An attempt to prove physics by making reality indubitable

Alexandre Harvey-Tremblay [1,2]

[1] Independent scientist
[2] ahtremblay@gmail.com

September 19, 2020

In this manuscript I attempt to produce the most fundamental description of physics and reality I believe to be possible. Specifically, using algorithmic information theory, statistical physics and entropy as my tools, my strategy is to define reality as the ensemble of all realized experiments (the ensemble of what "I" has indubitably proven), the domain of science as the set of all realizable experiments (the set of what "I" could prove) and finally nature as the group of all possible transformations of the ensemble. Then, at its most fundamental level, *physics* is —quite simply— the probability measure that makes reality maximally informative (to "I"/ the observer) within the domain of science and under the constraint of nature. The procedure yields novel physics in the form of the mathematical origin of (an extended version of) the Born rule as the probability measure connecting the domain of science to reality, while the geometry of space-time itself is automatically emergent in the structure of said extended Born rule; a process which, notably, is self-limited to precisely four space-time dimensions. Thus producing, automatically, uniquely and necessarily, the equations for the familiar four-dimensional geometric quantum mechanical universe as the inescapable result of applying a formalization of the practice of science to an indubitable description of reality.

Contents

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 $f : X \to \mathbb{R}$ will be written as $f[x]$ for $x \in X$. S will denote the entropy, $\mathcal{S}$ the action, L the Lagrangian, and $\mathcal{L}$ the Lagrangian density. Sets, unless a prior convention assigns it another symbol, will be written using the blackboard bold typography (ex: $\mathbb{L}, \mathbb{W}, \mathbb{Q}$, etc.). Matrices will be in bold upper case (ex: $\mathbf{A}, \mathbf{B}$), whereas vectors and multivectors will be in bold lower case (ex: $\mathbf{u}, \mathbf{v}, \mathbf{g}$) 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 $\mathbf{I}$ 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 $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ (such that $\mathbf{e}_\nu \cdot \mathbf{e}_\mu = g_{\mu\nu}$) and if they are orthonormal as $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n$ (such that $\hat{\mathbf{x}}_\mu \cdot \hat{\mathbf{x}}_\nu = \eta_{\mu\nu}$). The asterisk z^* denotes the complex conjugate of z, and the dagger $\mathbf{A}^\dagger$ denotes the conjugate transpose of $\mathbf{A}$. A geometric algebra of n dimensions over a field $\mathbb{F}$ is noted as $\mathcal{G}_n(\mathbb{F})$. We note the matrix representation of a multivector $\mathbf{g}$ as $M[\mathbf{g}]$, defined as a map $M : \mathcal{G}_n(\mathbb{F}) \to \mathbb{M}(n, \mathbb{F})$

which equates the geometric product to the matrix product, and thus benefits from group isomophism. The grades of a multivector will be denoted as $\langle \mathbf{v} \rangle_k$. Specifically, $\langle \mathbf{v} \rangle_0$ is a scalar, $\langle \mathbf{v} \rangle_1$ is a vector, $\langle \mathbf{v} \rangle_2$ is a bivector, $\langle \mathbf{v} \rangle_{n-1}$ is a pseudovector and $\langle \mathbf{v} \rangle_n$ is a pseudoscalar. Furthermore, a scalar and a vector $\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_1$ is a paravector, and a combination of even grades $(\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_2 + \langle \mathbf{v} \rangle_4 + \dots)$ or odd grades $(\langle \mathbf{v} \rangle_1 + \langle \mathbf{v} \rangle_3 + \dots)$ are even-multivectors or odd-multivectors, respectively. The commutator is defined as $[\mathbf{A}, \mathbf{B}] := \mathbf{AB} - \mathbf{BA}$ and the anti-commutator as $\{\mathbf{A}, \mathbf{B}\} := \mathbf{AB} + \mathbf{BA}$. We use the symbol $\cong$ to relate two sets that are related by a group isomorphism (ex: $\mathcal{G}_4(\mathbb{C}) \cong \mathbb{M}(4, \mathbb{C})$).

1 *Anti-pattern: postulating a path to reality*

Nearly all currently available physical theories are formulated as a list of axioms such that each corresponds to a law, and said laws are justified based on the analysis of empirical data. The archetypal example is special relativity which postulates an invariant interval between any two space-time events, justified primarily under the repeated failure of the Michelson-Morley experiments. Another example would be electricity and magnetism that was empirically identified decades, if not centuries, before they were integrated in a unified basis by Maxwell, such that four (relatively simple) laws are now the axioms. Finally, a third example would be Dirac–von Neumann axioms that defines quantum mechanics in terms of operators acting on a Hilbert space, justified under at least two decades of prior accumulated empirical data.

These physical laws are made formal within a choice of mathematical framework: popular ones include manifolds, linear algebra or Hilbert spaces. Since these mathematical frameworks are formulated as structures layered on top of structures (for instance manifolds are layered on topology which is layered on set theory) and that the laws of physics are at the very end, one seems invited to ask: can we eventually get reality by postulating enough stacks? As Exhibit A, consider a recent book, titled "The Road to Reality: A complete guide to the laws of the universe" by Roger Penrose[3], 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

[3] Roger Penrose. *The road to reality: A complete guide to the laws of the universe*. Random house, 2006

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 and most complete "tower of postulates" produced thus far, and then, as expected, concludes that he does not see a road to reality in any of the proposals. So... why does tower-building not achieve the desired goal of bringing us closer to reality? What's truly missing?

I've nearly always harbored the intuition that postulating an equation (or generally a law, or set of laws) to be true (in the axiomatic sense) is a fatal mistake that necessarily disconnects the formalism from reality, because —in reality— it is the laws that are derived and it is reality that is axiomatic. That is, postulating an empirically-derived law is not only an oxymoron (i.e. a postulated derived law), but more than that, it is the fundamental anti-pattern that erases reality from the equations. If you will allow, I will give my best attempt to explain my reasoning behind why I think that to be the case.

> "An anti-pattern is a common response to a recurring problem that is usually ineffective and risks being highly counterproductive."[4]

[4] D Budgen. Software design . harlow, uk, 2003; and Scott W Ambler. *Process patterns: building large-scale systems using object technology*. Cambridge university press, 1998

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 text — I just had the intense intuition that she misunderstood something fundamental about reality and therefore her statements *had* to be verified before they could be trusted. So during the lunch break (for about 1.5 hours) I set out to do just that. I picked up rocks from the schoolyard and added them all out, permuting over the different arrangements I could construct and by so doing, verified a (tiny) subset of arithmetic. Okay, so I have established that it works with rocks, but does it work with... branches? So I went to get branches, and verified it again, and sure it worked for branches too. One of the other kids asked me what I was doing and I told him that I was trying to verify that what the teacher had said was true. He asked, surprised, "oh. You don't believe her?". I responded along the lines of: "I am almost certain that she is right, but I cannot take the risk to take it on faith". Eventually, the bell rang and I ran out of time. Back in class began a long process of ruminating over what had transpired. Before I could continue with the program, I had to somehow grind away at the claim that all arithmetical permutations of numbers holds in reality. Clearly, it works with small numbers; I in fact just recently verified it in the schoolyard. For numbers larger than what I could personally verify, I convinced myself of the somewhat reasonable argument that possibly millions of other human beings where taught these equations before me and themselves have surely verified very exhaustively the claims. However, I reckoned that there was still a limit (a very large one indeed) beyond which 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 stringly 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. Then, when I was presented with string theory (and its competitors) as the next and possibly final step, I called quit. It would take me years to merely acquire the technical language sufficient to describe the problem, then to pinpoint exactly what causes it, and finally to produce a proposal able 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, and if it doesn't, then they reckon that it simply means that the tower must be patched and improved until it does. This reversed 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:

$$\sum_{\mathbf{F}\in\mathbb{F}} \mathbf{F} = m\mathbf{a} \tag{1}$$

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 to we pay when we erase that 'something' by writing "$F = ma$" as an axiom instead of as a theorem? Specifically, we create the logical equivalent of 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 the relationship 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[5] 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. This preserves the universality of the problem yet makes it tractable.

