for the inputs to a map. For instance a map \( f : X \rightarrow \mathbb{R} \) will be written as \( f[x] \) for \( x \in X \). \( S \) will denote the entropy, \( S \) the action, \( L \) the Lagrangian, and \( L \) the Lagrangian density. Sets, unless a prior convention assigns it another symbol, will be written using the blackboard bold typography (ex: \( L, W, Q \), etc.). Matrices will be in bold upper case (ex: \( A, B \)), whereas vectors and multivectors will be in bold lower case (ex: \( u, v, g \)) and most other constructions (ex.: scalars, functions) will have plain typography (ex. \( a, A \)). The identity matrix is \( I \), the unit pseudoscalar (of geometric algebra) is \( I \) and the imaginary number is \( i \). The Dirac gamma matrices are \( \gamma_0, \gamma_1, \gamma_2, \gamma_3 \) and the Pauli matrices are \( \sigma_x, \sigma_y, \sigma_z \). The basis elements of an arbitrary curvilinear geometric basis will be denoted \( e_0, e_1, e_2, \ldots, e_n \) (such that \( e_\nu \cdot e_\mu = g_{\mu\nu} \)) and if they are orthonormal as \( \hat{x}_0, \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n \) (such that \( \hat{x}_\mu \cdot \hat{x}_\nu = \eta_{\mu\nu} \)). The asterisk \( * \) denotes the complex conjugate of \( z \), and the dagger \( A^\dagger \) denotes the conjugate transpose of \( A \). A geometric algebra of \( n \) dimensions over a field \( \mathbb{F} \) is noted as \( G_n(\mathbb{F}) \). We note the matrix representation of a multivector \( g \) as \( M[g] \), defined as a map \( M : G_n(\mathbb{F}) \rightarrow M(n, \mathbb{F}) \).
which equates the geometric product to the matrix product, and thus benefits from group isomorphism. The grades of a multivector will be denoted as $\langle v \rangle_k$. Specifically, $\langle v \rangle_0$ is a scalar, $\langle v \rangle_1$ is a vector, $\langle v \rangle_2$ is a bivector, $\langle v \rangle_{n-1}$ is a pseudovector and $\langle v \rangle_n$ is a pseudoscalar. Furthermore, a scalar and a vector $\langle v \rangle_0 + \langle v \rangle_1$ is a paravector, and a combination of even grades $(\langle v \rangle_0 + \langle v \rangle_2 + \langle v \rangle_4 + \ldots)$ or odd grades $(\langle v \rangle_1 + \langle v \rangle_3 + \ldots)$ are even-multivectors or odd-multivectors, respectively. The commutator is defined as $[A, B] := AB - BA$ and the anti-commutator as $\{A, B\} := AB + BA$. We use the symbol $\cong$ to relate two sets that are related by a group isomorphism (ex: $\mathbb{G}_4(C) \cong M(4, C)$).

2 Towards a mathematical model of reality

My primary goal with this work is to construct two mathematical models; first, a model of reality and, second, a model of science. Furthermore, these constructions will have the key feature that applying the second model (science) to the first (reality) produces a third model (physics). The constructions not only provides an account for the mathematical origin of the laws of physics as a mere theorem of the practice of science, but it almost further trivializes this origin to a tautology. This effectively reduces the search space for an ultimate theory of physics to a pulp which is one of the key motivation of this effort.

To achieve this goal, we must begin with a number of modifications to our understanding of the usual practice of science such that it is conductive to a mathematical formulation. Let us start with a quick summary of the very familiar current practice. As illustrated in Figure 1, in the current practice both theoretical and empirical physics work in tandem to eventually (and hopefully) converge towards a correct model of reality via an iterative falsification/refinement process. In the one hand, postulated laws are compared to empirical laws, and in the other, measured experimental states are compared to predicted experimental states. Any discrepancy then ought to trigger a modification of the postulated laws, and the process begins again with the new postulated laws. From the figure, we note that ‘empirical physics’ measures experimental states then derives empirical laws, and theorectal physics postulates laws, then derives states. As such, they are logically entailed in opposite directions.

The modification that I propose is summarized in Figure 2. The key change is that the usual relationship between postulated laws and predicted states will now be reversed. Instead of postulating laws we postulate states, and instead of solving for states, we solve
for laws.

How does one solves for laws? The laws derived by my method are 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 law in our setup is derived as the universal pattern found by permuting over all possible arrangements and rearrangements of postulated states.

An equivalence thesis between the set of all possible thought experiments and all possible real experiments is supported as a direct consequence of the mathematical universality and other completeness considerations which applies to both sets — remarkably, it cannot be the case that there exists a real experiment that cannot be formulated as a (properly constructed) thought experiment, and vice-versa. With the equivalence thesis, it is then implied that my method is a mathematical copy of empirical physics. Using this copy, one recovers the laws of physics by applying science to the set of all possible thought experiments, for the same reason that applying science to all possible
real experiments also produces them. In the present case however, the laws of physics are derived within having to leave the realm of mathematics.

2.1 Anti-pattern: postulated laws

An anti-pattern is a common response to a recurring problem that is usually ineffective and risks being highly counterproductive.³

I’ve nearly always harbored the feeling that postulating an equation (or generally a law, or set of laws) to be true (in the axiomatic sense) is a critical mistake; that is, we are postulating the wrong kind of object and that this mistake produces problems at the most fundamental level. That is, postulated laws are an anti-pattern. I will now give my best attempt to explain my intuition behind this.

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 Roger 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, Roger 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."

Penrose erects what is possibly the highest "tower of postulates" produced thus far, and then concludes that he does not see a road to reality in any of the proposals. So... why does tower-building not achieve the stated goal of bringing us closer to reality? What’s truly missing?

My earliest memory of thinking about this was on my second day of school, but before I can explain what happened and why it happened I need to lay out the context leading up to it. So, please allow me to share this anecdote — I promise it will be relevant. My father’s strategy of choice to prepare me for the world was, I would summarize, to “sync my mind to reality” 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 would not develop an idea that is "disconnected from reality". Essentially, he attempted to falsify whatever expectations of reality I would derive from my internal model of 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, causing me to ponder on whose memories can be trusted more; his or mine, and how would I know. This is just a small example, but ‘tricks’ of this nature were made on a daily basis. 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, instead to simply accept that the present state of the world is the undeniable foundation to reality; any expectations I might have of its future states (and in the extreme case even my memory of past states could be questioned) can be no less than the set of all ‘physically-permissible’ rearrangement of the environment. Consequently, I reasoned that the best case strategy to adopt in the wild was to assign a likelihood to each scenario and to preferably have a backup plan for most undesirable scenarios so as to amortize the risk/fluctuations over time. My intuitive mental foundation to reality was that the instantaneous state of the system is the only arbiter of truth for the system — everything else is up to questioning.

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. 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
the universe exists as indubitable structured information — 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 arithmetic 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 holds some statement to be true that are outside the scope of reality. Thus I held as strongly suspect even something as seemingly banal as the inclusion of ‘unbounded’ arithmetic within the "tower of postulates" of reality.

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 in its infinite scope. 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; from natural numbers, to classical mechanics to eventually quantum field theory and everything in between. In
my mind each additional layer takes me 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 do not share the belief that building a tower of postulates will ever bring us closer to reality. All "tower-builders" make the same fundamental critical mistake: they assume that reality comes from the tower. This reserved logical entailment came to be the primary mode of mathematical construction of physical theories; from Newton to quantum field theory and almost everything in between. As an example, let us consider Newton’s second law of motion, mathematically expressed as follows:

$$\sum_{F \in F} F = ma$$

(1)

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)? What price do we pay when we erase that ‘something’ by writing "$F = ma" as an axiom instead of as a theorem? Specifically, we create what seems to me as a cargo cult, and allow me to explain.

A cargo cult is characterized as a belief, by a technologically less advanced 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. Wrong: reality is at the street level, not at the penthouse.

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\(^4\) 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; 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 of the highest level. 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 tower-building as a whole, of which Penrose’s book just happens to be the single most complete embodiment of such. I stress that this is not meant as a critique of Roger Penrose’s book specifically.

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 confirm-a

tion/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\(^5\) (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 perform science, because the process could, in principle, eventually just halt on it (without actually converging towards it).

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. As long as the thesis holds, the laws derived from my process ought not to diverge from reality.

2.2  Math offers no free lunches

Proving not only the laws of physics, but also the existence of a structure which obeys these laws —sometimes called physical reality— purely mathematically runs counter to expectations and therefore it is widely assumed to be an impossibility. However, contrary to ex-

\(^5\) 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
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 the technique that I use, which illustrates the key concepts, then further in the paper we will thoroughly investigate the technique to produce a mathematical model 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\(^6\) perfect\(^7\) hash function.

For example, suppose my hash produces the following result:

\[
\text{hash}[1 + 1 = 1] = \text{fa1869db4bfbf1767a5446b6a9290243} \quad (2)
\]

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

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

I also define the inverse function:

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

The bruteforce function finds the solution by brute force: it hashes all statements of \(\mathbb{L}_{\text{PA}}\) 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.

Now, let me define a new axiom as follows:

**Definition (Axiom of rot).**

\[
\text{rot} := \text{bruteforce}[\text{fa1869db4bfbf1767a5446b6a9290243}] \quad (5)
\]

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).

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

(6)

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 bruteforce function. But consider instead the following axiom:

**Definition** (Axiom-X).

**Axiom-X :** = bruteforce[0cfae383362bc63d7ac429a5755fe05] (7)

Now I ask you, knowing the hash but not the statement, is the formal theory comprised of PA \( \cup \text{Axiom-X} \), 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 significant expenditure of computing 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\(^8\) estimates that there are approximately \( 10^{122} \) bits (and approximately the same amount of operations) available for computations in the universe. What if our bruteforce 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\(^9\), then a contradiction could be injected,

---


the universe exists as indubitable structured information

accidentally or on purpose, at any level of computational complexity within a theory. My specific example with a brute force function shows how to purposefully inject a contradiction at a tunable level of computational 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 with computing resources 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.

Let us take another example by asking the question: "Is ZFC consistent?". An answer along the following lines is often provided: "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 random article was published 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? More specifically, why does 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 a framework able to formalise this intuition. 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, a ground state and an equation of state. The elements of the quartet are related to each other as shown on Figure 3.

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

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 equation of state that quantifies the amount of mathematical work required to excite the system as the manifest is moved further along the axis.

2.3 The fundamental structure of reality

Why is this approach able to transpose properties to reality? Specifically, it boils down to the relationship between manifest and mathematical work:

**Syllogism 1 (The fundamental structure of reality).**

1. All manifests are contingent on mathematical work.
2. At least one manifest exists indubitably.
3. Therefore, reality is contingent on mathematical work.

(We note that statement 1 was argued for in the previous section and will be formalized in section 5.2, and statement 2 will be proven here.)

To fundamentally understand the power of this syllogism, and to prove statement 2, we have to start at the very beginning of the rational inquiry. Allow me first to lay out the groundwork using Cartesian philosophy, then we will use our tools to modernize the argument. We will recall the philosophy of René Descartes (1596–1650), the famous 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 address 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 (a good part) of a hundred pages manuscript. 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”12. We note that bio-electrical signals were characterized in 1791 by Luigi Galvani, as bioelectro-

magnetics 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 remained.

The proof of the ‘cogito ergo sum’ by Descartes is the only proof
I have encountered that I feel 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 the construction that I have proposed in
terms of axioms, manifest and mathematical work along with the
above syllogism, such that proving statements from a finitely ax-
iomatic system is contingent on the consumption of mathemati-
cal work is a repeat, in disguise, of the universal doubt method of
Descartes (including its conclusion), but extended to the structure
of reality (“the body”). Descartes essentially constructed a universal
proof checker, and applied it to the set of all mathematical statements
(although in an informal manner, and consequently missed out on
this massive opportunity).

Let us now do an exercise that will reveal this missed opportu-
nity. 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 for this
purpose.

Let us start by defining what a language is:

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

$$L := \{ \emptyset, s_1, s_2, \ldots \} \quad (9)$$

For instance, the sentences of the binary language are:

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

and its alphabet is:
\[ \Sigma_b = \{0, 1\} \quad \text{(11)} \]

**Definition 2** (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:

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

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

**Definition 3** (Minimal proof checker).

\[
\text{MPC} : \quad L \rightarrow \{1, \#\} \\
\text{sentence} \mapsto \text{result} \quad \text{(13)}
\]

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 of the input sentence using said 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 here 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 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 the behavior of MPC using an example. Say Descartes feeds the sentence \((1 + 1 = 2)\) to the MPC. Will the machine find a proof for it? Well let's 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 fed to MPC:

\[
\text{PA} \vdash (1 + 1 = 2) \quad \text{(14)}
\]

The statement states: PA proves that one plus one equals two. As we did with the previous example, we also ask here will \( \text{MPC}[\text{PA} \vdash \cdot] \)
(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 required to become a necessary true statement.

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 the means to describe reality to any level of complexity or expressivity desired or required while guaranteeing that we tell the truth and nothing but the truth, and furthermore that we do so without having to construct a preliminary "tower of postulates" to do so. Manifests constructed from these statements are necessarily the most fundamental description of reality possible; indeed, all other representations that include at least one statement not of this type must adopt said statement as an axiom (without proof), and thus invariably ends up being less fundamental.

Let us define a set \( M \), which we call a manifest, of \( n \) statements verifiable by MPC:

\[
M := \{ \\
A_1 \vdash T_1, \\
A_2 \vdash T_2, \\
A_3 \vdash T_3, \\
\vdots \\
A_n \vdash T_n, \\
\}
\]

(15)

where the letter \( A \) designate the axiomatic part of the statement, and the letter \( T \) designate the theorem part. Finally, the symbol \( \vdash \) states that \( A \) logically entails, or proves, \( T \).