[5] Seth Lloyd. Computational capacity of the universe. *Physical Review Letters*, 88(23):237901, 2002

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.

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 connecting the second model (science) to the first (reality) via entropy produces a third model (physics). The construction is thus able to provide a mathematical account of the origin of the laws of physics, without having to postulate said laws as axioms.

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 theoretical physics postulates laws, then derives states. As such, they are logically entailed in opposite directions.

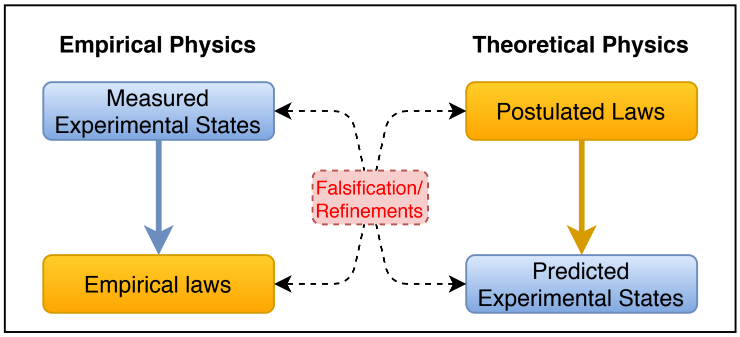

Figure 1: The current practice of physics, summarized.

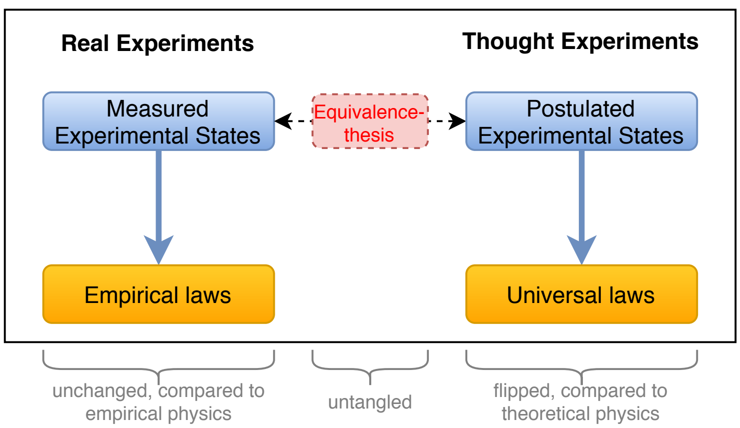

Figure 2: In my proposal, the falsification/refinement "algorithm" of science is untangled to a mere equivalence-thesis by reversing the typical relationship between states and laws used in theoretical physics.

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 computational universality which applies to both sets — remarkably, if reality is the purview of science, then 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 without having to leave the realm of mathematics, yielding physics as a mathematical theorem.

2.1 *Math is work*

The possibility of creating a purely mathematical structure that nonetheless has properties normally associated only to physical objects runs counter to expectations and therefore the philosophically safe bet has usually been to assume it to be an impossibility. However, contrary to expectations, a number of years ago I was able to lay out a precise path able to do so. In this section, I will provide a simplified example of the technique that I use, which illustrates the key concepts.

I will now 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:

$$\text{hash}[1 + 1 = 1] = \text{fa1869db4bfbf1767a5446b6a9290243} \tag{2}$$

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:

$$\begin{aligned} \text{hash} \quad : \quad \mathbb{L}_{\text{PA}} &\longrightarrow \mathbb{L}_{\text{hex}} \\ \textit{statement} &\longmapsto \textit{hash} \end{aligned} \tag{3}$$

[6] A secure hash function can only be inverted by brute force.

[7] A perfect hash function is an injective function that maps each input to an hash, with no collisions.

I also define the inverse function:

$$\begin{aligned} \text{bruteforce} \quad : \quad \mathbb{L}_{\text{hex}} &\longrightarrow \mathbb{L}_{\text{PA}} \\ \textit{hash} &\longmapsto \textit{statement} \end{aligned} \tag{4}$$

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)**.**

$$\text{rot} := \text{bruteforce}[\text{fa1869db4bfbf1767a5446b6a9290243}] \tag{5}$$

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)**.**

$$\text{rot} \cup \text{PA} \tag{6}$$

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)**.**

$$\text{Axiom-X} := \text{bruteforce}[\text{0cfae383362bc63d7ac429a5755fef05}] \tag{7}$$

Now I ask you, knowing the hash but not the statement, is the formal theory comprised of PA ∪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.

[8] Seth Lloyd. Computational capacity of the universe. *Physical Review Letters*, 88(23):237901, 2002

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, accidentally or on purpose, at any level of computational complexity within a theory. My specific example with a bruteforce 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.

[9] Alan M Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1):230–265, 1937; and Gregory J Chaitin. Meta math! the quest for omega. *arXiv preprint math/0404335*, 2004

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 recent non-peer verified results, 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.

Intuitively the answer is obvious (finding mistakes takes effort), but 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[10], but also on the more recent works of Baez and Stay regarding algorithmic thermodynamics[11] 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.

[10] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975

[11] John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012

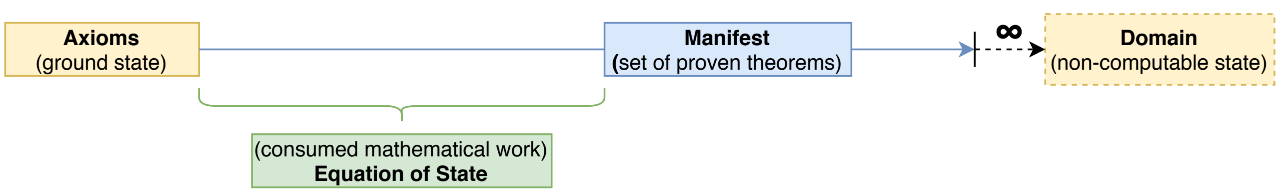

Figure 3: The equation of state quantifies the amount of mathematical work required to excite the system to a given manifest.

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

$$\text{Axioms} \subseteq \text{Manifest} \subset \text{Dom}[\text{Axioms}] \tag{8}$$

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.

2.2 *The fundamental structure of reality*

Why is this approach able to transpose a structure 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 addressed the mind-body problem? In the *Treatise of man* (written before 1637, but published posthumously) and later in the *The passions of the soul* (1649), Descartes presents his picture of man as composed of both a body and a soul. The body is a 'machine made of earth' and the soul is the center of thoughts. Communication between the two parts would be handled by the 'infamous' pineal gland, whose inner working he covers in great detail in (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 bioelectromagnetics some $\approx$150 years later. Of course, the Cartesian pineal gland theory was incorrect, but the problem on how to connect the mind and the body remained.

[12] Gert-Jan Lokhorst. Descartes and the pineal gland. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018

The proof of the 'cogito ergo sum' by Descartes 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. This is a pitfall that one must side-step.

Let me clarify that I consider the term 'body' to be a euphemism for the material universe, and not to be the literal human body as Descartes used the term. I also consider the term 'mind' to be a euphemism for the domain of mathematics. The problem then translates to a modern form: how does one obtain a structure equivalent to the material universe from a purely mathematical starting point? I am claiming that an ensemble of indubitable statements, along with the mathematical work it is contingent upon, is a concept than once paired with entropy can solve the problem up to equivalence. Specifically, the framework attributes to the material universe a description as, quite remarkably and perhaps even surprisingly, a (special kind of) probability measure that is exactly four-dimensional, geometric and quantum mechanical. All of these properties elegantly and automatically come out of the framework without having to inject any prior physical or logical baggage. We note that such a probability measure is a purely mathematical object, but nonetheless, as per conventional quantum mechanical wisdom, it constitutes a complete description of the system — thereby, recovering the material universe, as we said, 'up to equivalence'. Interpreted consistently with this derivation, it is then suggestive that what an observer reports as the material universe can be construed as a simple side-effect of understanding the verification of indubitable facts via this probability measure —which is its maximally informative interpretation—, and the reported perceived inviolability of the laws of physics is also a side-effect of, in this case, mathematical work, the structure which obeys these laws, serving as the constraints of the probability measure. The proof along with the accompanying argument is somewhat substantial but each step is remarkably intuitive, and primarily involves formalising the concepts normally associated with the practice of science. The method thus produces the laws of physics as a result of formally applying the practice of science to an indubitable description of reality, which I would argue is precisely where one could excepts to find the laws of physics within mathematical space in the first place.

The first step is to produce an indubitable description of reality. Specifically, this will take the form of a universal/Turing-complete indubitable framework. Descartes helps us here — he essentially constructed what amounts to the first universal proof checker, and applied it to the set of all mathematical statements. However, since it was done in an informal manner, he missed out on this massive opportunity.

Let us start with the definition of language:

Definition 1 (Language)**.** *A language $\mathbb{L}$, with alphabet Σ, is the set of all sentences $(s_1, s_2, \dots)$ that can be constructed from the elements of Σ and it includes the empty sentence $\varnothing$:*

$$\mathbb{L} := \{\varnothing, s1, s2, \dots\} \tag{9}$$

For instance, the sentences of the binary language are:

$$\mathbb{L}_b := \{\varnothing, 0, 1, 00, 01, 10, 11, 000, \dots\} \tag{10}$$

and its alphabet is:

$$\Sigma_b = \{0, 1\} \tag{11}$$

Definition 2 (Shortlex ordering)**.** *A shortlex ordering is a list of the sentences of* $\mathbb{L}$*, first ordered by length from shortest to longest, then alphabetically.*

For instance, the shortlex ordering of $\mathbb{L}_b$ is:

$$(\varnothing, 0, 1, 00, 01, 10, 11, 000, \dots) \tag{12}$$

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

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

$$\begin{aligned} \text{MPC} \quad : \qquad \mathbb{L} &\longrightarrow \{1, \nexists\} \\ \textit{sentence} &\longmapsto \textit{result} \end{aligned} \tag{13}$$

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 MPC[*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* $\mathbb{L}$ *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 lets see. To prove $(1 + 1 = 2)$, one requires PA (or an equivalent). However, since the machine only contains logical axioms, it will never halt because no proof will ever be found. On the surface, it may seem that the conclusion of Descartes regarding the idea that one can doubt of all mathematical theorems because the axioms they rely upon need not be true, is sound.

However, once in a while something quite interesting happens. Let's say we feed all sentences of $\mathbb{L}$ in shortlex to MPC using dovetailing scheduling. Eventually this statement will be feed to MPC:

$$\text{PA} \vdash (1 + 1 = 2) \tag{14}$$

The statement states: PA proves that one plus one equals two. As we did with the previous example, we also ask here will MPC[PA $\vdash$ $(1 + 1 = 2)$] 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 soundness, and to further do so without constructing a preliminary "tower of postulates". 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. The set of all indubitable statements is a universal mathematical rock bottom.

Specifically, the significance of this set is as follows:

> The set of all indubitable statements is the only mathematical construction which is both necessarily sound over reality, and universal in the computational sense.

First, we re-iterate that soundness over reality is guaranteed by the fact that each statement is indubitable. Second, we state, via the Church-Turing thesis, that Turing completeness is the highest level of completeness available to mathematics. This description thus gives us the highest mathematical level of expressiveness available to describe reality in a necessarily sound manner.

Let us define a set $\mathbb{M}$, which we call a manifest, of n statements verifiable by MPC:

$$\begin{aligned} \mathbb{M} := \{ & \\ & A_1 \vdash T_1, \\ & A_2 \vdash T_2, \\ & A_3 \vdash T_3, \\ & \vdots \\ & A_n \vdash T_n, \\ \} & \end{aligned} \tag{15}$$

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 — structures that behave as physical objects will be 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*

The previous 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, for a number of reasons. 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:

$$\mathbb{M} := \{p_1, p_2, p_3, \ldots, p_n\} \tag{16}$$

To each such manifest one can then define at least one Turing machine TM that halts for (and only for) each $p \in \mathbb{M}$. In more technical terms, $\mathbb{M}$ 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 $\mathbb{M}$ are verified.

To obtain manifest theories (a mathematically formal definition of a scientific theory) one first selects a universal Turing machine UTM as the baseline (sometimes referred to as the choice of language), then for each program $p \in \mathbb{M}$ a Turing machine, understood as an abstraction layer and referred to in this context as a manifest theory, is also provided as input to the universal Turing machine. Specifically, all programs of $\mathbb{M}$ are then verified as follows:

$$
\begin{aligned}
\{ & \\
& \mathrm{UTM}[\mathrm{TM}_1, p_1] \\
& \mathrm{UTM}[\mathrm{TM}_2, p_2] \\
& \mathrm{UTM}[\mathrm{TM}_3, p_3] \\
& \vdots \\
& \mathrm{UTM}[\mathrm{TM}_n, p_n] \\
\} &
\end{aligned} \tag{17}
$$

Although this is not what we typically think of when we think of a physical theory, a universal Turing machine UTM that recursively enumerated $\mathbb{M}$ 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 $\mathrm{TM}_3 = \mathrm{TM}_4 = \mathrm{TM}_9$, 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 disciples (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 programs of the manifest are within its computational reach.

The set of all manifest 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, manifest 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 manifest 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 manifest 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 computation 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[13] limit, an expression for the minimum amount of energy required to erase one bit of information for a system at thermodynamic equilibrium:

[13] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961; and Rolf Landauer et al. Information is physical. *Physics Today*, 44(5):23–29, 1991

$$E \geq T k_B \ln 2 \tag{18}$$

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 gravity 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 surface-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 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;

[14] John Archibald Wheeler. Frontiers of time. In *Problems in the Foundations of Physics*. North-Holland for the Societa italiana di fisica, 1978; and John A Wheeler. Information, physics, quantum: The search for links. *Complexity, entropy, and the physics of information*, 8, 1990

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

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

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

For John A. Wheeler, it makes little sense to speak of the photon existing (or not existing) until a detector registers a count. But he goes further and suggests that even after the registration of a count, deducing that the photon existed in-between the counts is a "blown-up version of the raw fact, a count". Here, John A. Wheeler implies that the counts are what is real, not the theory that explains the counts. The theory is one hypothesis among many alternatives and is, at best, a mathematical tool to make some sense of the counts, which by themselves define the world irrespectively of the theory.

In "*Frontiers of time*" (about a decade before 'it from bit'), John A. Wheeler lays out multiple attempts to derive some form of physical behavior/law from the study of experimentally-derived bits, but his approaches suffer from introducing physical baggage to get them started. Taking a specific example, on page 150, he reasons that time should emerge out of entropy. So far so good, but then he argues that because the universe goes from Big Bang to Big Stop, to Big Crunch, the statistics of entropy must be time-symmetric. Therefore, he concludes that the acceptable rules of statistics to describe the dynamics of this entropy are those that he calls "double-ended statistics" which works in both directions of time (pages 150-155). The argument has, of course, an obvious fatal error: if time is derived from the bits, then so should the cosmos — why would one not be allowed to refer to time apriori (it must be derived from entropy), but be allowed to refer to the cosmos' hypothetical future time-reversal to justify some properties on the bits? Thirty-nine years later, the results of the Planck Collaboration[15] indicate a critical density consistent with flat topology and eternal expansion, possibly contradicting Wheeler's argument relying upon the necessity of some upcoming future cosmological reversal. Obviously, the eventual correct approach is only appealing if all physical statements (the 'its') follow from the bits such that the future time reversal, if any, ought to be derived from the 'bits'. John A. Wheeler's book presents a myriad of similarly constructed arguments. John A. Wheeler does understand this to be a problem, and in his defense, he does present "double-ended statistics" only as an example of what might be done. Some 11 years later he corrects his approach to the participatory-universe hypothesis.

[15] N Aghanim, Y Akrami, M Ashdown, J Aumont, C Baccigalupi, M Ballardini, AJ Banday, RB Barreiro, N Bartolo, S Basak, et al. Planck 2018 results. vi. cosmological parameters. *arXiv preprint arXiv:1807.06209*, 2018

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[16].