In this case, the manifest is a set of indubitable statements of the kind that the universal doubt method of Descartes fails to invalidate. A specific instance of such a manifest can be constructed simply by conducting a small number of thought experiments; it could be as simple as doing basic arithmetic in one’s head:

**Syllogism 2** (At least one manifest exists indubitably). :

1. That which survives the universal doubt method exists indubitably.

2. The statements of the manifest comprised of \{PA \vdash (1 + 1 = 2)\} survives the universal doubt method.
3. Therefore, at least one manifest exists indubitably (as a provable fact of reality).

We note that existence is proven in a similar sense to how Descartes uses the universal doubt method to prove his own existence only as a thinking being, but fails to prove that he has a body; here indubitably proving \( \{ PA \vdash (1 + 1 = 2) \} \) implies the statements exists as a provable fact of reality (it does not imply that the statement exists as a physical object because physical substance is the purview of mathematical work, not of statements).

This completes the proof of statement 2 of syllogism 1.

3 Towards a mathematical model of science

This definition of a manifest as a set of statements of the type \( A \vdash T \), although conceptually simpler, is not the best possible definition for a manifest because it does not produces the ‘cleanest’ of theorems. For instance, the dependence on the symbol \( \vdash \) is a hindrance, notably, because it rules out all formal languages that do not admit the symbol even if they may be completely legitimate otherwise. Furthermore, the dependence on MPC on some specific logical axioms is also an undesirable that should be removed. The formulation should be fee of linguistic features. The preferred framework to formalize these definitions will be that of algorithmic information theory and that of theoretical computer science including the formalism of Turing machines. In this language, a manifest is simply a finite set of programs:

\[
M := \{ p_1, p_2, p_3, \ldots, p_n \}
\]  

(16)

To each such manifest one can then define at least one Turing machine \( TM \) that halts for (and only for) each \( p \in M \). In more technical terms, \( M \) is the domain of \( TM \). Mathematical work, that we now prefer to call computational work, is produced by this Turing machine as the elements of \( M \) are verified.

To obtain scientific theories one first selects a universal Turing machine \( UTM \) as the baseline, then for each program \( p \in M \) a Turing machine, understood as an abstraction layer and referred to in this context as a ‘scientific theory’, is also provided as input to the universal Turing machine. Specifically, all programs of \( M \) are then verified as follows:
\[
\{ \\
    \text{UTM}[T_{M_1}, p_1] \\
    \text{UTM}[T_{M_2}, p_2] \\
    \text{UTM}[T_{M_3}, p_3] \\
    \vdots \\
    \text{UTM}[T_{M_n}, p_n] \\
\} \quad (17)
\]

Although this is not what we typically think of when we think of a physical theory, a universal Turing machine UTM that recursively enumerated $M$ does provide the means to verify all programs of the manifest, but it does so in a "patchy manner". Specifically, in the case where two or more programs are verifiable by the same Turing machine, say it happens to be the case that $T_{M_3} = T_{M_4} = T_{M_9}$, then the theory forms a "logical grain" such that the programs $p_3, p_4, p_9$ are entailed from the same axiomatic basis.

If computational work is abundant, large logical grains are favored. Intuitively, one with access to a very powerful quantum computer can, in principle, directly solve quantum field theory to get the higher scientific discipes (chemistry first, then biology and so on — the higher level theories are eventually made redundant by computational work abundance). Whereas; one with access to less computational work must then compensate by creating more abstraction layers as required such that every program is within its computational reach.

The set of all scientific theories for the system are the various logical grains that can be formed based on the availability of computational work. Like the grain structure that arises naturally when a crystal (of solid state physics) is formed, an initial choice of abstraction layer can then determine how future grains can be structured whilst fitting with the pre-existing arrangement of grains. Furthermore, scientific theories are subject to refinements as grains are fused or reorganized following a change in computational work availability.

One is free, even encouraged, to produce a plurality of logically independent scientific theories, each valid within their own domain of applicability and each corresponding to a grain and each verifying a subset of the manifest. In this sense, we may suspect that all scientific theories are subject to possible falsification or refinements as computational work abundance is increased, but some formulations are non-abstracting and universally valid for all arrangements and rearrangements of the system. For instance, the theory of computa-
tion is, this context, a meta-scientific theory which holds a privileged position (computational work is computational work regardless of its abundance), and, as we will eventually see, physics is another.

So, to be truly general we will think of statements as arbitrary programs that halt of a universal Turing machine, instead of as sentences with specific structures. Let us now investigate how it is possible for reality to be completely equivalent to a set of programs, as we enter the first hint.

### 3.1 Hint: John A. Wheeler

Information, physics and entropy have, of course, a long and rich history. Skipping over the very familiar Maxwell demon (for brevity) we take, as an example, the Landauer\textsuperscript{13} limit, an expression for the minimum amount of energy required to erase one bit of information for a system at thermodynamic equilibrium:

\[ E \geq T k_B \ln 2 \quad (18) \]

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 physical 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 of an horizon) and physical states (in the bulk of the enclosed volume). Lets 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 been 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

THE UNIVERSE EXISTS AS INDubITABLE STRUCTURED INFORMATION 22

c 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\(^\text{15}\) 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 provides general guidance on how to rectify this. It is there that he introduces the core idea that the bits are the result of the registering of equipment-evoked responses. With this John A. Wheeler discards the idea of referring to the cosmos at all to enforce any kind of properties on the distribution of the bits and instead refers to equipment evoked responses exclusively. After-all, evidence for both time and the cosmos are derived from the information provided to us by experimental devices (including the biological senses).

This completes our summary of the core concepts of John A. Wheeler’s participatory universe hypothesis\(^\text{16}\).

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:

\[11101011001001110101010101\]  \(19\)

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 programs is made. As shown on Figure 4, instead of connecting information to some complex pre-existing physical 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.

3.2 **Hint: Gregory Chaitin**

But before we can formalize science within mathematics, it helps 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 \(\mathcal{U}\) be the set of all universal Turing machines, then:


The universe exists as indubitable structured information

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

\[ UTM \longmapsto \sum_{p \in \text{Dom}[UTM]} 2^{-|p|} \] (20)

The image of \( \Omega \) is the set of all real numbers that are normal, incompressible and provably algorithmically random due to their connection to the halting problem in computer science. We note to the reader that we offer a more detailed primer on \( \Omega \) in the few paragraphs of our technical introduction (Section 5.2) on algorithmic information theory.

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

Finally, Gregory Chaitin further suggests:

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

In another more recent article, 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.

4 Foundation

4.1 The Axioms of Science

The fundamental object of study of science is not the electron, the quark or even super-strings, but the experiment. An experiment represents an ‘atom’ of verifiable knowledge.

Definition 4 (Experiment). An experiment $p$ is a tuple comprising two sentences of $L$. The first sentence, $h$, is called the hypothesis. The second sentence, $TM$, is called the protocol. Let $UTM: L \times L \rightarrow L \cup \{\#\}$ be a universal Turing machine, then we say that the experiment holds if $UTM[TM, h]$ halts, and fails otherwise:

\[
UTM[TM, h] \begin{cases} \text{halts} \Rightarrow p \text{ holds} \\ \# \quad \neg \text{halts} \Rightarrow p \text{ fails} \end{cases}
\]

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

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

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 5** (Domain of science). We note $D$ as the domain (Dom) of science. We can define $D$ in reference to a universal Turing machine $UTM$ as follows:

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

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

(We note that the choice of $UTM$ determines the language/structure of the programs of the domain, however the formalism will be independent of this choice.)

**Definition 6** (Manifest). A manifest $M$ is a subset of $D$:

$$M \subset D$$

We note that the set of all possible manifests is the power set of $D$:

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

**Definition 7** (Observer). An observer $\mathcal{O}$ is a Turing machine that recursively enumerates the domain of science. Given a sentence $s$ as input, the observer eventually halts for $s$ iff $s$ is an element of the domain of science, otherwise it never halts.

### 4.2 The Axioms of Reality

The fundamental object of study of reality is the manifest.

**Assumption 1** (The fundamental assumption of reality). The state of affairs of the world is describable as a set of 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. In other words, experiments are complete with respect to reality.
Axiom 1 (Existence of the reference manifest). As the world is in a given state of affairs, then there exists, as a brute fact, a manifest $\hat{\mathbb{M}}$ which corresponds to its state:

$$\exists! \hat{\mathbb{M}}$$  \hspace{1cm} (25)

- $\hat{\mathbb{M}}$ is called the ‘reference manifest’.

- The symbol $\mathbb{M}$ will denote any manifest in $S$, whereas $\hat{\mathbb{M}}$ 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 $\mathbb{M}$ exists ($\exists$), but in reality, only the state of affairs described by $\hat{\mathbb{M}}$ exists ontologically as verified facts (whereas any $\mathbb{M} \neq \hat{\mathbb{M}}$ exists as verifiable facts but that are not yet verified).

- Unique to $\hat{\mathbb{M}}$, and unlike other manifests, 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.

4.3 The Axioms of Physics

The axioms of physics, comprises the axioms of reality, those of science and the following:

Axiom 2 (Equivalence thesis). All experiments verified by $\mathcal{O}$ are elements of $\hat{\mathbb{M}}$, and all elements of $\hat{\mathbb{M}}$ are verified by $\mathcal{O}$.

Intuition: The reference manifest is the set of all observations and of all experiments made by the observer.

Intuition 2: This idea is closely related to the concept in ordinary quantum physics that a quantity may exist if and only if it is measured.

Technical note: Using a result of usual quantum physics, that the set of all observations need, a-priori, only be defined in reference to one observer is supported in a very general sense in the form of the Wigner’s friend thought experiment: An observer that made a measurement, but his hiding this information from other observers, is acting as a glorified hidden variable theory, which is ruled out by Bell’s inequality. Consequently, it follows that no observer can in principle hide measurement results from other observers.
Personal note: I had initially assumed that the starting point would include multiple observers and that the reference manifest would be defined in reference to the union of all observations made by all observers. Then I attempted to extend this initial idea using a theory of observer-communication & agreement which equated the set of all agreed upon observations to the foundation of reality. Eventually, I realized I was going down the wrong path. I realized that I only needed to start with one observer, because the existence or non-existence of other observers is simply a fact, like any other, itself subject to experimentation and falsification by inspection of the elements of the reference manifest. Indeed, a newborn baby will eventually deduce that other observers exist by inspecting the evidence — it is not a-priori knowledge.

The requirement that the elements of the reference manifest are verified implies, by syllogism 1, the existence of mathematical work in quantities exactly sufficient to verify them. In the context of the reference manifest, we give mathematical work the special name of nature.

Definition 8 (Nature). Nature $N$ is a system of mathematical work used to verify $\hat{M}$. Thus, experiments are verified in nature.

Syllogism 3 (Existence of Nature).

1. All manifests are contingent on mathematical work.
2. There exists the reference manifest (Axiom 1).
3. Therefore, reality is contingent on nature.

We note that since the state of affairs represents the axiomatic basis of the model, it cannot be derived from more fundamental principles. 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.

Assumption 2 (The fundamental assumption of physics). The reference manifest $\hat{M}$ is randomly selected from the set of all possible manifests $P[D]$.

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. However, as existentially 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 $\mathcal{P}[D]$. Using $\rho[M]$ as the probability measure, we will be able to quantify the information in the message $\hat{M}$.

It is from this connection to information that we will find our opportunity to create a physical theory out of nature. For this purpose, we will investigate the framework of statistical physics which is able to constrain a probability distribution, or more precisely the entropy of such, with a set of constraints, as the candidate to recover physics. Here, the manifests will serve as the microstates, and nature will be the macroscopic constraint.

However, we will find that statistical physics, in its usual form, comes short of the goal. It will be in fact, using an extension to statistical physics, that I call universal statistical physics, that we will be able to reformulate physics as entirely emergent in the sense of statistical physics.

**Definition 9** (Physics). We define physics as the probability measure that maximizes the information $O$ gains by knowing $\hat{M}$ as an element that is randomly selected from $\mathcal{P}[D]$, under the constraint of nature $N$.

As we will see with these axioms and definitions, our goal to reduce physics to its simplest and purest expression, such that the recovery of the laws of physics is incidental to this information maximization procedure, will have been achieved.

The fundamental object of study of physics is the probability measure that connects science to reality.

5 Towards a mathematical proof of physics

To precisely quantify the relationship between entropy, mathematical work, and how this implies the laws of physics as the probability measure of a system of verified experiments, we will eventually construct a statistical ensemble of universal statistical physics. But before we introduce this framework, we will provide a recap of ordinary statistical physics, and then of algorithmic thermodynamics.

5.1 Recap: Statistical Physics

The applicability of statistical physics to a given physical system relies primarily upon two assumptions$^{21}$.