[16] John A Wheeler. Information, physics, quantum: The search for links. *Complexity, entropy, and the physics of information*, 8, 1990; and John Archibald Wheeler. Frontiers of time. In *Problems in the Foundations of Physics*. North-Holland for the Societa italiana di fisica, 1978

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:

$$1110101100010011101010101010 \tag{19}$$

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. Instead of connecting information to some complex pre-existing physical quantity, we here connect it to the general concept of 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 Ω construction, in the book "*Meta Math!*"[18]. Let $\mathbb{U}$ be the set of all universal Turing machines, then:

[17] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975

[18] Gregory J Chaitin. Meta math! the quest for omega. *arXiv preprint math/0404335*, 2004

$$\begin{array}{rrcl} \Omega : & \mathbb{U} & \longrightarrow & [0,1] \\ & \text{UTM} & \longmapsto & \sum_{p\in\text{Dom}[\text{UTM}]} 2^{-|p|} \end{array} \tag{20}$$

The image of this Ω function 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 Ω 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 Ω 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 $H[\text{FAS}]$ can never determine more than $H[\text{FAS}] + c$ bits of the value for Ω."
>
> 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 Ω 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[19], Gregory Chaitin provides concrete examples of how the incompleteness phenomenon can enter some fields of mathematics. Specifically, he states:

[19] Gregory Chaitin. Doing mathematics differently. https://inference-review.com/article/doing-mathematics-differently, 02 2019. Accessed: 2019-12-04

> "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[20], "*Proving Darwin: making biology mathematical*", which references many of these concepts.

[20] Gregory Chaitin. *Proving Darwin: making biology mathematical*. Vintage, 2012

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 $\mathbb{L}$. The first sentence, h, is called the hypothesis. The second sentence,* TM*, is called the protocol. Let* $\text{UTM}\colon \mathbb{L} \times \mathbb{L} \to \mathbb{L} \cup \{\nexists\}$ *be a universal Turing machine, then we say that the experiment holds if* $\text{UTM}[\text{TM}, h]$ *halts, and fails otherwise:*

$$\text{UTM}[\text{TM}, h] \begin{cases} = r & \text{halts} \implies p \text{ holds} \\ \nexists & \neg\text{halts} \implies p \text{ fails} \end{cases} \tag{21}$$

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

An experiment, so defined, is formally reproducible. Indeed, for the protocol 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 $\mathbb{D}$ as the domain* (Dom) *of science. We can define $\mathbb{D}$ in reference to a universal Turing machine* UTM *as follows:*

$$\mathbb{D} := \mathrm{Dom}[\mathrm{UTM}] \tag{22}$$

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

(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 $\mathbb{M}$ is a subset of $\mathbb{D}$:*

$$\mathbb{M} \subset \mathbb{D} \tag{23}$$

We note that the set of all possible manifests is the power set of $\mathbb{D}$:

$$\mathbb{S} := \mathcal{P}[\mathbb{D}] \tag{24}$$

Definition 7 (Observer)**.** *An observer $\mathcal{O}$ is a Turing machine that recursively enumerates the domain of science. Given a program p as input, the observer eventually halts for p iff p 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* $\mathring{\mathbb{M}}$ *which corresponds to its state:*

$$\exists!\mathring{\mathbb{M}} \tag{25}$$

- $\mathring{\mathbb{M}}$ *is called the 'reference manifest'.*
- *The symbol* $\mathbb{M}$ *will denote any manifest in* $\mathcal{P}[\mathbb{D}]$*, whereas* $\mathring{\mathbb{M}}$ *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* $\mathring{\mathbb{M}}$ *exists ontologically as* <u>*verified*</u> *facts (whereas any* $\mathbb{M} \neq \mathring{\mathbb{M}}$ *exists as* <u>*verifiable*</u> *facts but that are not yet verified).*
- *Unique to* $\mathring{\mathbb{M}}$*, and unlike other manifests, its elements are verified, and is thus subject to syllogism 1.*

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* $\mathring{\mathbb{M}}$*, and all elements of* $\mathring{\mathbb{M}}$ *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 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. Intuitively, 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 $\mathcal{N}$ is a system of mathematical work used to verify $\mathring{\mathbb{M}}$. 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 $\mathbb{M}$ can be constructed from the elements of $\mathbb{D}$, one may wonder why it is the reference manifest $\mathring{\mathbb{M}}$ that is actual and not any other.

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

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 $\mathring{\mathbb{M}}$ and not $\mathbb{M}$ 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 $\mathring{\mathbb{M}}$ to information, and it is precisely because $\mathring{\mathbb{M}}$ is randomly selected from a larger set that this is possible. We briefly recall the mathematical theory of information of Claude Shannon: Specifically, $\mathring{\mathbb{M}}$ will be interpreted as a message randomly selected from the set $\mathcal{P}[\mathbb{D}]$. Using $\rho[\mathbb{M}]$ as the probability measure, we will be able to quantify the information in the message $\mathring{\mathbb{M}}$.

It is from this connection to information that we will find our opportunity to create a physical theory. 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 $\mathcal{O}$ gains by knowing $\mathring{\mathbb{M}}$ as an element that is randomly selected from $\mathcal{P}[\mathbb{D}]$, under the constraint of nature $\mathcal{N}$.*

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.

5 Towards a mathematical proof of physics

To precisely quantify the relationship between entropy, mathematical work, and how this produces physics as a theorem, 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].

[21] Jos Uffink. Compendium of the foundations of classical statistical physics. *Philosophy of Physics*, 03 2006

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]

[22] Victor S. Batista. Postulates of statistical mechanics. http://xbeams.chem.yale.edu/~batista/vaa/node20.html. Accessed: 2019-12-17

The first assumption is responsible for implying a number of fixed macroscopic quantities, known as the constraints. Let $\mathbb{Q}$ be a set of micro-states and $\mathcal{N}$ 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:

$$\mathbb{P} := \left\{ \rho \colon \mathbb{Q} \to [0,1] \,\middle|\, \sum_{q \in \mathbb{Q}} \rho[q] = 1 \,\middle|\, \mathcal{N} \right\} \tag{26}$$

The observables, in general, are functions defined as:

$$\begin{array}{rcccl} \overline{O}_i & : & \mathbb{P} & \longrightarrow & \mathbb{R} \\ & & \rho & \longmapsto & \sum_{q \in \mathbb{Q}} \rho[q] O_i[q] \end{array} \tag{27}$$

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

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

Table 1: Typical thermodynamic quantities

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

$$\begin{array}{rcccl} S & : & \mathbb{P} & \longrightarrow & [0,\infty[\\ & & \rho & \longmapsto & -k_B \sum_{(q \in \mathbb{Q})} \rho[q] \ln \rho[q] \end{array} \tag{28}$$

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:

$$\begin{array}{rccl} \rho & : & \mathbb{Q} \times \mathbb{R}^n & \longrightarrow & [0,1] \\ & & (q, \alpha_1, \ldots, \alpha_n) & \longmapsto & Z^{-1} \exp\left(-\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q]\right) \end{array} \tag{29}$$

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

$$\begin{array}{rccl} Z & : & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & & (\alpha_1, \ldots, \alpha_n) & \longmapsto & \sum_{(q \in \mathbb{Q})} \exp\left(-\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q]\right) \end{array} \tag{30}$$

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

$$\overline{O}_i = Z^{-1} \sum_{(q \in \mathbb{Q})} O_i[q] \exp\left(-\alpha_1 O_1[q] - \cdots - \alpha_n O_n[q]\right) \tag{31}$$

The constraints are also equivalently given by the following relations:

$$\frac{\partial \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i} = \overline{O}_i \tag{32}$$

And the variance by the following n relations:

$$\frac{\partial^2 \ln Z[\alpha_1, \ldots, \alpha_n]}{\partial \alpha_i^2} = \overline{(\Delta O_i)^2} \tag{33}$$

The entropy for this ensemble is:

$$S[\alpha_1, \ldots, \alpha_n] = k_B(\ln Z + \alpha_1 \overline{O}_1 + \cdots + \alpha_n \overline{O}_n) \tag{34}$$

Taking the total derivative of the entropy, we obtain:

$$\mathrm{d}S[\alpha_1, \ldots, \alpha_n] = k_B(\alpha_1 \, \mathrm{d}\overline{O}_1 + \cdots + \alpha_n \, \mathrm{d}\overline{O}_n) \tag{35}$$

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 $\mathbb{Q}$. Under modification of $\mathbb{Q}$, usually by cross product: $\mathbb{Q} \times \mathbb{Q}_1 = \mathbb{Q}_2$, or by set complement $\mathbb{Q} \setminus \mathbb{Q}_3 = \mathbb{Q}_4$, quantities which are invariant $\{\alpha_1, \ldots, \alpha_n\}$ are called intensive, and quantities which are variant $\{O_1, O_2, \ldots, O_n\}$ are called extensive.