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."$^{22}$

$^{21}$ Jos Uffink. Compendium of the foundations of classical statistical physics. Philosophy of Physics, 03 2006

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 $N$ be a set of $n$ constraints (identified as $O_1, O_2, \ldots, O_n$), then set of all probability measures compatible with the constraints is:

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

(26)

The observables, in general, are functions defined as:

$$O_i : P \rightarrow \mathbb{R}$$

$$\rho \mapsto \sum_{q \in Q} \rho[q] O_i[q]$$

(27)

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

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

Table 1: Typical thermodynamic quantities

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

$$S : P \rightarrow [0, \infty]$$

$$\rho \mapsto -k_B \sum_{q \in Q} \rho[q] \ln \rho[q]$$

(28)

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:

$$\rho : Q \times \mathbb{R}^n \rightarrow [0,1]$$

$$(q, \alpha_1, \ldots, \alpha_n) \mapsto Z^{-1} \exp \left( -\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q] \right)$$

(29)

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)

(30)

and the observables (which includes the \( n \) constraints) are expressed as follows:

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

(31)

The constraints are also equivalently given by the following relations:

$$
\frac{\partial \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i} = \overline{O}_i
$$

(32)

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

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

(33)

The entropy for this ensemble is:

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

(34)

Taking the total derivative of the entropy, we obtain:

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

(35)

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 (35). Thermodynamic changes (and cycles) can be realized by changing the quantities \( \{\alpha_1, \ldots, \alpha_n\} \) and/or by modifications of \( Q \). Under modification of \( Q \), usually by cross product: \( Q \times Q_1 = Q_2 \), or by set complement \( Q \setminus Q_3 = Q_4 \), quantities which are invariant \( \{\alpha_1, \ldots, \alpha_n\} \) are called intensive, and quantities which are variant \( \{O_1, O_2, \ldots, O_n\} \) are called extensive.

As an example, consider the following typical thermodynamic quantities taken from Table 1:
\[ a_1 := \beta \]  
\[ a_2 := \gamma \]  
\[ a_3 := \delta \]  
\[ O_1[q] := E[q] \]  
\[ O_2[q] := V[q] \]  
\[ O_3[q] := N[q] \]

the partition function would be:

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

The Gibbs measure is:

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

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:

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

and the equation of state is:

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

In the case where the constraints are continuous, the partition function may be replaced by an integral:

\[ Z[\beta] = \int \exp \left( -\beta H[O_1, O_2, \ldots, O_n] \right) dO_1 dO_2 dO_3 \]
Finally, in the case where the constraints are uncountable and are functions, the partition function is to be replaced by a functional integral:

$$Z[\beta] = \int \mathcal{D}\phi \exp (-\beta H[\phi])$$  \hspace{1cm} (50)

5.2 Recap: Algorithmic Thermodynamics

Many authors[18, 6, 19, 20, 21, 22, 23, 24, 25] 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\textsuperscript{23}. First let us introduce $\Omega$. Let $\mathcal{U}$ be the set of all universal Turing machines, and let $\text{UTM}$ be an element of $\mathcal{U}$. Then, the usual definition of $\Omega$ is:

$$\Omega := \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|}$$  \hspace{1cm} (51)

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 $\text{UTM}$. Chaitin’s construction\textsuperscript{24} (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\textsuperscript{25}). As the sum does not erase halting information, knowing $\Omega$ is enough to know the programs that halt and those that do not on $\text{UTM}$. 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\textsuperscript{26} calculated the first 64 bits of $\Omega$ for some specific universal Turing machine $u \in \mathcal{U}$ as:

$$\Omega_u = 0.0000001000000100000110\ldots_2$$  \hspace{1cm} (52)

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


\textsuperscript{24} Gregory J. Chaitin. A theory of program size formally identical to information theory. J. ACM, 22(3):329–340, July 1975


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\(^{27}\) suggested augmenting \( \Omega \) with a multiplication constant \( D \), which acts as an ‘algorithmic decompression’ term on \( \Omega \).

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

\[
\Omega = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-|p|} \quad \rightarrow \quad \Omega[D] = \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|} \tag{53}
\]

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

\[
\begin{align*}
\text{Gibbs ensemble} & \quad \rightarrow \quad \text{Tadaki ensemble} \\
Z[\beta] &= \sum_{q \in \mathbb{Q}} e^{-\beta E[q]} & \Omega[D] &= \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-D|p|} \\
\end{align*}
\tag{54}
\]

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

\[
|p| = \sum_{p \in \text{Dom}[\text{UTM}]} |p|2^{-D|p|} \tag{55}
\]

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

\[
S[D] = \ln \Omega + D|\bar{p}| \ln 2 \tag{56}
\]

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\(^{28}\) 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 aim 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:

$$
\Omega : \mathbb{R}^3 \rightarrow \mathbb{R}
(\beta, \gamma, \delta) \mapsto \sum_{p \in \text{Dom}[\text{UTM}]} 2^{-\beta E[p] - \gamma V[p] - \delta N[p]} \quad (57)
$$

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

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

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

"–John C. Baez and Mike Stay

From (Equation 57), they derive analogs of Maxwell’s relations and consider thermodynamic cycles, such as the Carnot cycle or Stoddard cycle. For this, they introduce the concepts of algorithmic heat and algorithmic work. Finally, we note that other authors have suggested other alternative mappings in other but related contexts.29

5.3 Attempt 1: Literal system

Let me start by giving out two attempts and then the retained solution.

My first attempt consisted of taking algorithmic thermodynamic at face-value and to apply it to the presently introduced model of reality. For this purposes we will use quantities consistent with the

computer-theoretic origin of algorithmic thermodynamics. 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 types of 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 $k$, the computing repetency conjugated with $L[p]$ the program size, and $f$ the computing frequency conjugated with $T[p]$ the program runtime. The partition function is:

$$Z : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(k, f) \mapsto \sum_{p \in \text{Dom}[UTM]} 2^{-kL[p] - fT[p]}$$

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L[p]$</td>
<td>program size</td>
<td>[bit]</td>
<td>extensive</td>
</tr>
<tr>
<td>$k$</td>
<td>computing repetency</td>
<td>[1/bit]</td>
<td>intensive</td>
</tr>
<tr>
<td>$T[p]$</td>
<td>program runtime</td>
<td>[operation]</td>
<td>extensive</td>
</tr>
<tr>
<td>$f$</td>
<td>computing frequency</td>
<td>[1/operation]</td>
<td>intensive</td>
</tr>
<tr>
<td>$T$</td>
<td>average clock usage</td>
<td>[operation]</td>
<td>macroscopic</td>
</tr>
</tbody>
</table>

Table 2: Algorithmic quantities of the canonical ensemble of programs

The second partition function is a grand canonical ensemble. It is obtained by multiplying multiple partition functions of the canonical type:

$$Z = \left( \sum_{p \in \text{Dom}[UTM]} 2^{-kL[p] - fT[p]} \right)^n$$

Distributing the terms of the sums results in a sum that is the equivalent of a grand partition function describing an ensemble of sets of programs. The resulting partition function is over manifests:

$$Z = \sum_{M \in \text{Dom}[UTM]} ^{g[M]} 2^{-kL[M] - fT[M]}$$

where $g[M]$ is the degeneracy of the state $M$. Executing a manifest of programs on a universal Turing machine refers to a specific computation involving multiple programs. In the grand canonical ensemble, it is customary to add a quantity, such as $\mu$ the computing overhead and conjugate it to $N[M]$, the quantity of programs in the manifest to account for ‘equilibrium-preserving’ changes in quantities of
programs within the manifests. This new quantity is shown in Table 3.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Units</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>size of programs in the manifest</td>
<td>[bit]</td>
<td>extensive</td>
</tr>
<tr>
<td>$k$</td>
<td>computing repetency</td>
<td>[1/bit]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>average tape usage</td>
<td>[bit]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$T$</td>
<td>running time of programs in the manifest</td>
<td>[operation]</td>
<td>extensive</td>
</tr>
<tr>
<td>$f$</td>
<td>computing frequency</td>
<td>[1/operation]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\bar{T}$</td>
<td>average clock usage</td>
<td>[operation]</td>
<td>macroscopic</td>
</tr>
<tr>
<td>$N$</td>
<td>quantity of programs in the manifest</td>
<td>[program]</td>
<td>extensive</td>
</tr>
<tr>
<td>$\mu$</td>
<td>computing overhead</td>
<td>[1/program]</td>
<td>intensive</td>
</tr>
<tr>
<td>$\bar{N}$</td>
<td>average concurrency</td>
<td>[program]</td>
<td>macroscopic</td>
</tr>
</tbody>
</table>

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

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

- The computing repetency: $k$ counts how many times the average tape usage $\Lambda$ must be doubled to double the entropy of the ensemble while holding the average clock usage $\bar{T}$ and the average concurrency $\bar{N}$ fixed.

- The computing frequency: $f$ counts how many times the average clock usage $f$ must be doubled to double the entropy of the ensemble while holding the average tape usage $\Lambda$ and the average concurrency $\bar{N}$ fixed.

- The computing overhead: $\mu$ counts how many times the average concurrency $\bar{N}$ must be doubled to double the entropy of the ensemble while holding the average clock time $\bar{T}$ and the average tape usage $\bar{L}$ fixed.

Flexibility is available in the form of the various systems of natural computing that can be produced by defining other computing resources, or set filtering conditions. Let us give a few examples.

1. Computing time to program frequency formulation:

$$Z' : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(k,t) \rightarrow \sum_{p \in \text{Dom}[UTM]} 2^{-kL[p]} - tf[p]$$  \hspace{1cm} (61)

To formulate this relation, we introduce the program frequency $F[p]$ as the inverse of the program time $T[p]$, thus $F[p] := 1/T[p]$. This formulation fixes an average clock frequency $\bar{F}$ by having the programs executed under a constant computing time $t$. 

• The **Computing time**: $t$ counts how many times the average clock frequency $F$ must be doubled to double the entropy of the ensemble while holding the average tape usage $L$ and the average concurrency $N$ fixed.

2. **Size-cutoff formulation**:

$$Z'' : \mathbb{R}^2 \rightarrow \mathbb{R} \quad (k, l) \mapsto \sum_{p \in \{|q: \text{Dom}[UTM]| L[q] < l\}} 2^{-kL[p]} \tag{62}$$

The sum $Z''$ only includes programs with size less than or equal to $l$. $\Omega$ is recovered in the limit when $l \rightarrow \infty$ (and with $k = 1$).

3. **Time-cutoff formulation**:

$$Z''' : \mathbb{R}^2 \rightarrow \mathbb{R} \quad (k, t) \mapsto \sum_{p \in \{|q: \text{Dom}[UTM]| T[q] < t\}} 2^{-kL[p]} \tag{63}$$

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 with $k = 1$).

4. **Arbitrary filter cutoff formulation**:

Let $\mathcal{O} \subset \text{Dom}[UTM]$:

$$Z'''' : \mathbb{R} \rightarrow \mathbb{R} \quad k \mapsto \sum_{p \in \mathcal{O}} 2^{-kL[p]} \tag{64}$$

The sum only includes programs that halt further filtered by an arbitrary selection process $S : \mathcal{O} \rightarrow \text{Dom}[UTM]$.

So, how close are we to any real physics with this? Let us brainstorm:

1. **Feasible computing complexity**:

Usual 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.

A possible application of this framework is to construct a theory of feasible computational complexity. Indeed, 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. The grand canonical partition function of algorithmic thermodynamics has the following equation of state:

\[
dS = k d\bar{L} + f d\bar{T} + \mu d\bar{N}
\]  

(65)

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’.

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 \( Z(\beta) = \sum_{q \in Q} \exp(-\beta E[q]) \). It’s average energy is given by \( \bar{E} = -\partial \ln \ Z/\partial \beta \) and its fluctuations are \( \bar{(\Delta E)^2} = \partial^2 \ln Z/\partial \beta^2 \). 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’.

By taking algorithmic thermodynamics at face value, we have recovered a system of computation that maximizes the entropy over its domain of computation and subject to a variety of resource constraints. So far so good; but why not a quantum computation? Where is quantum mechanics, the qubit, the geometry of space-time... where is the richness of modern physics?

Quantum computations rests primarily on the idea that one can define a sequence of unitary operators such that each member of the sequence is usually (but not necessarily) associated with a computationally simple operation (often called a quantum gate). The complexity of the sequences one can form by combining these gates eventually allows one to perform arbitrary computations upon some initial state. The end result is constructed by measuring multiple copies (or re-runs) of the computation and taking an average over the observables.

No matter how much I played with and rearranged algorithmic thermodynamics, it seamed that the quantum computing description was outside its scope; or that if I ostentatiously tried to made it fit regardless, it had to be altered with such artificially that it would feel like I was just fixing the axioms to give me what I wanted to get in the first place.

Something exceedingly fundamental was surely missing.

5.4 Attempt 2: Designer Ensemble

My second series of attempts could be grouped under a simple concept: I 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.

In 2002, Lloyd\textsuperscript{30} 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, Lloyd obtains the number $\approx 10^{122}$. This number is consistent with other approaches; for instance, the Bekenstein-Hawking entropy\(^{31}\) of a 'holographic surface' at the cosmological horizon\(^{32}\) (also $\approx 10^{122}$).

How did Lloyd derive these numbers? First, he calculated the value for these quantities while ignoring the contribution of gravity and he obtained $\approx 10^{90}$. It is only by including the degrees of freedom of gravity that the number $\approx 10^{122}k_B$ is obtained, which he does in the second part of his paper. The main relation he obtains is:

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

where $\rho_c$ is the critical density and $t_p$ is the Planck time and $t$ is the age of the universe. With present-day values of $t$, the result is $\approx 10^{122}$. 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{2^t}{t_p} = \frac{t^2}{t_p^2}^\star."

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 summation over all available degrees of freedom in space-time). Interestingly these computational quantities are related to the square of $x$ and $t$, and thus grow as area laws.

A general 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\(^{33}\) 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\(^{34}\) 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: $S = k_B c^3 / (4\hbar G) A$.

We mention Ted Jacobson\(^{35}\) 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\(^3\) 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 \( T dS = F \, d\mathbf{x} \), 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\(^4\) 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\(^3\), the Coulomb force\(^3\), Yang-Mills gauge fields\(^3\), and cosmology\(^4\). Some criticism has, however, been voiced\(^4\) including by Visser\(^4\).

Even more recently, a connection between entanglement entropy and general relativity has been supported by multiple publications\(^5\),\(^6\),\(^7\),\(^8\),\(^9\),\(^10\),\(^11\),\(^12\),\(^13\),\(^14\),\(^15\),\(^16\),\(^17\),\(^18\),\(^19\),\(^20\),\(^21\),\(^22\),\(^23\),\(^24\),\(^25\),\(^26\),\(^27\),\(^28\),\(^29\),\(^30\),\(^31\),\(^32\),\(^33\),\(^34\),\(^35\),\(^36\),\(^37\),\(^38\),\(^39\),\(^40\),\(^41\),\(^42\),\(^43\),\(^44\),\(^45\),\(^46\),\(^47\),\(^48\),\(^49\),\(^50\),\(^51\),\(^52\),\(^53\),\(^54\),\(^55\),\(^56\),\(^57\),\(^58\),\(^59\),\(^60\),\(^61\),\(^62\),\(^63\),\(^64\),\(^65\),\(^66\),\(^67\).

I initially joined in to this effort. However, in the end, we felt that there was a general problem with this approach and I eventually scratched about 3 years of work in this direction.

5.5 Anti-pattern: axiomatic fixing

The problem with the second attempt, even if it successfully lead to some set of valid physical laws, is that any results would be specific to the constructed ensemble. With this approach, “ensemble-building” replaces “tower-building” but the same limitations applies. My choice of designer ensembles was contingent on my prior knowledge of the laws of physics, as provided to me by the experimental sciences and by available modern physics literature. That is; I knew before hand what I was supposed to get, therefore I had the opportunity to fix my axiomatic-basis/designer-ensemble based on this knowledge.

Since the dept of mathematical complexity (and flexibility) has no bounds, one will eventually be able to construct a fundamental basis for physics using almost any (Turing-complete) framework, if one tries hard enough. But will the basis be sound? Axiomatic fixing is a trap that occurs when one adjust and re-adjust a candidate fundamental basis until all known relevant experimental knowledge is derivable or integrated with it, but (and perhaps quite shockingly to the theorists who did the work) the basis fails to surviving falsifi-

---


cation each and every-time it is tested outside the domain that was initially available to it for fixing. Axiomatic fixing suggests that this, rather than being shocking, is actually nearly unavoidable.

Axiomatic fixing is an anti-pattern that has emerged in theoretical physics in a predominant manner over the last four decades or so; ever since the novelty of experimental particle physics has thinned out. Before then, scientific revolutions had the benefit of novel data available but not yet integrated within the the existing laws of physics. Thus, the existing laws were modified to account for this new incompatible data, which resulted in a new more fundamental basis. However, today, we have two theories (GR/QM) that are incompatible with one other, yet are compatible with all experimental data collected thus far. Unification ought to be possible, but since there is no experimental data available to reduce the search space, axiomatic fixing is the predominantly used anti-pattern.

To avoid this anti-pattern, one must derive the laws of physics in an incidental manner, for instance as incidental to the goal of deriving the probability measure that makes reality maximally informative. This goal can be reached without referencing the already known laws of physics. Following this goal I can thus claim; I am not trying to recover the laws of physics per se, instead I just want to maximize the information I can get out of reality. Then, any derivation of the laws of physics is incidental to my maximization procedure. Now, it may well be the case that I will derive the familiar laws of physics at the end of this process, nonetheless said laws would have been derived without prior axiomatic fixing. Finally, as the anti-pattern is avoided, the produced fundamental basis is likelier to be sound.

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

\[ dE = T dS + p dV - \mu dN \]  

To capture this generality, my retained solution was not to define a specific system of statistical physics (a.k.a. a designer ensemble), but instead to increase the generality of statistical physics such that the default probability measure automatically acquires the structure of the laws of physics, without effort.
6 Physics as the ultimate probability measure of reality

6.1 Intuition: Universal Thermodynamics

The entropy of a typical thermodynamic system is constrained by a set of quantities $E, V, N$, etc. Such a system will admit scalar macroscopic transformations over its constraints via an appropriate equation of state. For instance a change from state a to b involves a scalar equation of state:

\[
\begin{align*}
\text{state-a} & \quad \rightarrow \quad \text{state-b} \\
\begin{pmatrix}
E_a \\
V_a \\
N_a
\end{pmatrix} & \rightarrow T \, dE + p \, dV + \mu \, dN \\
\begin{pmatrix}
E_b \\
V_b \\
N_b
\end{pmatrix}
\end{align*}
\]  

(68)

During my exploratory attempts, I initially identified the potential to generalize thermodynamics as I attempted to create a thermodynamic system whose macroscopic transformations are consistent with the symmetries of space-time. By doing so, I realized that such transformations could be produced if the relevant thermodynamic constraints admitted linear transformations.

To achieve linear transformations in the equation of state, I have defined a thermodynamic system able to describe a transformation that is described by matrices. A transformation from state a to b would thus be made as follows:

\[
\begin{align*}
\text{state-a} & \quad \rightarrow \quad \text{state-b} \\
\begin{pmatrix}
X_{11}^a & \ldots & X_{1n}^a \\
\vdots & \ddots & \vdots \\
X_{n1}^a & \ldots & X_{nn}^a
\end{pmatrix} & \rightarrow \begin{pmatrix}
\tilde{k} \, dX_{11} & \ldots & \tilde{k} \, dX_{1n} \\
\vdots & \ddots & \vdots \\
\tilde{k} \, dX_{n1} & \ldots & \tilde{k} \, dX_{nn}
\end{pmatrix} \\
\begin{pmatrix}
X_{11}^b & \ldots & X_{1n}^b \\
\vdots & \ddots & \vdots \\
X_{n1}^b & \ldots & X_{nn}^b
\end{pmatrix}
\end{align*}
\]  

(69)

This extension, although on the face conceptually quite simple, has remarkable consequences.

6.2 The mathematical origin of the Born rule

The main result of universal statistical physics, is a mathematical derivation (and extension) of the Born rule. In fact, all systems of universal statistical physics produces an extended Born rule which is used to connect the domain of science (the set of all verifiable statements) to reality (the set of verified statements).
We recall that in ‘scalar/typical’ statistical physics, one obtains the Gibbs ensemble using the method of the Lagrange multipliers by solving for the probability measure which maximizes the entropy under a set of constraints. Taking $E$ as the constraint, the Lagrange equation would be:

$$L = -k_B \sum_{p \in P} \rho[p] \ln \rho[p] + \alpha_1 (-1 + \sum_{p \in P} \rho[p]) + \alpha_2 (-E + \sum_{p \in P} \rho[p]E[p])$$

(70)

where $\alpha_1, \alpha_2$ are the Lagrange multipliers. Then, extremalizing it, one obtains the Gibbs measure:

$$\frac{\partial L}{\partial \rho[p]} = 0 \Rightarrow \rho[p] = \frac{1}{Z} \exp(-\beta E[p])$$

(71)

We will now repeat the usual treatment of entropy in statistical physics, but now for a system of multiple Lagrange equations. We will show that an extended Born rule appears in the form of the partition function whenever we have a thermodynamic system described as a system of multiple Lagrange equations. The thermodynamic constraints will be the eigenvalues of a matrix: specifically, $n$ eigenvalues implies $n$ constraints which implies $n$ Lagrange equations to extremalize.

Since we will be using the eigenvalues of a matrix as the constraints, and such eigenvalues can be complex, the $\rho[p]$ which is usually a probability measure between zero and one and normalized to one, will here be changed to a complex amplitude whose sum over its domain is that of a finite complex value. We now define a few key quantities:

1. Let $M[p]$ be a $n \times n$ matrix-valued function from $P \rightarrow M(n, \mathbb{C})$
2. Let $\lambda_1[p], \ldots, \lambda_n[p]$ be the $n$ eigenvalues of $M[p]$.
3. Let $\rho_1[p], \ldots, \rho_n[p]$ be function $P \rightarrow \mathbb{C}$, each called a complex amplitude, normalized to a finite complex value $\sum_{p \in P} \rho_1[p] = A_1, \ldots, \sum_{p \in P} \rho_n[p] = A_n$.
4. Let each eigenvalues of $M[p]$ be a thermodynamic constraint; $\sum_{p \in P} \rho_1[p] \lambda_1[p] = \lambda_1, \ldots, \sum_{p \in P} \rho_n[p] \lambda_n[p] = \lambda_n$

Now, for each of $n$ eigenvalues, we can define a Lagrange equation as follows:
The universe exists as indubitable structured information.

\[ L_1 = -k_B \sum_{p \in \mathcal{P}} \rho_1[p] \ln \rho_1[p] + \alpha_1(-1 + \sum_{p \in \mathcal{P}} \rho_1[p]) + \alpha_2(-\overline{\lambda}_1 + \sum_{p \in \mathcal{P}} \rho_1[p] \lambda_1[p]) \]

(72)

\[ L_n = -k_B \sum_{p \in \mathcal{P}} \rho_n[p] \ln \rho_n[p] + \alpha_1(-1 + \sum_{p \in \mathcal{P}} \rho_n[p]) + \alpha_2(-\overline{\lambda}_n + \sum_{p \in \mathcal{P}} \rho_n[p] \lambda_n[p]) \]

(73)

We can extremize each Lagrange equations individually. To do so we take an element \( q \) in \( \mathcal{P} \) and we take the partial derivative of \( L \) with respect to \( \rho_1[q] \):

\[ 0 = \frac{\partial L_1}{\partial \rho_1[q]} = -k_B - k_B \ln \rho_1[q] + \alpha_1 + \alpha_2 \lambda_1[q] \]

(74)

\[ 0 = \frac{\partial L_n}{\partial \rho_n[q]} = -k_B - k_B \ln \rho_n[q] + \alpha_1 + \alpha_2 \lambda_n[q] \]

(75)

Solving for \( \rho_1[q] \), we obtain

\[ \rho_1[p] = \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) \exp\left(\frac{\alpha_2}{k_B} \lambda_1[p]\right) \]

(76)

\[ \vdots \]

\[ \rho_n[p] = \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) \exp\left(\frac{\alpha_2}{k_B} \lambda_n[p]\right) \]

(77)

Using the normalization constraint on \( \rho[p] \), we find \( n \) ‘eigen-partition function’:

\[ 1 = \frac{1}{A_1} \sum_{p \in \mathcal{P}} \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) \exp\left(\frac{\alpha_2}{k_B} \lambda_1[q]\right) \Rightarrow \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) = \left(\frac{1}{A_1} \sum_{p \in \mathcal{P}} \exp\left(\frac{\alpha_2}{k_B} \lambda_n[p]\right)\right)^{-1} \]

(78)

\[ \vdots \]

\[ 1 = \frac{1}{A_n} \sum_{p \in \mathcal{P}} \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) \exp\left(\frac{\alpha_2}{k_B} \lambda_n[q]\right) \Rightarrow \exp\left(-\frac{k_B + \alpha_1}{k_B}\right) = \left(\frac{1}{A_n} \sum_{p \in \mathcal{P}} \exp\left(\frac{\alpha_2}{k_B} \lambda_n[p]\right)\right)^{-1} \]

(79)

Finally, the extended Born rule appears, from universal thermodynamics, by multiplying the eigen-partition functions.
Definition 10 (Extended Born rule (diagonal case)). We multiply the eigen-partition functions:

\[ \|Z\| = Z_1 Z_2 \ldots Z_n \] (80)

Let us now apply this result to a few examples:

Theorem 1 (Scalar Thermodynamics). In the case where \( M[p] \) is a \( 1 \times 1 \) matrix, one recovers the scalar partition function of usual statistical physics.

Proof. Trivial \( \square \)

Theorem 2 (Grand-partition function). In the case where \( M[p] \) is the product of a \( n \)-dimensional identity matrix \( I_n \) and a scalar constraint, then the eigen-partition function multiplication produces a grand partition function of identical particles.

Proof. Let \( M[p] := \beta E[p] I_n \), where \( E : \mathbb{P} \to \mathbb{R} \) and where \( I_n \) is the \( n \)-dimension identity matrix. In this case \( M[p] \) is:

\[
M[p] = \begin{pmatrix}
-\beta E[p] & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & -\beta E[p]
\end{pmatrix}
\] (81)

And the partition function is:

\[
\|Z\| = \left( \sum_{p \in \mathbb{P}} \exp \left( -\beta E[p] \right) \right)^n
\] (82)

Theorem 3 (Quantum probabilities (diagonal case)). In the case where \( M[p] \) is the matrix representation of the complex numbers, one recovers the familiar probability amplitude and Born rule of quantum mechanics.