As an example, consider the following typical thermodynamic quantities taken from Table 1:

$$\alpha_1 := \beta \tag{36}$$
$$\alpha_2 := \gamma \tag{37}$$
$$\alpha_3 := \delta \tag{38}$$
$$O_1[q] := E[q] \tag{39}$$
$$O_2[q] := V[q] \tag{40}$$
$$O_3[q] := N[q] \tag{41}$$

the partition function would be:

$$Z[\beta, \gamma, \delta] = \sum_{q \in \mathbb{Q}} \exp\left(-\beta E[q] + \gamma V[q] + \delta N[q])\right) \tag{42}$$

The Gibbs measure is:

$$\rho(q, \beta, \gamma, \delta) = \frac{1}{Z} \exp\left(-\beta E[q] - \gamma V[q] - \delta N[q]\right) \tag{43}$$

The observables are:

$$\overline{E} = \frac{1}{Z} \sum_{q \in \mathbb{Q}} E[q] \exp\left(-\beta E[q] - \gamma V[q] - \delta N[q]\right) \tag{44}$$
$$\overline{V} = \frac{1}{Z} \sum_{q \in \mathbb{Q}} V[q] \exp\left(-\beta E[q] - \gamma V[q] - \delta N[q]\right) \tag{45}$$
$$\overline{N} = \frac{1}{Z} \sum_{q \in \mathbb{Q}} N[q] \exp\left(-\beta E[q] - \gamma V[q] - \delta N[q]\right) \tag{46}$$

The entropy is:

$$S[\beta, \gamma, \delta] = k_B(\ln Z + \beta \overline{E} + \gamma \overline{V} + \delta \overline{N}) \tag{47}$$

and the equation of state is:

$$\mathrm{d}S[\beta, \gamma, \delta] = k_B(\beta \, \mathrm{d}\overline{E} + \gamma \, \mathrm{d}\overline{V} + \delta \, \mathrm{d}\overline{N}) \tag{48}$$

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

$$Z[\beta] = \int \exp\left(-\beta H[O_1, O_2, \ldots, O_n]\right) \mathrm{d}O_1 \, \mathrm{d}O_2 \, \mathrm{d}O_3 \tag{49}$$

where, in the general case H is a scalar valued function of the constraints.

Finally, in the case where the constraints are uncountable and are functions, the partition function is to be replaced by a functional integral:

$$Z[\beta] = \int \mathcal{D}\phi \exp\left(-\beta H[\phi]\right) \tag{50}$$

5.2 *Recap: Algorithmic Thermodynamics*

Many authors[19, 7, 20, 21, 22, 23, 24, 25, 26] have discussed the similarity between the Gibbs entropy $S = -k_B \sum_{q\in\mathbb{Q}} \rho[q] \ln \rho[q]$ and the entropy in information theory $H = -\sum_{q\in\mathbb{Q}} \rho[q] \log_2 \rho[q]$. Furthermore, the similarity between the halting probability Ω and the Gibbs ensemble of statistical physics has also been studied[23]. First let us introduce Ω. Let $\mathbb{U}$ be the set of all universal Turing machines, and let UTM be an element of $\mathbb{U}$. Then, the usual definition of Ω is:

$$\Omega := \sum_{p\in\text{Dom}[\text{UTM}]} 2^{-|p|} \tag{51}$$

Here, $|p|$ denotes the length of p, a computer program. The domain, Dom[UTM], is the domain of the universal Turing machine (the set of all programs that halt for it). The sum represents the probability that a random program will halt on UTM. Chaitin's construction[24] (a.k.a. Ω, 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, Ω, is between zero and one (The Kraft inequality[25]). As the sum does not erase halting information, knowing Ω is enough to know the programs that halt and those that do not on UTM. Since the halting problem is unsolvable, Ω must, therefore, be non-computable. Ω'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 Ω. As such, Calude[26] calculated the first 64 bits of Ω for some specific universal Turing machine $u \in \mathbb{U}$ as:

$$\Omega_u = 0.000000100000010000011 0\ldots_2 \tag{52}$$

Running the calculation for a handful of bits is certainly possible, however, any finitely axiomatic systems will eventually run out of

[23] Ming Li and Paul M.B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008; Cristian S. Calude and Michael A. Stay. Natural halting probabilities, partial randomness, and zeta functions. *Inf. Comput.*, 204(11):1718–1739, November 2006; John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012; and Kohtaro Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002

[24] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975

[25] L. G. Kraft. A device for quanitizing, grouping and coding amplitude modulated pulses. Master's thesis, Mater's Thesis, Department of Electrical Engineering, MIT, Cambridge, MA, 1949

[26] Cristian S Calude, Michael J Dinneen, Chi-Kou Shu, et al. Computing a glimpse of randomness. *Experimental Mathematics*, 11(3):361–370, 2002

steam and hit a wall. Calculating the digits of π, for instance, will not hit this kind of limitation. For π, the axioms of arithmetic are sufficiently powerful to compute as many bits as we wish to calculate, limited only by the physical resources of the computers at our disposal. To understand why this is not the case for Ω, we have to realize that solving Ω 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 Ω with a multiplication constant D, which acts as an 'algorithmic decompression' term on Ω.

[27] Kohtaro Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002

$$\begin{array}{lcl} \text{Chaitin construction} & \rightarrow & \text{Tadaki ensemble} \\ \Omega = \sum\limits_{p \in \mathrm{Dom}[\mathrm{UTM}]} 2^{-|p|} & \rightarrow & \Omega[D] = \sum\limits_{p \in \mathrm{Dom}[\mathrm{UTM}]} 2^{-D|p|} \end{array} \tag{53}$$

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

$$\begin{array}{ll} \text{Gibbs ensemble} & \text{Tadaki ensemble} \\ Z[\beta] = \sum\limits_{q \in \mathrm{Q}} e^{-\beta E[q]} & \Omega[D] = \sum\limits_{p \in \mathrm{Dom}[\mathrm{UTM}]} 2^{-D|p|} \end{array} \tag{54}$$

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 $\overline{|p|}$;

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

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

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

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

[28] John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012

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 Ω with algorithmic quantities to obtain the Baez-Stay ensemble:

$$\begin{array}{cccccc} \Omega & : & \mathbb{R}^3 & \longrightarrow & \mathbb{R} \\ & & (\beta, \gamma, \delta) & \longmapsto & \sum_{p \in \mathrm{Dom[UTM]}} 2^{-\beta E[p] - \gamma V[p] - \delta N[p]} \end{array} \tag{57}$$

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].

[29] Ming Li and Paul M.B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008; and Kohtaro Tadaki. A statistical mechanical interpretation of algorithmic information theory. In *Local Proceedings of the Computability in Europe 2008 (CiE 2008)*, pages 425–434. University of Athens, Greece, Jun 2008

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:

$$\begin{array}{cccc} Z & : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & & (k, f) & \longmapsto & \sum_{p \in \mathrm{Dom}[\mathrm{UTM}]} 2^{-kL[p]-fT[p]} \end{array} \tag{58}$$

Symbol	Name	Units	Type
$L[p]$	program size	[bit]	extensive
k	computing repetency	[1/bit]	intensive
$\overline{L}$	average tape usage	[bit]	macroscopic
$T[p]$	program runtime	[operation]	extensive
f	computing frequency	[1/operation]	intensive
$\overline{T}$	average clock usage	[operation]	macroscopic

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:

$$Z = \left(\sum_{p \in \mathrm{Dom}[\mathrm{UTM}]} 2^{-kL[p]-fT[p]} \right)^n \tag{59}$$

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:

$$Z = \sum_{\mathbb{M} \in (\mathrm{Dom}[\mathrm{UTM}])^n} g[\mathbb{M}] 2^{-kL[\mathbb{M}]-tT[\mathbb{M}]} \tag{60}$$

where $g[\mathbb{M}]$ is the degeneracy of the state $\mathbb{M}$. 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 μ the *computing overhead* and conjugate it to $N[\mathbb{M}]$, 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.