Proof. Let us now show that quantum probabilities are a special case of the extended Born rule. Let \( \mathbb{P} := \{p_1, p_2\} \). We also use the maps \( r : \mathbb{P} \to \mathbb{R} \) and \( \theta : \mathbb{P} \to \mathbb{R} \) as the matrix entries, as follows:

\[
M'[p] = \begin{pmatrix}
r[p] & \theta[p] \\
-\theta[p] & r[p]
\end{pmatrix}
\] (83)

We note that here \( M[p] \) is a matrix representation of the complex numbers, via the group isomorphism of \( a + ib \cong \left( \begin{array}{cc}
a & b \\
b & -a
\end{array} \right) \). The eigenvalues of \( M[p] \) are:
\begin{align*}
\lambda_1 &= r[p] + i\theta[p] \\
\lambda_2 &= r[p] - i\theta[p]
\end{align*} (84)

The universal Born rule becomes:

\[||Z|| = \left( \sum_{p \in \mathcal{P}} \exp(r[p] + i\theta[p]) \right) \left( \sum_{p \in \mathcal{P}} \exp(r[p] - i\theta[p]) \right)\] (86)

If we now take a two state ensemble \(\mathcal{P} := \{p_1, p_2\}\), we get:

\[||Z|| = \left( \exp(r[p_1] + i\theta[p_1]) + \exp(r[p_2] + i\theta[p_2]) \right) \left( \exp(r[p_1] - i\theta[p_1]) + \exp(r[p_2] - i\theta[p_2]) \right)\] (87)

With straightforward algebraic manipulation and simplifications, we get:

\[= (e^{r[p_1]})^2 + (e^{r[p_1]})^2 + 2e^{r[p_1]}e^{r[p_2]} \cos[\theta[p_2] - \theta[p_1]]\] (88)

Finally, by posing \(|\psi_1| := e^{r[p_1]}, |\psi_2| := e^{r[p_2]}, \varphi_1 = \theta[p_1]\) and \(\varphi_2 = \theta[p_2]\) we get the typical quantum probability of a two-state system along with the interference term in its usual form:

\[= |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \cos[\varphi_2 - \varphi_1]\] (89)

\[\square\]

In statistical physics, the partition function contains, by definition and within the sum, all possible states of the system. When we multiply \(n\) identical partition functions to obtain a grand-partition function (theorem 2), the result is a single partition function that sums over the cross product of the microstates of each individual partition function. For instance, a partition function over a set \(\mathcal{X}\) will, when multiplied by a partition function of a set \(\mathcal{Y}\), produce a grand-partition function over the set \(\mathcal{X} \times \mathcal{Y}\). In all three cases the partition functions \(Z[\mathcal{X}], Z[\mathcal{Y}]\) and \(Z[\mathcal{X} \times \mathcal{Y}]\) produce a constant numerical value that can be used to normalize a Gibbs probability measure. However, in theorem 3, the resulting partition function—as it is produced from the product of two non-identical partition function—contains a new free variable \(\cos[\theta[p_2] - \theta[p_1]]\) and thus it is not a constant numerical value. We have in fact obtained an infinite continuum of partition functions.

One can then wonder, if we perform the experimental realization of such a multiplication, which of the infinitely many partition functions will the system produce? Which one will we observed in the
lab? We can assign the neutral probability measure as follow; If we assume that each partition functions is statistically weighted according to its numerical value, but are otherwise equally likely, then each partition function correspond to an element of a Hilbert space, and quantum mechanics is recovered. One integrates all possible partition functions over the emergent free variables to obtain the normalization constant. The general result of this integral is:

$$\int Z[X \times Y] dx < \infty$$ (90)

which in the case of the matrix representation of the complex, is:

$$\int Z[X]Z[X]^* dx < \infty$$ (91)

The multiplication of the partition functions of a system of universal thermodynamic are elements normalizable within a Hilbert space. Thus, partition functions of universal thermodynamics are quantum mechanical states.

**Theorem 4 (Path integral).** If $M[p]$ is the matrix representation of the imaginary numbers, then $\|Z\|$ can be equated to the path integral by the usual limiting process.

**Proof.** Let

$$M[p] = \begin{pmatrix} 0 & -S[p] \\ S[p] & 0 \end{pmatrix}$$ (92)

its eigenvalues are:

$$\lambda_1[p] = -iS[p]$$ (93)
$$\lambda_2[p] = +iS[p]$$ (94)

Posing $\alpha_2/k_B \to 1/h$, the extended Born rule produces

$$\|Z\| = \left(\frac{1}{A_1} \sum_{p \in P} \exp(iS[p]/\hbar)\right) \left(\frac{1}{A_2} \sum_{p \in P} \exp(-iS[p]/\hbar)\right)$$ (95)

Then, by taking the set $P$ to be infinite, and $S[p]$ to be a sum of infinitesimals $S[p] := S_1[p] + S_2[p] + S_3[p] + \ldots$, then we get

$$\|Z\| = \left(\int dp \exp\left(\frac{i}{\hbar} \int L[t] dt\right)\right) \left(\int dp \exp\left(-\frac{i}{\hbar} \int L[t] dt\right)\right)$$ (96)

provided that the modulus squared of the complex amplitudes is normalized: $\frac{1}{A_1} \frac{1}{A_2} = 1$.

Now, lets make this interesting by kicking it up a notch:
6.3 The connection between science and reality

The probability of a system of universal thermodynamics to occupy a microstate is given by a generalization of the Born probability rule, which extends classical probabilities to matrices (including non-diagonal matrices):

**Definition 11** (Extended Born rule). Let $\mathcal{P}$ be a countable set and let $M[p]$ be a $n \times n$ matrix. We define the extended Born rule as follows:

$$\|Z\| = \det \sum_{p \in \mathcal{P}} \exp M[p]$$  \hspace{1cm} (97)

In the case where $M[p]$ is diagonal, we recover definition 10.

**Theorem 5** (Quantum probabilities). In the case where $M[p]$ is the matrix representation of the complex numbers, one recovers the familiar probability amplitude and Born rule of quantum mechanics, even if $M[p]$ is non-diagonal.

**Proof.** Let $\mathcal{P} := \{p_1, p_2\}$. We also use the maps $r : \mathcal{P} \rightarrow \mathbb{R}$ and $\theta : \mathcal{P} \rightarrow \mathbb{R}$ as the matrix entries, as follows:

$$M'[p] = \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix}$$  \hspace{1cm} (98)

We note that here $M[p]$ is a matrix representation of the complex numbers, via the group isomorphism of $a + ib \cong \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. The universal Born rule becomes:

$$\|Z\| = \det \sum_{p \in \mathcal{P}} \exp \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix}$$  \hspace{1cm} (99)

(100)

The matrix exponential reduces to the following expression:

$$\exp \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix} = \begin{pmatrix} i \sqrt{\frac{1}{2}} & -i \sqrt{\frac{1}{2}} \\ i \sqrt{\frac{1}{2}} & i \sqrt{\frac{1}{2}} \end{pmatrix} \exp \begin{pmatrix} r[p] - i\theta[p] & 0 \\ 0 & r[p] + i\theta[p] \end{pmatrix} \begin{pmatrix} i \sqrt{\frac{1}{2}} & i \sqrt{\frac{1}{2}} \\ i \sqrt{\frac{1}{2}} & i \sqrt{\frac{1}{2}} \end{pmatrix}$$  \hspace{1cm} (101)

$$= \begin{pmatrix} e^{r[p]} \cos[\theta[p]] & e^{r[p]} \sin[\theta[p]] \\ -e^{r[p]} \sin[\theta[p]] & e^{r[p]} \cos[\theta[p]] \end{pmatrix}$$  \hspace{1cm} (102)

If we now take a two state ensemble $\mathcal{P} := \{p_1, p_2\}$, we get:
\|Z\| = \det \left[ \begin{pmatrix} e^{i[p_1]} \cos[\theta[p_1]] & e^{i[p_1]} \sin[\theta[p_1]] \\ -e^{i[p_1]} \sin[\theta[p_1]] & e^{i[p_1]} \cos[\theta[p_1]] \end{pmatrix} + \begin{pmatrix} e^{i[p_2]} \cos[\theta[p_2]] & e^{i[p_2]} \sin[\theta[p_2]] \\ -e^{i[p_2]} \sin[\theta[p_2]] & e^{i[p_2]} \cos[\theta[p_2]] \end{pmatrix} \right] \\
(103)

With straightforward algebraic manipulation and simplifications, we get:
\[
= (e^{i[p_1]})^2 + (e^{i[p_1]})^2 + 2e^{i[p_1]}e^{i[p_2]} \cos[\theta[p_2] - \theta[p_1]] \\
(104)
\]

Finally, by posing \(|\psi_1| := e^{i[p_1]}, |\psi_2| := e^{i[p_2]}, \varphi_1 = \theta[p_1] and \varphi_2 = \theta[p_2]\) we get the typical quantum probability of a two-state system along with the interference term in its usual form:
\[
= |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \cos[\varphi_2 - \varphi_1] \\
(105)
\]

Our goal now is to extend this methodology to matrices having even more structure than just the complex numbers. One can interpret the previous result in the usual sense that \(M[p]\) is the matrix representation of a complex number. However, it is also possible to interpret \(M[p]\) as the matrix representation of the even sub-algebra of \(G_2(\mathbb{R})\) (which is group isomorphic to the complex). In this interpretation the complex number are a geometric object (one of many possible geometric objects), and the extended Born rule is simply an extension of the Born to any geometric object. As we will see with the next theorem, the second interpretation can be extended to a probability rule having an even richer structure. Let us now take the even sub-algebra of a geometric algebra of higher dimensions.

As we said, in the case of \(G_2(\mathbb{R})\), an element of the even sub-algebra is:
\[
v = r + \theta I \\
(106)
\]
which is group isomorphic with the complex. For \(G_{3,1}(\mathbb{R})\), an element of the even sub-algebra is:
\[
v = r + \mathbf{F} + \theta I \\
(107)
\]
where \(\mathbf{F}\) is a bivector.

The extended Born rule applied to the matrix representation of the even sub-algebra element leads us directly to the relativistic wavefunction, formulated in the language of geometric algebra as suggested by David Hestenes:

\[ \psi = \exp \left( \frac{1}{2} (r + F + \theta I) \right) = R \exp \left( \frac{1}{2} (r + \theta I) \right) \]  
(108)

where \( R = \exp F/2 \) is a rotor.

**Theorem 6** (Quantum wavefunction (4D)). *We now use the matrix representation of the even subalgebra of the Clifford algebra in 3+1 spacetime.*

**Proof.** The basis used for the matrix representation of a complete even multi-vector of \( \mathcal{G}_{3,1}(\mathbb{R}) \), expressed in terms of the Dirac matrix, is:

\[
I = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
\end{pmatrix}
\]  
(109)

\[
s_{01} = \begin{pmatrix}
0 & 0 & 0 & -1 \\
0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
\end{pmatrix} \quad s_{02} = \begin{pmatrix}
0 & 0 & i & 0 \\
0 & -i & 0 & 0 \\
i & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix} \quad s_{03} = \begin{pmatrix}
0 & 0 & -1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
\end{pmatrix} \quad s_{12} = \begin{pmatrix}
i & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
\end{pmatrix} \quad s_{31} = \begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\]  
(110)

\[
I = \begin{pmatrix}
0 & 0 & -i & 0 \\
0 & 0 & 0 & -i \\
-i & 0 & 0 & 0 \\
0 & -i & 0 & 0 \\
\end{pmatrix}
\]  
(111)

Using this basis, the matrix representation of the most general even multi-vector of \( \mathcal{G}_{3,1}(\mathbb{R}) \) is:

\[
M[p] = r[p]I + F_{01}[p]s_{01} + F_{02}[p]s_{02} + F_{03}[p]s_{03} + F_{23}[p]s_{23} + F_{13}[p]s_{13} + F_{12}[p]s_{12} + \theta[p]I
\]  
(112)

\[
= \begin{pmatrix}
\end{pmatrix}
\]  
(113)

One then completes the Born rule by taking the determinant of the matrix representation of the even element. The key identity is:

\[
\det \exp \frac{1}{2} M[p] = \exp \text{Tr} \frac{1}{2} M[p] = \exp 2r[p]
\]  
(114)

Which yields the same result as in the complex case. \( \Box \)

The extended Born rule then automatically cancels out the rotor (via the relation \( R\tilde{R} = 1 \)) as well as the complex part (via the relation the square modulus) and maps \( \psi \) directly to a real probability.
value. This cancellation implies a prior automatic inclusion of the space-time geometric structure of the wavefunction as part of the extended Born rule. Here the wavefunction is a natural consequence of applying the definition of information (via entropy) to geometry (geometric algebra represented by matrices). Its meaning as a ‘wave of probability’ is cemented by this fundamental relation to universal statistical physics.

The essential insight to this series of result is that the entropy maximization procedure applies the exponential map to $M(n, C)$. The effect is to reduce the microscopic domain to the set of all general linear matrices via the well-known correspondence:

$$\exp : M(n, C) \rightarrow GL(n, C)$$ (115)

that maps exponentials of arbitrary matrices to general linear transformations. The exponential map is the minimum “filter” required such that each element has an inverse due to the identity $\exp M \exp - M = I$. Consequently, arbitrary macroscopic transformations are mapped to the general linear group in the microscopic sector of the ensemble.

We have previously investigated the role of the even sub-algebra of $G_4(R)$ and we have seen how this recovers the wavefunction in 3+1 spacetime as a pure probabilistic object; the geometric amplitude. Including the odd algebra terms so as to produce a complete multi-vector simply extends the rotor component to the group of versors which will now accounts for all possible Lorentz transformation, including reflections and inversions.

**Theorem 7** (Extended Born rule for a complete multi-vector of $G_4(R)$).

*Proof.* The inclusion of the odd part to the microscopic element allows us to express all possible Lorentz transformations, extending the rotors made available by the even sub-algebra to now include those of space and time inversions. The extended Born rule is able to account for this general case. It is in this case that 4 unique eigenvalues are produced.

Let us choose the geometric algebra $G_4(C)$ as the representation for $M(4, C)$ matrices. With it we can associate the determinant of a general $4 \times 4$ matrix to a universal norm of space-time. To support the applicability of this choice, we will rely on the fact that the matrix representation of geometric algebra $G_4(C)$ comprises the full set of $M(4, C)$; the set of $4 \times 4$ matrix with complex entries.

First, let us note that the Dirac matrices form the generators of the basis of $M(4, C)$. There are 16 elements of the basis:
1. The identity matrix $I$

2. Four matrix $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$.

3. Six matrix $\sigma_{\mu\nu} = -\frac{1}{2}(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu)$

4. One pseudoscalar matrix $\gamma_5 = \gamma_0 \gamma_1 \gamma_2 \gamma_3$

5. Four matrix $\gamma_\mu = \gamma_5 \gamma_\mu$

(where $\gamma_a \gamma_b$ is the usual matrix multiplication.)

Explicitly, the 16 matrices are:

\begin{align*}
I &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
\gamma_0 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\
\gamma_1 &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \\
\gamma_2 &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \\
\gamma_3 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\
\sigma_{01} &= \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\sigma_{02} &= \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & -i \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\sigma_{03} &= \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & -i \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\sigma_{12} &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\sigma_{31} &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
v_0 &= \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ -1 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \end{pmatrix} \\
v_1 &= \begin{pmatrix} 0 & i & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & 1 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} \\
v_2 &= \begin{pmatrix} 0 & i & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} \\
v_3 &= \begin{pmatrix} 0 & i & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 1 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}
\end{align*}

One can write any matrix $\mathbf{M} \in \mathbb{M}(4, \mathbb{C})$ using a linear combination of these 16 matrices over the complex (parenthesis added for clarity):

\begin{align*}
\mathbf{M} &= (X_0)\gamma_0 + (X_1)\gamma_1 + (X_2)\gamma_2 + (X_3)\gamma_3 \\
&+ (E_1)v_0 + (E_2)v_0 + (E_3)v_0 + (B_1)v_0 + (B_2)v_0 + (B_3)v_0 \\
&+ (V_0)v_0 + (V_1)v_0 + (V_2)v_0 + (V_3)v_0 \\
&+ (a) + (R)\gamma_5
\end{align*}

Likewise, one can write any multivector $\mathbf{u} \in \mathcal{G}_4$ using a linear combination of the 16 basis elements of $\mathcal{G}_4$, also over the complex, as:

\begin{align*}
\mathbf{u} &= (X_0)\hat{x}_0 + (X_1)\hat{x}_1 + (X_2)\hat{x}_2 + (X_3)\hat{x}_3 \\
&+ (E_1)\hat{x}_0 \wedge \hat{x}_1 + (E_2)\hat{x}_0 \wedge \hat{x}_2 + (E_3)\hat{x}_0 \wedge \hat{x}_3 + (B_1)\hat{x}_1 \wedge \hat{x}_3 + (B_2)\hat{x}_2 \wedge \hat{x}_3 + (B_3)\hat{x}_1 \wedge \hat{x}_2 \\
&+ (V_0)\hat{x}_1 \wedge \hat{x}_2 \wedge \hat{x}_3 + (V_1)\hat{x}_0 \wedge \hat{x}_2 \wedge \hat{x}_3 + (V_2)\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_3 + (V_3)\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 \\
&+ (a) + (R)\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 \wedge \hat{x}_3
\end{align*}
There exists an bijective map between the elements of $\mathcal{G}_4(\mathbb{C})$ and those of $\mathcal{M}(4, \mathbb{C})$:

$$M = M[g] \quad \quad M^{-1}[M] = u \quad (123)$$

such that the matrix multiplication of one is the geometric product of the other:

$$M^2 = M[u^2] \quad (124)$$

The map is realized by replacing the basis $\hat{x_i}$ by the gamma matrix $\gamma_i$, and vice-versa. The two representation are group isomorphic over the multiplication.

We now reduce the domain to the reals, as we take the microscopic element as the exponential of a complete multi-vector of $\mathcal{G}_4(\mathbb{R})$:

$$\psi[p] = e^{r[p]} e^{\theta[p]} \exp(X + F + V) = e^{r[p]} e^{\theta[p]} L \quad (125)$$

where $F$ is the previously defined bivector, and where:

$$X = X_0\hat{x}_0 + X_1\hat{x}_1 + X_2\hat{x}_2 + X_3\hat{x}_3 \quad (126)$$

$$V = V_0\hat{x}_1 \wedge \hat{x}_2 \wedge \hat{x}_3 + V_1\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 + V_2\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_3 + V_3\hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 \quad (127)$$

This expression, as it is an exponential, is reversible, and consequently represents the versors $L$, and the transformations are done by sandwiching: $\mathbf{x}' = LxL^{-1}$, where $LL^{-1} = 1$. In the matrix representation, $M[p]$ is:

$$M[p] = \begin{pmatrix}
    a + X_0 - iB_3 - iV_3 & B_2 - iB_1 + V_2 - iV_1 & -ib + X_3 + E_3 - iV_0 & X_1 - iX_2 + E_1 - iE_2 \\
    -B_2 - iB_1 - V_2 - iV_1 & a + X_0 + iB_3 + iV_3 & X_1 + iX_2 + iE_1 + E_2 & -ib - X_3 + E_3 - iV_0 \\
    -ib - X_3 - E_3 + iV_0 & X_1 + iX_2 + iE_1 - iE_2 & a - X_0 - iB_3 + iV_3 & B_2 - iB_1 - V_2 + iV_1 \\
    -X_1 - iX_2 + E_1 + iE_2 & -ib + X_3 - E_3 + iV_0 & -B_2 - iB_1 + V_2 + iV_1 & a - X_0 + iB_3 - iV_3
\end{pmatrix} \quad (128)$$

As before, the det of exp is related to the trace:

$$\text{det} \exp \frac{1}{2} M[p] = \exp \text{Tr} \frac{1}{2} M[p] = e^{2r[p]} \quad (129)$$

In fact, if we consider the constraint of a system of universal thermodynamics to be an arbitrary matrix $\mathcal{M}(n, \mathbb{C})$, and possibly even with the subdomain $\mathcal{M}[\mathcal{G}_4(\mathbb{R})]$, we get a no-go theorem regarding the number of dimensions the system can have:
Theorem 8 (Loss of structure beyond 4 space-time dimensions (no-go theorem)).: If;

1. we attribute physical significance to the eigenvalues of the matrix representation of multi-vectors (such as; for the construction of a thermodynamic system of equations, for a change of basis, etc.), and;

2. we require the laws of physics to be expressible as general solutions in radicals, and;

3. we require the laws of physics to remain invariant with respect to a change of numerical value within the entry of the matrix representing the system (Lorentz invariance, coordinate-change invariance, etc), and;

4. we require the matrix representation to be square so as to be able to use the determinant, and

5. we consider a system of universal thermodynamics constrained to an arbitrary $M(n, C)$ matrix,

then for a general/arbitrary matrix, the dimensions stops at $4 \times 4$ because of the Abel–Ruffini theorem.

Proof. We note:

1. The Abel–Ruffini theorem states that there exists no solutions in radicals to a general polynomial equation of degree 5 or higher with arbitrary coefficient.

2. Obtaining the eigenvalues of a $n \times n$ matrix requires one to solve the roots of its characteristic polynomial.

3. The characteristic polynomial, for a $n \times n$ matrix with arbitrary coefficient is of degree up to $n$.

4. The general multi-vectors of $G_4(C)$ form a complete representation of any elements of $M(4, C)$.

Then, it follows that the characteristic polynomial associated with the matrix representation of $4 \times 4$ matrices is a general polynomial of degree 4 with arbitrary coefficient. It further follows that since above 4 dimensions, one requires a matrix representation higher than $4 \times 4$, the corresponding characteristic polynomial will be of degree 5 or higher and will have no general solutions expressible in radicals. Thus, with the extended Born rule, it follows that no wavefuntions (defined as, roughly, an information bearing single invariant equation expressible in radicals) can exist beyond 4 dimensions. The extended Born rule, together with the correspondence between $M(4, C)$ and $G_4(C)$ —allowing for a purely geometric interpretation of $M(4, C)$—,
produces this no-go theorem beyond 4 space-time dimensions. In the language of this paper, we will say that microstate of the eigen-partition functions have no general structure beyond four dimensions. Thus, beyond four dimensions the informational backbone of the wavefunction fails.

We reiterate of course that if one allows eigenvalues not expressible in radicals in the definition of the constraints of the entropy, then the no-go theorem fails.

We note that in the case where we use only a subset of the $G_4(C)$ algebra, we can obtain radical solutions for the roots of a system of more than 4 dimensions. For instance, if we take a 1-vector of $G_4(C)$:

$$v = X_0 \hat{x}_0 + X_1 \hat{x}_1 + X_2 \hat{x}_2 + X_3 \hat{x}_3$$  (130)

then the characteristic polynomial of its matrix representation has reduced complexity and is not an arbitrary polynomial of degree 4.

The matrix representation of $v$ is:

$$M[v] = \begin{pmatrix}
X_0 & 0 & X_3 & X_1 - iX_2 \\
0 & X_0 & X_1 + iX_2 & -X_3 \\
-X_3 & -X_1 + iX_2 & -X_0 & 0 \\
-X_1 - iX_2 & X_3 & 0 & -X_0
\end{pmatrix}$$  (131)

The characteristic polynomial $\det [M[v] - \lambda I] = 0$ reduces to:

$$\lambda^2 = X_0^2 - X_1^2 - X_2^2 - X_3^2$$  (132)

which is a polynomial of degree 2. Thus, special relativity, by itself, does not limit space-time to 4 dimensions.

The key to limiting the dimensionality of spacetime to 4 is to consider the constraint to include all geometric degrees of freedom of spacetime (totalling 16 geometric degrees of freedom in the case of 3+1 space-time), which is enough for the characteristic polynomial to be an arbitrary polynomial of degree 4 and, consequently, to barely fly under the radar of the Abel-Ruffini theorem.

6.4 Space-time as an interference pattern

Here, we investigate new classes of interference patterns produced by the extended Born rule.

Theorem 9 (Hyperbolic interference (1D)). Now we will apply the geometric properties of the matrix representation of $G_2(R)$ to the universal Born rule:
Taking an ensemble of two elements

\[
\mathbf{I} \cong \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \hat{x}_1 \cong \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \hat{x}_2 \cong \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad I \cong \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (133)
\]

For the 1D case, we will only use \( I \) and \( \hat{x}_1 \). In this case, we obtain an interference pattern that uses the \( \cosh \) instead of the cos.

**Proof.** To start, we define two maps:

\[
r : \mathbb{P} \to \mathbb{R} \\
x : \mathbb{P} \to \mathbb{R}
\]

And we construct \( \mathbf{M}[p] \) as the sum \( r[p] \mathbf{I} + x[p] \hat{x}_1 \):

\[
\mathbf{M}[p] = r[p] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x[p] \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} r[p] + x[p] & 0 \\ 0 & r[p] - x[p] \end{pmatrix}
\]

The universal Born rule is:

\[
||\mathbf{Z}|| = \det \sum_{p \in \mathbb{P}} \begin{pmatrix} \exp(r[p] + x[p]) & 0 \\ 0 & \exp(r[p] - x[p]) \end{pmatrix}
\]

\[
= \left( \sum_{p \in \mathbb{P}} \exp(r[p] + x[p]) \right) \left( \sum_{p \in \mathbb{P}} \exp(r[p] - x[p]) \right)
\]

Taking an ensemble of two elements \( \mathbb{P} := \{ p_1, p_2 \} \) we get:

\[
= \left( \exp(r[p_1] + x[p_1]) + \exp(r[p_2] + x[p_2]) \right) \left( \exp(r[p_1] - x[p_1]) + \exp(r[p_2] - x[p_2]) \right)
\]

\[
= \exp(r[p_1] + x[p_1]) \exp(r[p_1] - x[p_1]) + \exp(r[p_2] + x[p_2]) \exp(r[p_2] - x[p_2]) + \exp(r[p_1] + x[p_1]) \exp(r[p_2] - x[p_2]) + \exp(r[p_2] + x[p_2]) \exp(r[p_1] - x[p_1])
\]

\[
= (e^{r[p_1]} + e^{r[p_2]})^2 + 2e^{r[p_1]}e^{r[p_2]} \cosh(x[p_1] - x[p_2])
\]

Here, we obtain a hyperbolic interference term in lieu of the cosine term normally present in quantum mechanics. We note that this probability becomes classical when \( x[p_1] - x[p_2] = 0 \) because \( \cosh 0 = 1 \). Indeed, with \( \cosh 0 \) the probability is classical:

\[
||\mathbf{Z}||^2 = \psi_1^2 + \psi_2^2 + 2|\psi_1||\psi_2| \cosh 0
\]

\[
\implies ||\mathbf{Z}|| = |\psi_1| + |\psi_2|
\]

\[\square\]
Theorem 10 (Geometric interference (2D+)). Let us repeat the same exercise, but this time we will use two dimensions. We will show that the interference pattern references the inner product between two vectors.