Symbol	Name	Units	Type
$L[\mathbb{M}]$	size of programs in the manifest	[bit]	extensive
k	computing repetency	[1/bit]	intensive
$\overline{L}$	average tape usage	[bit]	macroscopic
$T[\mathbb{M}]$	running time of programs in the manifest	[operation]	extensive
f	computing frequency	[1/operation]	intensive
$\overline{T}$	average clock usage	[operation]	macroscopic
$N[\mathbb{M}]$	quantity of programs in the manifest	[program]	extensive
μ	computing overhead	[1/program]	intensive
$\overline{N}$	average concurrency	[program]	macroscopic

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

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

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

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

1. Computing time to program frequency formulation:

$$\begin{array}{rccl} Z' : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (k,t) & \longmapsto & \sum_{p\in\mathrm{Dom}[\mathrm{UTM}]} 2^{-kL[p]-tF[p]} \end{array} \tag{61}$$

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 $\overline{F}$ by having the programs executed under a constant computing time t:

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

2. Size-cutoff formulation:

$$\begin{array}{rcccl} Z'' & : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & & (k,l) & \longmapsto & \sum_{p\in\{q:\mathrm{Dom}[\mathrm{UTM}]\,|\,L[q]<l\}} 2^{-kL[p]} \end{array} \tag{62}$$

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

3. Time-cutoff formulation:

$$\begin{array}{rcccl} Z''' & : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & & (k,t) & \longmapsto & \sum_{p\in\{q:\mathrm{Dom}[\mathrm{UTM}]\,|\,T[q]<t\}} 2^{-kL[p]} \end{array} \tag{63}$$

The sum Z''' only includes programs that halt within a time cutoff t. Thus, Z''' contains no "non-halting information" and is computable. Ω is recovered in the limit when $t \to \infty$ (and with $k = 1$).

4. Arbitrary filter cutoff formulation:

Let $\mathbb{O} \subset \mathrm{Dom}[\mathrm{UTM}]$:

$$\begin{array}{rcccl} Z'''' & : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & k & \longmapsto & \sum_{p\in\mathbb{O}} 2^{-kL[p]} \end{array} \tag{64}$$

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

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. <u>Entropy as a measure of computational 'distance'</u>

   Consider an equation of state based on computing resources. The grand canonical partition function of algorithmic thermodynamics has the following equation of state:

$$\mathrm{d}S = k\,\mathrm{d}\overline{L} + f\,\mathrm{d}\overline{T} + \mu\,\mathrm{d}\overline{N} \tag{65}$$

   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. <u>Reservoirs of computing resources:</u>

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

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

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

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[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 $\approx 10^{122}$).

[30] Seth Lloyd. Computational capacity of the universe. *Physical Review Letters*, 88(23):237901, 2002

[31] Stephen W Hawking. Black hole explosions? *Nature*, 248(5443):30, 1974; and Damien A. Easson, Paul H. Frampton, and George F. Smoot. Entropic inflation. *International Journal of Modern Physics A*, 27(12):1250066, 2012

[32] Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, 1995

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:

$$\#\text{ops} \approx \frac{\rho_c c^5 t^4}{\hbar} \approx \frac{t^2 c^5}{G\hbar} = \frac{1}{t_p^2} t^2 \tag{66}$$

where ρ_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 $\approx \frac{c^2t^2}{l_p^2} = \frac{t^2}{t_p^2}$."

A particularly interesting consequence of this result is that these relations appear to imply conservation of both information and operations in space-time (the numerical quantity of 10^{122} is obtained by summing over all available degrees of freedom in space-time). 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: $S = k_B c^3/(4\hbar G)A$.

[33] Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333, 1973; Jacob D Bekenstein. Generalized second law of thermodynamics in black-hole physics. *Physical Review D*, 9(12):3292, 1974; and Jacob D Bekenstein. Black-hole thermodynamics. *Physics Today*, 33(1):24–31, 1980

[34] Stephen W Hawking. Black hole explosions? *Nature*, 248(5443):30, 1974

We mention Ted Jacobson[35] and his derivation of the Einstein field equation as an equation of state of a suitable thermodynamic system.

[35] Ted Jacobson. Thermodynamics of spacetime: The einstein equation of state. *Phys. Rev. Lett.*, 75:1260–1263, Aug 1995

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 η, and some legwork, the Einstein field equations are eventually recovered.

Recently, Erik Verlinde[36] 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 $T\,\mathrm{d}S = F\,\mathrm{d}x$, 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[37] 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 ([38]), the Coulomb force ([39]), Yang-Mills gauge fields ([40]), and cosmology ([41]). Some criticism has, however, been voiced[42], including by Visser[43].

Even more recently, a connection between entanglement entropy and general relativity has been supported by multiple publications[51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68].

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 carries the same philosophical limitations. My choice of designer ensembles was contingent on my prior knowledge of the laws of physics, as provided to me by the experimental sciences. 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 fun-

[36] Erik P. Verlinde. On the origin of gravity and the laws of newton. *Journal of High Energy Physics*, 2011(4):29, Apr 2011

[37] Matt Visser. Conservative entropic forces. *Journal of High Energy Physics*, 2011(10):140, 2011

[38] Lee Smolin. Newtonian gravity in loop quantum gravity. *arXiv preprint arXiv:1001.3668*, 2010

[39] Tower Wang. Coulomb force as an entropic force. *Physical Review D*, 81(10):104045, 2010

[40] Peter GO Freund. Emergent gauge fields. *arXiv preprint arXiv:1008.4147*, 2010

[41] Rong-Gen Cai, Li-Ming Cao, and Nobuyoshi Ohta. Friedmann equations from entropic force. *Physical Review D*, 81(6):061501, 2010; Miao Li and Yi Wang. Quantum uv/ir relations and holographic dark energy from entropic force. *Physics Letters B*, 687(2-3):243–247, 2010; and Damien A Easson, Paul H Frampton, and George F Smoot. Entropic accelerating universe. *Physics Letters B*, 696(3):273–277, 2011

[42] Sabine Hossenfelder. Comments on and comments on comments on verlinde's paper" on the origin of gravity and the laws of newton". *arXiv preprint arXiv:1003.1015*, 2010; Archil Kobakhidze. Gravity is not an entropic force. *Physical Review D*, 83(2):021502, 2011; Shan Gao. Is gravity an entropic force? *Entropy*, 13(5):936–948, 2011; BL Hu. Gravity and nonequilibrium thermodynamics of classical matter. *International Journal of Modern Physics D*, 20(05):697–716, 2011; and Archil Kobakhidze. Once more: gravity is not an entropic force. *arXiv preprint arXiv:1108.4161*, 2011

[43] Matt Visser. Conservative entropic forces. *Journal of High Energy Physics*, 2011(10):140, 2011

damental 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 survive falsification each and every-time it is tested outside the experimental 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 emergent 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, then so are its defects.

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:

$$\mathrm{d}\overline{E} = T\,\mathrm{d}S + p\,\mathrm{d}\overline{V} - \mu\,\mathrm{d}\overline{N} \tag{67}$$

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 fixing.

6 Physics as the probability measure that makes reality maximally informative

Personal note: It took me approximately four years to go from attempt one to the retained solution (these are just the attempts at physics mind you; the axiomatic foundation took decades of investigation). In the interim I produced approximately 50 draft papers, and over 2000 pages of notes and calculations searching for the 'right way' to think about how to fundamentally derive the physics from reality. Once I discovered the solution I am about to present, I became so convinced that it is the right way to think about physics that I almost immediately and without regret scratched (nearly) everything I had done before.