Proof. To start, we define three maps:

\[ r : \mathcal{P} \to \mathbb{R} \quad (144) \]
\[ x : \mathcal{P} \to \mathbb{R} \quad (145) \]
\[ y : \mathcal{P} \to \mathbb{R} \quad (146) \]

The resulting matrix is obtained by summing:

\[ M[p] = r[p]I + x[p]\hat{x}_1 + y[p]\hat{x}_2 \quad (147) \]

we get:

\[ M[p] = \begin{pmatrix} r[p] + x[p] & y[p] \\ y[p] & r[p] - x[p] \end{pmatrix} \quad (148) \]

The universal Born rule is:

\[ \|Z\| = \det \sum_{p \in \mathcal{P}} \exp \left( \frac{r[p] + x[p]}{y[p]} y[p] \right) \quad (149) \]

With straightforward algebraic manipulations (omitted), the exponentiation yields:

\[ \|Z\| = \det \sum_{p \in \mathcal{P}} e^{r[p]} \begin{pmatrix} \cosh \sqrt{x[p]^2 + y[p]^2} + \frac{x[p]\sinh \sqrt{x[p]^2 + y[p]^2}}{\sqrt{x[p]^2 + y[p]^2}} & \frac{y[p]\sinh \sqrt{x[p]^2 + y[p]^2}}{\sqrt{x[p]^2 + y[p]^2}} \\ \frac{y[p]\sinh \sqrt{x[p]^2 + y[p]^2}}{\sqrt{x[p]^2 + y[p]^2}} & \cosh \sqrt{x[p]^2 + y[p]^2} - \frac{x[p]\sinh \sqrt{x[p]^2 + y[p]^2}}{\sqrt{x[p]^2 + y[p]^2}} \end{pmatrix} \quad (150) \]

If we take a two-state system \( \mathcal{P} = \{p_1, p_2\} \), the again with straightforward algebraic manipulations (omitted), (and by posing \( x_i = x[p_i], y_i = y[p_i] \)) we obtain:

\[ \|Z\| = \|\psi_1\|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \left( \cosh \sqrt{x_1^2 + y_1^2} \cosh \sqrt{x_2^2 + y_2^2} - \frac{(x_1x_2 + y_1y_2) \sinh \sqrt{x_1^2 + y_1^2} \sinh \sqrt{x_2^2 + y_2^2}}{\sqrt{x_1^2 + y_1^2} \sqrt{x_2^2 + y_2^2}} \right) \quad (151) \]

where \( x_1x_2 + y_1y_2 \) is an inner product between the two states. We make a number of observations:
1. No geometry: When the geometry is annulled by posing \( \sqrt{x_1^2 + y_1^2} = 0 \) and \( \sqrt{x_2^2 + y_2^2} = 0 \), then \( M[p] \) reduces to the scalar matrix \( M[p] = a[p]I \) and the probability becomes classical:

\[
\|Z\| = |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \quad \Rightarrow \quad P = |\psi_1| + |\psi_2| \quad (152)
\]

2. When the two-states are parallel, the probability reduces to the 1D case involving hyperbolic interference:

\[
\|Z\| = |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \cosh \left( \sqrt{x_1^2 + y_1^2} - \sqrt{x_2^2 + y_2^2} \right) \quad (153)
\]

3. When the two-states are orthogonal, the inner product becomes 0, and the probability reduces to:

\[
\|Z\| = |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \cosh \sqrt{x_1^2 + y_1^2} \cosh \sqrt{x_2^2 + y_2^2} \quad (154)
\]

Then, finally, to a space-time event.

**Definition 12** (Probabilities of space-time events). Let us now repeat the same exercise, but with the gamma matrices and for a paravector. The gamma matrices along with the identity matrix produces the following basis:

\[
\begin{align*}
I &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
\gamma_0 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\
\gamma_1 &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\gamma_2 &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix} \\
\gamma_3 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}
\end{align*}
\]

One introduces fours maps:

\[
\begin{align*}
r : \mathbb{P} &\to \mathbb{R} \quad (156) \\
X_0 : \mathbb{P} &\to \mathbb{R} \quad (157) \\
X_1 : \mathbb{P} &\to \mathbb{R} \quad (158) \\
X_2 : \mathbb{P} &\to \mathbb{R} \quad (159) \\
X_3 : \mathbb{P} &\to \mathbb{R} \quad (160)
\end{align*}
\]

Then, the matrix \( M[p] \) is:

\[
M[p] = r[p]I + X_0[p]\gamma_0 + X_1[p]\gamma_1 + X_2[p]\gamma_2 + X_3[p]\gamma_3 \quad (161)
\]

and its representation is:
Using the matrix representation leads to a substantially verbose proof. Instead, we will remain in the language of geometric algebra. We can write \( \exp M[p] \) as:

\[
M[p] = \begin{pmatrix}
  r + X_0 & 0 & X_3 & X_1 - iX_2 \\
  0 & r + X_0 & X_1 + iX_2 & -X_3 \\
  -X_3 & -X_1 + iX_2 & r - X_0 & 0 \\
  -X_1 - iX_2 & X_3 & 0 & r - X_0
\end{pmatrix}
\]

(162)

Now, we construct an ensemble of two states \( \mathbb{P} = \{p_1, p_2\} \) and we apply the universal Born rule to it:

\[
\|Z\| = \det \left( \exp (X[p_1]) + \exp (X[p_2]) \right)
\]

(165)

We note three observations:

1. In the case of a paravector, we define the norm as follows:

\[
\|v\| = \sqrt{(r + x)(r - x)}
\]

(166)

\[
= \sqrt{r^2 + x^2}
\]

(167)

2. We further note that the exponential of a vector is

\[
\exp x = \cosh \|x\| + \frac{x}{\|x\|} \sinh \|x\|
\]

(168)

3. We note that the Clifford conjugate of \( \exp x \) is:

\[
(\exp x)^\square = \cosh \|x\| - \frac{x}{\|x\|} \sinh \|x\|
\]

(169)

4. Finally, we adopt the notation \( X[p_i] = x_i \) and \( r[p_i] = r_i \).

With these observations, we can now find the expression of \( \|Z\| \):

\[
\exp M[p] = e^{r[p]} \exp (X[p_0] + X_1[p]\hat{x}_1 + X_2[p]\hat{x}_2 + X_3[p]\hat{x}_3)
\]

(163)

\[
= e^{r[p]} \exp (X[p])
\]

(164)
\[ \sqrt{Z} = \left( e^t \cosh \| x_1 \| + e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| + e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ \left( e^t \cosh \| x_1 \| - e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| - e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ = e^t \cosh \| x_1 \| \left( e^t \cosh \| x_1 \| - e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| - e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ + e^t \cosh \| x_2 \| \left( e^t \cosh \| x_1 \| - e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| - e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ + e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| \left( e^t \cosh \| x_1 \| - e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| - e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ + e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \left( e^t \cosh \| x_1 \| - e^t \frac{x_1}{\| x_1 \|} \sinh \| x_1 \| + e^t \cosh \| x_2 \| - e^t \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \right) \]
\[ = e^{2t} \cosh^2 \| x_1 \| - e^{2t} \frac{x_1}{\| x_1 \|} \cosh \| x_1 \| \sinh \| x_1 \| + e^{2t} \cosh \| x_2 \| \cosh \| x_2 \| + e^{2t} \frac{x_2}{\| x_2 \|} \cosh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \cosh \| x_2 \| \sinh \| x_2 \| \]
\[ + e^{2t} \cosh \| x_2 \| \sinh \| x_2 \| + e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \sinh \| x_2 \| \]
\[ + e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \cosh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \sinh \| x_2 \| \]
\[ = e^{2t} \cosh^2 \| x_1 \| - e^{2t} \frac{x_1}{\| x_1 \|} \cosh \| x_1 \| \sinh \| x_1 \| + e^{2t} \cosh \| x_2 \| \cosh \| x_2 \| + e^{2t} \frac{x_2}{\| x_2 \|} \cosh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \cosh \| x_2 \| \sinh \| x_2 \| \]
\[ + e^{2t} \cosh \| x_2 \| \sinh \| x_2 \| + e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \sinh \| x_2 \| \]
\[ + e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \cosh \| x_2 \| - e^{2t} \frac{x_2}{\| x_2 \|} \sinh \| x_2 \| \sinh \| x_2 \| \]
\[ (170) \]
\[ (171) \]

To simplify this expression, we note the following observations:

1. First, we note the following well-known identity:

\[ \cosh^2 x - \sinh^2 x = 1 \]  
\[ (173) \]

2. we also note that:

\[ \frac{x^2}{\| x \|^2} = 1 \]  
\[ (174) \]

Proof. Let \( x = x_0 \hat{x}_0 + x_1 \hat{x}_1 + x_2 \hat{x}_2 + x_3 \hat{x}_3 \). Then the geometric product is \( x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2 \) which is the same as the square of the norm. \( \Box \)
We proceed with our simplifications as follows:

\[ = e^{2r_1} + e^{2r_2} - e^{2r_1} \frac{x_1}{||x_1||} \cosh ||x_1|| \sinh ||x_1|| + e^{r_1} e^{r_2} \cosh ||x_1|| \cosh ||x_2|| - e^{r_1} e^{r_2} \frac{x_2}{||x_2||} \cosh ||x_1|| \sinh ||x_2|| + e^{r_2} e^{r_2} \cosh ||x_2|| \cosh ||x_1|| - e^{r_1} e^{r_2} \frac{x_1}{||x_1||} \cosh ||x_1|| \sinh ||x_2|| - e^{r_2} e^{r_2} \frac{x_2}{||x_2||} \cosh ||x_1|| \sinh ||x_2|| + e^{2r_1} \frac{x_1}{||x_1||} \sinh ||x_1|| \cosh ||x_1||
\]

\[ + e^{r_1} e^{r_2} \frac{x_1}{||x_1||} \sinh ||x_1|| \cosh ||x_2|| - e^{r_1} e^{r_2} \frac{x_1 x_2}{||x_1|| ||x_2||} \sinh ||x_1|| \sinh ||x_2|| + e^{r_2} \frac{x_2}{||x_2||} \sinh ||x_2|| \cosh ||x_2|| \]

\[ = e^{2r_1} + e^{2r_2} + 2e^{r_1} e^{r_2} \cosh ||x_1|| \cosh ||x_2|| - 2e^{r_1} e^{r_2} \frac{x_1 x_2}{||x_1|| ||x_2||} \sinh ||x_1|| \sinh ||x_2|| \]

For even greater simplicity, we can look at the colinear case where

\[ \frac{x_1 x_2}{||x_1|| ||x_2||} = 1: \]

\[ \sqrt{||z||} = e^{2r_1} + e^{2r_2} + 2e^{r_1} e^{r_2} \cosh ||x_1|| \cosh ||x_2|| - 2e^{r_1} e^{r_2} \sinh ||x_1|| \sinh ||x_2|| \]

\[ = e^{2r_1} + e^{2r_2} + 2e^{r_1} e^{r_2} (||x_1|| - ||x_2||) \]  \hspace{2cm} (177)

The quantities \(|x_1|| and ||x_2|| correspond to the Lorentz norm. Since this norm is able to produce an imaginary number for space-like separation, the probability rule is able to produce three different regimes of interference:

1. (Quantum): In the case where \(|x_1| - ||x_2|| is imaginary (space-like separation), the hyperbolic cosine is converted to a trigonometric cosine, and the probability rules are quantum mechanical.

2. (Classical): In the case where \(|x_1| - ||x_2|| = 0 (light-like separation or equi-distance), then the probability rules are classical:

\[ |z_1|^2 + |z_2|^2 + 2|z_1||z_2| \cosh 0 \implies |z_1| + |z_2| \]  \hspace{2cm} (179)

3. (Hyperbolic): In the case where \(|x_1| - ||x_2|| \in \mathbb{R} (time-like separation), the probability rules are hyperbolic:

\[ |z_1|^2 + |z_2|^2 + 2|z_1| |z_2| \cosh (||x_1|| - ||x_2||) \]  \hspace{2cm} (180)

To combine both the even part and the odd part together, let us introduce the universal norm:
6.5 The Universal Norm

**Definition 13** (Universal Norm). We take the norm of the geometric algebra $G_4(\mathbb{R})$ to be a function $\|\cdot\| : G_4(\mathbb{R}) \rightarrow \mathbb{R}$ with the requirement that it’s output be the same as that of the determinant of its matrix representation (If we include the complex $G_4(\mathbb{C})$, the norm remains the same but its domain is now $\|\cdot\| : G_4(\mathbb{C}) \rightarrow \mathbb{C}$. It attributes no new geometry to the complexification of the pre-factors and they simply “pass-through” the norm.):

$$\|u\| := \sqrt[4]{\det M[u]} \quad (181)$$

We can equivalently define this norm\(^{45}\) fully in the language of geometric algebra, as follows:

$$\|u\| = \sqrt{[\left(u^{\ominus}\right)u]_{\{3,4\}} [\left(u^{\ominus}\right)u]} \quad (182)$$

Let us explain the notation. First, $u^{\ominus}$ is the geometric conjugate (also called the Clifford conjugate) of a multivector defined, in $G_4$, as follows:

$$u^{\ominus} = \langle u \rangle_0 - \langle u \rangle_1 - \langle u \rangle_2 + \langle u \rangle_3 + \langle u \rangle_4 \quad (183)$$

Furthermore, the notation $[\left(u^{\ominus}\right)u]_{\{3,4\}}$ represents the $\{3,4\}$-grade conjugate. It is defined as follows:

$$[u]_{\{3,4\}} = \langle u \rangle_0 + \langle u \rangle_1 + \langle u \rangle_2 - \langle u \rangle_3 - \langle u \rangle_4 \quad (184)$$

For reference, the grades of $u$ are:

$$\langle u \rangle_0 = r \quad (185)$$

$$\langle u \rangle_1 = (X_0)x_0 + (X_1)x_1 + (X_2)x_2 + (X_3)x_3 \quad (186)$$

$$\langle u \rangle_2 = (E_1)x_0 \wedge x_1 + (E_2)x_0 \wedge x_2 + (E_3)x_0 \wedge x_3 + (B_1)x_2 \wedge x_3 + (B_2)x_3 \wedge x_1 + (B_3)x_1 \wedge x_2) \quad (187)$$

$$\langle u \rangle_3 = (V_0)x_1 \wedge x_2 \wedge x_3 + (V_1)x_0 \wedge x_2 \wedge x_3 + (V_2)x_0 \wedge x_1 \wedge x_3 + (V_3)x_0 \wedge x_1 \wedge x_2) \quad (188)$$

$$\langle u \rangle_4 = (\theta)x_0 \wedge x_1 \wedge x_2 \wedge x_3 \quad (189)$$