6.1 Intuition: the thermodynamics of 'clicks'

Ordinary statistical physics, in its mathematical construction, is surprisingly close to its experimental foundation. Indeed, the 'ontological backbone' of statistical physics is based, almost entirely, on the existence of instruments and measurements. For instance, these instruments could be the thermometer, the barometer, the energy-meter, the eudiometer (volume-meter), etc. Then, for a given system, an observer will make a series of measurements using one or more of these instruments. Each measurement produces a real[44] value that has a physical unit conveyed to it by the instrument used to measure it. For instance, a thermometer produces values given in Kelvin, a eudiometer produces values given in milliliters, etc. Let's call a real value with units, a *physical value*. It is then the set of measured physical values that form the 'ontological backbone' of the theory.

One can produce an 'epistemological backbone' for the system by maximizing the entropy using the method of the Lagrange multiplier, under some constraints. Specifically, constraints are defined as the average value that infinitely-many measurements will converge to. We stress, and this is very important, that one cannot make infinitely-many measurements to verify this assumption, and therefore the epistemological backbone is (slightly) disconnected from reality as a result. Nonetheless, in this description, the interpretation is clear: what exists are the measured physical values, and what is derived is the probability measure over an infinite idealization of these measurements. Yet another "micro-disconnection" from reality occurs as one usually applies the epistemological backbone to a new "virgin system", whereas in reality, the backbone has been derived from system that as already undergone some amount of measurements. Keeping track of all of these micro-disconnections will be quite im-

[44] Actually, significant figures of measurements make it such that each physical quantity are expressible as whole numbers in some equivalent system of units. For instance, a length measurement of 5.002 meter becomes a whole number when expressed in millimeters: it is 5002 millimeters. That measuring devices are able to produce "infinitely deep" numbers to recover the reals is, as far as I know, empirically unsubstantiated. Saying 'the equations I postulated requires real numbers!' is not an empirically substantiated claim. Saying 'the equations I postulated requires real numbers and they produce very good predictions as a result!' *may* render the claim eventually falsifiable, but it is still not an empirically substantiated claim. Nonetheless to keep some level of sanity and brevity but also because I doubt that it is worth the effort to rebuild everything along those lines, we will define our frameworks using the reals as it usually done is almost all physical theories. This can be interpreted as an assumption, unfortunately another possible "micro-disconnection" from reality, that significant figures can in principle be reduced as small as we want.

portant in the discussion section.

The constraints of ordinary statistical physics are scalars, and as such they are able to produce scalar macroscopic transformations via an appropriate equation of state. For instance a change from state a to b involves a scalar equation of state:

$$\overbrace{\begin{pmatrix}\overline{E}_a\\ \overline{V}_a\\ \overline{N}_a\end{pmatrix}}^{\text{state-a}} \to \underbrace{T\,\mathrm{d}\overline{E} + p\,\mathrm{d}\overline{V} + \mu\,\mathrm{d}\overline{N}}_{\text{scalar transformation}} \to \overbrace{\begin{pmatrix}\overline{E}_b\\ \overline{V}_b\\ \overline{N}_b\end{pmatrix}}^{\text{state-b}} \tag{68}$$

During my exploratory attempts, I eventually realized that the key to recover the richness of modern physics purely from statistical physics, was to extend the definition of the measurement to a *universal measurement*, that I call the 'click'. A click, like John A. Wheeler envisioned, is simply the measurement produced by a type of *universal instrument*, which we call a detector. Clicks have more structure than ordinary scalar measurements. Clicks admit a few representations: the two that we will work with are a matrix representation and a geometric algebra representation. A transformation of a system measured by clicks from state a to b, and represented as matrices, would thus be made as follows:

$$\overbrace{\begin{pmatrix}\overline{X}^a_{11} & \dots & \overline{X}^a_{1n}\\ \vdots & \ddots & \vdots\\ \overline{X}^a_{n1} & \dots & \overline{X}^a_{nn}\end{pmatrix}}^{\text{state-a}} \to \underbrace{\begin{pmatrix}\tilde{k}\,\mathrm{d}\overline{X}_{11} & \dots & \tilde{k}\,\mathrm{d}\overline{X}_{1n}\\ \vdots & \ddots & \vdots\\ \tilde{k}\,\mathrm{d}\overline{X}_{n1} & \dots & \tilde{k}\,\mathrm{d}\overline{X}_{nn}\end{pmatrix}}_{\text{linear transformation}} \to \overbrace{\begin{pmatrix}\overline{X}^b_{11} & \dots & X^b_{1n}\\ \vdots & \ddots & \vdots\\ \overline{X}^b_{n1} & \dots & \overline{X}^b_{nn}\end{pmatrix}}^{\text{state-b}} \tag{69}$$

We thus add the 'detector' to the list of thermodynamics instruments as a new type of universal instrument, and the 'click' as the universal measurement associated to this instrument, then we will make the necessary modifications to statistical physics such that we can maximize the entropy of a system constrained by clicks. Like in ordinary statistical physics, here the click forms the ontological backbone of the framework. Its derived epistemological backbone will be a probability measure obtained by maximizing the entropy of a system constrained by clicks. As we will see, this extension has remarkable consequences.

We recall that a volume constraint in thermodynamics is defined as:

$$\overline{V} = \sum_{x \in \mathbb{X}} \rho[x] V[x] \tag{70}$$

In comparison, a click will be defined as follows:

Definition 10 (Click)**.** *A click is defined as a system of equations:*

$$\overline{\lambda_1} = \sum_{x \in \mathbb{X}} \rho_1[x] \lambda_1[x] \tag{71}$$

$$\overline{\lambda_2} = \sum_{x \in \mathbb{X}} \rho_2[x] \lambda_2[x] \tag{72}$$

$$\vdots$$

$$\overline{\lambda_n} = \sum_{x \in \mathbb{X}} \rho_n[x] \lambda_n[x] \tag{73}$$

where $\lambda_1[x], \lambda_2[x], \ldots, \lambda_n[x]$ are the eigenvalue functions of a $n \times n$ matrix-valued function $\mathbf{M}[x]$. We note that each of these equations has individually the same mathematical structure as that of the scalar volume or energy constraint.

6.2 *The mathematical origin of the Born rule*

The main result of the thermodynamics of clicks 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 $\overline{E}$ as the scalar constraint, the Lagrange equation would be:

$$\mathcal{L} = -k_B \sum_{p \in \mathbb{P}} \rho[p] \ln \rho[p] + \alpha_1 \Big(-1 + \sum_{p \in \mathbb{P}} \rho[p]\Big) + \alpha_2 \Big(-\overline{E} + \sum_{p \in \mathbb{P}} \rho[p] E[p]\Big) \tag{74}$$

where α_1, α_2 are the Lagrange multipliers. Then, extremalizing it, one obtains the Gibbs measure:

$$\frac{\partial \mathcal{L}}{\partial \rho[q]} = 0 \implies \rho[p] = \frac{1}{Z} \exp\big(-\beta E[p]\big) \tag{75}$$

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

Now, for each of n eigenvalues, we can define a Lagrange equation as follows:

$$\mathcal{L}_1 = -k_B \sum_{p\in\mathbb{P}} \rho_1[p] \ln \rho_1[p] + \alpha_1(-1 + \sum_{p\in\mathbb{P}} \rho_1[p]) + \alpha_2(-\overline{\lambda}_1 + \sum_{p\in\mathbb{P}} \rho_1[p]\lambda_1[p]) \tag{76}$$

$$\vdots$$

$$\mathcal{L}_n = -k_B \sum_{p\in\mathbb{P}} \rho_n[p] \ln \rho_n[p] + \alpha_1(-1 + \sum_{p\in\mathbb{P}} \rho_n[p]) + \alpha_2(-\overline{\lambda}_n + \sum_{p\in\mathbb{P}} \rho_n[p]\lambda_n[p]) \tag{77}$$

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

$$0 = \frac{\partial \mathcal{L}_1}{\partial \rho_1[q]} = -k_B - k_B \ln \rho_1[q] + \alpha_1 + \alpha_2 \lambda_1[p] \tag{78}$$

$$\vdots$$

$$0 = \frac{\partial \mathcal{L}_n}{\partial \rho_n[q]} = -k_B - k_B \ln \rho_n[q] + \alpha_1 + \alpha_2 \lambda_n[p] \tag{79}$$