Now, let us write out the "half-product" of the norm: $(u^{\ominus})u \in G_0 \oplus G_3 \oplus G_4$.
\[(u \Box) u =
\]
\[
r^2 - \theta^2 + B_1^2 + B_2^2 + B_3^2 - E_1^2 - E_2^2 - E_3^2 + V_0^2 - V_1^2 - V_2^2 - V_3^2 - X_0^2 + X_1^2 + X_2^2 + X_3^2
\]
\[
+ \hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 (2 E_3 V_0 + 2 B_2 V_1 - 2 B_1 V_2 + 2r V_3 - 2 B_3 X_0 + 2 E_2 X_1 - 2 E_1 X_2 - 2 \theta X_3)
\]
\[
+ \hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_3 (-2 E_2 V_0 + 2 B_3 V_1 - 2r V_2 - 2 B_1 V_3 + 2 B_2 X_0 + 2 E_3 X_1 + 2 \theta X_2 - 2 E_1 X_3)
\]
\[
+ \hat{x}_0 \wedge \hat{x}_2 \wedge \hat{x}_3 (2 E_1 V_0 + 2r V_1 + 2 B_3 V_2 - 2 B_2 V_3 - 2 B_1 X_0 - 2 \theta X_1 + 2 E_3 X_2 - 2 E_2 X_3)
\]
\[
+ \hat{x}_1 \wedge \hat{x}_2 \wedge \hat{x}_3 (2r V_0 + 2 E_1 V_1 + 2 E_2 V_2 + 2 E_3 V_3 - 2 \theta X_0 - 2 B_1 X_1 - 2 B_2 X_2 - 2 B_3 X_3)
\]
\[
+ \hat{x}_0 \wedge \hat{x}_1 \wedge \hat{x}_2 \wedge \hat{x}_3 (2r \theta - 2 B_1 E_1 - 2 B_2 E_2 - 2 B_3 E_3 - 2 V_0 X_0 + 2 V_1 X_1 + 2 V_2 X_2 + 2 V_3 X_3)
\]
\[(190)\]

If we now complete the full product we end up with the following norm applicable to a general multivector of \(G_4(\mathbb{C})\), which we call the universal norm:

\[
\left(\left\langle (u \Box) u \right\rangle_{(3,4)} (u \Box) u \right)^4 = 
\]
\[
(r^2 - \theta^2 + B_1^2 + B_2^2 + B_3^2 - E_1^2 - E_2^2 - E_3^2 + V_0^2 - V_1^2 - V_2^2 - V_3^2 - X_0^2 + X_1^2 + X_2^2 + X_3^2)^2
\]
\[
+ (2 E_3 V_0 + 2 B_2 V_1 - 2 B_1 V_2 + 2r V_3 - 2 B_3 X_0 + 2 E_2 X_1 - 2 E_1 X_2 - 2 \theta X_3)^2
\]
\[
+ (-2 E_2 V_0 + 2 B_3 V_1 - 2r V_2 - 2 B_1 V_3 + 2 B_2 X_0 + 2 E_3 X_1 + 2 \theta X_2 - 2 E_1 X_3)^2
\]
\[
+ (2 E_1 V_0 + 2r V_1 + 2 B_3 V_2 - 2 B_2 V_3 - 2 B_1 X_0 - 2 \theta X_1 + 2 E_3 X_2 - 2 E_2 X_3)^2
\]
\[
- (2r V_0 + 2 E_1 V_1 + 2 E_2 V_2 + 2 E_3 V_3 - 2 \theta X_0 - 2 B_1 X_1 - 2 B_2 X_2 - 2 B_3 X_3)^2
\]
\[
+ 4 (r \theta - B_1 E_1 - B_2 E_2 - B_3 E_3 - V_0 X_0 + V_1 X_1 + V_2 X_2 + V_3 X_3)^2
\]
\[(191)\]

Let us now take a few examples. Starting with a scalar:

**Example 1** (Universal norm applied to a real).

\[
v := r \Rightarrow \|v\| = \sqrt{(u \Box)u} = r
\]
\[(192)\]
\[(193)\]

**Example 2** (Universal norm applied to a complex). Let

\[
v := a + b \mathbf{i}
\]
\[
\Rightarrow \|v\| = \sqrt{a^2 + b^2}
\]
\[(194)\]
\[(195)\]

**Example 3** (Universal norm applied to a 1-vector (Euclid)).

\[
v := X_1 \hat{x}_1 + X_2 \hat{x}_2 + X_3 \hat{x}_3
\]
\[
\Rightarrow \|v\| = \sqrt{X_1^2 + X_2^2 + X_3^2}
\]
\[(196)\]
\[(197)\]

**Example 4** (Universal norm applied to a 1-vector (Lorentz)).

\[
v := X_0 \hat{x}_0 + X_1 \hat{x}_1 + X_2 \hat{x}_2 + X_3 \hat{x}_3
\]
\[
\Rightarrow \|v\| = \sqrt{-X_0^2 + X_1^2 + X_2^2 + X_3^2}
\]
\[(198)\]
\[(199)\]
Example 5 (Universal norm applied to a 2-vector (Faraday tensor)).

\[ \mathbf{v} := E_1 \mathbf{x}_0 \wedge \mathbf{x}_1 + E_2 \mathbf{x}_1 \wedge \mathbf{x}_2 + E_3 \mathbf{x}_0 \wedge \mathbf{x}_3 + B_1 \mathbf{x}_2 \wedge \mathbf{x}_3 + B_2 \mathbf{x}_3 \wedge \mathbf{x}_1 + B_3 \mathbf{x}_1 \wedge \mathbf{x}_2 \]

\[
\implies \| \mathbf{v} \| = \sqrt[4]{(E_1^2 + E_2^2 + E_3^2 - B_1^2 - B_2^2 - B_3^2)^2 + 4(E_1 B_1 + E_2 B_2 + E_3 B_3)^2}
\]

We note that the quantities \( E_1^2 + E_2^2 + E_3^2 - B_1^2 - B_2^2 - B_3^2 \) and \( E_1 B_1 + E_2 B_2 + E_3 B_3 \) are the two Lorentz invariant of the Faraday tensor.

The universal norm is the foundation of the most general interference pattern that can be produced by the extended Born rule.

6.6 Quartic Hilbert spaces

First, we limit the domain of the universal norm to a) the reals and b) to the exponential map of multivectors:

\[ \| \cdot \| : \exp(\mathcal{G}_4(\mathbb{R})) \rightarrow \mathbb{R} \]

Limiting the norm only to exponentials of multi-vectors renders it positive-definite. We then use it to define a 'quartic' Hilbert space which maps the set of transformations to a normalizable probability distribution, via the extended Born rule. Specifically, the normalization conditions over a domain of integration \([D]\), and for a wavefunction \( \psi(x) \) is:

\[
\int_{[D]} |\psi^\square \psi| \, 3,4 \psi^\square \psi \, dx < \infty
\]

This definition is the geometric-algebra equivalent of the matrix representation:

\[
\int_{[D]} \det \exp M[u(x)] \, dx < \infty
\]

6.7 Octic Hilbert spaces

If we now want a probability norm with map:

\[ \| \cdot \| : \exp(\mathcal{G}_4(\mathbb{C})) \rightarrow \mathbb{R} \]

such that exponential map over the complexification of the space-time algebra, representing the general linear group of transformations,
the GL(4, C) group, is obtained by taking the square modulos of the universal norm:

\[ \| \psi \| = \left( |\psi^\square \psi|_{3,4} \psi^\square \psi \right) \left( |\psi^\square \psi|_{3,4} \psi^\square \psi \right)^* \] (206)

To maintain the group isomorphism of the geometric algebra to its matrix representation, the octic Hilbert space can be embedded into the ‘nearest-higher-dimensional’ algebra that shares said group isomorphism, which is \( G_6(C) \cong M(8, C) \). The physical interpretation and consequences implied by the octic Hilbert space and its embedding into a six-dimensional normalization space will be explored in a future paper.

6.8 Quantum computation

Finally, we note that experiments, defined as verified programs, are realized within the constrain of nature, producing a unitary evolving wavefunction capable of general quantum computation, such that the reference manifest remains verified throughout the evolution of the system.

7 Discussion

7.1 What is space-time?

Space-time is the interference pattern interpolated from \( N \) measurement events whose probability is given by the extended Born rule, and specifically from the inner-product/metric/universal-norm defining the interference pattern between two events. It occurs for the same reason that a complex interference pattern emerges out of a double-split experiment from the accumulation of detection events. Beyond four dimensions, the eigenvalues of the matrix representation of the general multi-vector have no solutions expressible in the radicals, and thus the wavefunction looses the structure of its informational backbone (as a partition function of universal statistical physics). This results in a no-go theorem for a wavefunction interpreted as a ‘wave of probability’ beyond four dimensions. Of course, we note that if we allow eigenvalues not expressible in radicals, the no-go theorem fails.

7.2 The measurement problem (recap)

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:

$$|\psi\rangle|\phi\rangle \rightarrow \sum_{n} c_{n} |\chi_{n}\rangle|\phi_{n}\rangle$$ (207)

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:

$$|\psi\rangle|\phi\rangle|e\rangle \rightarrow \sum_{n} c_{n} |\chi_{n}\rangle|\phi_{n}\rangle|e_{n}\rangle$$ (208)

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.

7.3 The measurement problem (solution)

Rather than taking some arbitrary set of laws as postulates, our methodology addresses the problem from the other direction by taking as its sole axiom the existence of the state of affairs referenced by \( \mathcal{M} \). The problem is not "how does a measurement causes one solution out of many to become actual", but rather "how does the existence of a single a-priori solution implies the laws of physics as a theorem? The second question is the correct one to ask.

The answer is that to define a probability measure such that the reference manifest is informative, one must extend the domain of reality (given as the reference manifest) to that of the domain of science. This extension is a mathematical construction, compatible with, but nonetheless unsupported by reality solely; the larger domain is constructed to make knowledge of reality informative. Nonetheless, the laws of physics do require this process lest they cannot be derived. This is why the laws of physics are a theorem of science applied to reality (not of reality alone). It is consequently inexact to claim that the laws of physics are the laws of reality: precisely, they are the laws that govern the random selection of reality (expressed as a reference manifest) from amongst the set of all possible realities (all possible manifests). To sum up, the measurement problem is the artefact caused by extending the domain of reality to that of science for the purposes of constructing a probability measure such that "I"/the-observer’s knowledge of reality is maximally informative.

An interpretation of quantum mechanics is available within this setup and follows from statistical physics by referencing the concepts of microstates and macrostates. 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 \( \mathcal{M} \), 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; but each from a difference perspective.

The observer, in our setup, corresponds to the microstate. This is an inversion on the typical association of the observer in ordinary statistical physics, where it is commonly associated to a device measuring quantities such as the volume or the energy, etc. In universal thermodynamics, due to this inversion, the association of the observer to the microscopic description represents $O$'s knowledge of the measurements in the form of a reference manifest. The macroscopic description is a logico-deductive model formulated by the observer under the principle of maximum entropy to assign the appropriate probability measure to the reference manifest such that $O$'s knowledge of it is maximally informative. The laws of physics, being the product of universal statistical physics, cannot tell us the result of a measurement, for the same reason that $pV = nRT$ cannot tell us the position of the molecules in the box.

The quantum measurement problem appears as unexplained only when one postulates the laws of physics rather than deriving them. When one postulates the laws, then solves them for solutions, one obtains a plurality of possible solutions, only one of which is actual. Since one intuitively expected that the laws of physics ought to produce the actual solution, one may then become baffled as to why the laws of physics produce a plurality of solutions.

Maximizing the entropy associated to the selection of the reference manifest from the domain of science releases the constraints imposed by a sole reference manifest, 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. Intuitively, if the laws of physics were restricted to have reality as their domain, then the laws of physics are a Turing machine whose domain is the reference manifest and nothing more. If instead, the domain of the laws of physics is the domain of science but conditional upon one manifest to be actual, then all of a sudden the laws of physics become a universal pattern that survive any transformations of the reference manifest. However, one cannot form a pattern from a single existing candidate (the reference manifest), unless one invents hypothetical alternatives (in this case, 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 one maximizes the entropy to produce 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 the information which was erased by maximizing the entropy. Consider if one would have instead said: "I am a physicist, but I could have been batman". How credible is that claim? Supposedly, we may admit that being batman violates the laws of physics (I am told it is actually superman that violates the laws of physics, my apologies), whereas being a doctor doesn't. Do we then want our laws of physics to rule out batman, 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, whilst ruling out only what would be considered 'truly' impossible.

In universal statistical physics there is no collapse (thus the Copenhagen interpretation is rejected), and also the reality is never in a superposition of many-worlds (thus the many-world interpretation is also rejected). Via the syllogisms, universal statistical physics limits the quantity of available mathematical work to allow for precisely the verification of a single reference manifest (defined as the set of all verified experiments). It then follows that all alternative manifests are mathematical creations used to facilitate the formulation of the laws of physics as a probability measure, and thus, have no ontological properties. Hypothetical manifest are comprised of experiments that are in principle verifiable but as of yet are unverified. Universal statistical physics 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. Universal statistical physics consequently identifies the measurement problem as an artefact of attributing information to a reference solution, which is a completely 'non-mystical' explanation.

References


A Sketch/Figure
Cartesian Paradigm

My Proposal

Physics is obtained, first, by creating a purely mathematical model of science, then by solving it.

Structure of Reality (inherited from reality, at the end of the road)

Universal Equation of State

Observable Universe

Physics (mathematical description thereof)

Statistical Ensemble of Experiments

Set of Realized Experiments

Observed Universe

Observer (mathematical description thereof)

The difference in entropy is the mathematical origin of the quantum measurement

I doubt → I think → I am

Pitfall @ Mind-body Problem

Figure 4: An indubitable description of reality sufficient to recover the laws of physics.