Solving for $\rho[q]$, we obtain

$$\rho_1[p] = \exp\left(\frac{-k_B + \alpha_1}{k_B}\right)\exp\left(\frac{\alpha_2}{k_B}\lambda_1[p]\right) \tag{80}$$

$$\vdots$$

$$\rho_n[p] = \exp\left(\frac{-k_B + \alpha_1}{k_B}\right)\exp\left(\frac{\alpha_2}{k_B}\lambda_n[p]\right) \tag{81}$$

Using the normalization constraint on $\rho[p]$, we find n 'eigen-partition function':

$$1 = \frac{1}{A_1}\sum_{p\in\mathbb{P}}\exp\left(\frac{-k_B + \alpha_1}{k_B}\right)\exp\left(\frac{\alpha_2}{k_B}\lambda_1[q]\right) \implies \exp\left(\frac{-k_B + \alpha_1}{k_B}\right) = \left(\frac{1}{A_1}\sum_{p\in\mathbb{P}}\exp\left(\frac{\alpha_2}{k_B}\lambda_n[p]\right)\right)^{-1} \tag{82}$$

$$\vdots$$

$$1 = \frac{1}{A_n}\sum_{p\in\mathbb{P}}\exp\left(\frac{-k_B + \alpha_1}{k_B}\right)\exp\left(\frac{\alpha_2}{k_B}\lambda_n[q]\right) \implies \exp\left(\frac{-k_B + \alpha_1}{k_B}\right) = \left(\frac{1}{A_n}\sum_{p\in\mathbb{P}}\exp\left(\frac{\alpha_2}{k_B}\lambda_n[p]\right)\right)^{-1} \tag{83}$$

Finally, the extended Born rule appears, from universal thermodynamics, by multiplying the eigen-partition functions.

Definition 11 (Extended Born rule (diagonal case)). *We multiply the eigen-partition functions:*

$$\|\mathbf{Z}\| = Z_1 Z_2 \dots Z_n \tag{84}$$

Let us now apply this result to a few examples:

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

Proof. Trivial □

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

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

$$\mathbf{M}[p] = \begin{pmatrix} -\beta E[p] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -\beta E[p] \end{pmatrix} \tag{85}$$

And the parition function is:

$$\|\mathbf{Z}\| = \left(\sum_{p \in \mathbb{P}} \exp -\beta E[p]\right)^n \tag{86}$$

□

Theorem 3 (Quantum probabilities (diagonal case))**.** *In the case where* $\mathbf{M}[p]$ *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:

$$\mathbf{M}'[p] = \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix} \tag{87}$$

We note that here $\mathbf{M}[p]$ is a matrix representation of the complex numbers, via the group isomorphism of $a + ib \cong \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. The eigenvalues of $\mathbf{M}[p]$ are:

$$\lambda_1 = r[p] + i\theta[p] \tag{88}$$

$$\lambda_2 = r[p] - i\theta[p] \tag{89}$$

The universal Born rule becomes:

$$\|\mathbf{Z}\| = \left(\sum_{p \in \mathbb{P}} \exp\big(r[p] + i\theta[p]\big)\right)\left(\sum_{p \in \mathbb{P}} \exp\big(r[p] - i\theta[p]\big)\right) \tag{90}$$

If we now take a two state ensemble $\mathbb{P} := \{p_1, p_2\}$, we get:

$$\|\mathbf{Z}\| = \Big(\exp\big(r[p_1] + i\theta[p_1]\big) + \exp\big(r[p_2] + i\theta[p_2]\big)\Big)\Big(\exp\big(r[p_1] - i\theta[p_1]\big) + \exp\big(r[p_2] - i\theta[p_2]\big)\Big) \tag{91}$$

With straightforward algebraic manipulation and simplifications, we get:

$$= (e^{r[p_1]})^2 + (e^{r[p_2]})^2 + 2e^{r[p_1]}e^{r[p_2]}\cos[\theta[p_2] - \theta[p_1]] \tag{92}$$

This is the typical quantum probability of a two-state system along with the interference term. □

6.3 Recovery of the formalism of quantum mechanics

We note that when we multiply n identical canonical partition functions to obtain a grand-partition function, a sum is made over tuples comprised of the states of each contributing canonical partition function. For instance, a canonical partition function over a set $\mathbb{X}$ will, when multiplied by a canonical partition function of a set $\mathbb{Y}$, produce a grand-canonical partition function over the set $\mathbb{X} \times \mathbb{Y}$. The resulting grand-canonical partition function retains its classical probability interpretation, because a Gibbs measure can be defined for it just as it was possible to do so for the states of each of the contributing canonical partition functions. Whether one multiplies canonical partition functions into grand-canonical partition functions, or splits out a canonical subset to eliminate it from a grand canonical partition function (via set complement $(\mathbb{X} \times \mathbb{Y})/\mathbb{Y} = \mathbb{X}$), the probability measure remains classical throughout the modifications.

This is not the case with eigen-partition functions. As such, the state produced by their multiplication, as it contains an interference term, cannot be understood generally as a the pairing of two classical states.

Let us see in more details. We recall that we have defined, in universal statistical physics, the partition function as the multiplication of each eigen-partition function. In the case of the matrix representation of complex numbers, we obtained:

$$\|\mathbf{Z}\| = Z(Z)^* \tag{93}$$

Now, consider that we define n such partition functions, one for each of a different system $\mathbb{P}_1, \ldots, \mathbb{P}_n$. The partition functions are:

$$\left\|\mathbf{Z}[\mathbb{P}_1]\right\| = Z[\mathbb{P}_1](Z[\mathbb{P}_1])^* \tag{94}$$

$$\vdots$$

$$\left\|\mathbf{Z}[\mathbb{P}_n]\right\| = Z[\mathbb{P}_n](Z[\mathbb{P}_n])^* \tag{95}$$

Now, we can define yet another partition functions as a sum of the previous partition functions. As we will see shortly, this definition is equivalent to the normalization condition of the wavefunction. Consequently, let us use $\langle\psi|\psi\rangle$ to refer to this partition function right away:

$$\langle\psi|\psi\rangle = \left\|\mathbf{Z}[\mathbb{P}_1]\right\| + \cdots + \left\|\mathbf{Z}[\mathbb{P}_n]\right\| \tag{96}$$

Here is an example as the sum of two partition functions:

$$\langle\psi|\psi\rangle = \Big(|\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2|\cos[\varphi_2 - \varphi_1]\Big) + \Big(|\psi_3|^2 + |\psi_4|^2 + 2|\psi_3||\psi_4|\cos[\varphi_4 - \varphi_3]\Big) \tag{97}$$

This definition recovers the same form as that of the quantum probability rules for n orthogonal states, usually expressed in the formalism of quantum mechanics as a column vector with n entries, of a discreet Hilbert space. For instance, a column vector given as:

$$|\psi\rangle = \begin{pmatrix} \psi_1+\psi_2 \\ \psi_3+\psi_4 \\ \vdots \end{pmatrix} \tag{98}$$

Will produce the following probability:

$$\begin{aligned}\langle\psi|\psi\rangle &= |\psi_1 + \psi_2|^2 + |\psi_3 + \psi_4|^2 + \dots & (99)\\ &= \Big(|\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2|\cos[\varphi_2 - \varphi_1]\Big) + \Big(|\psi_3|^2 + |\psi_4|^2 + 2|\psi_3||\psi_4|\cos[\varphi_4 - \varphi_3]\Big) + \dots & (100)\end{aligned}$$

which is the same result as that given by our partition function.

Likewise, our method via the partition function can easily be extended to the continuum under an appropriate continuous parametrization and limiting process (the sum is extended to infinitely many terms, and each term is proportionally reduced to a probability density), to obtain:

$$\langle\psi|\psi\rangle = \int_{-\infty}^{\infty} |Z[x]|^2 \,\mathrm{d}x \tag{101}$$

If the above integral yields a finite value, then we can further associate a probability measure representing the probability that the system is in a specific state within the range $[a, b]$ as:

$$\rho[a, b] = \frac{1}{\langle\psi|\psi\rangle} \int_a^b |Z[x]|^2 \,\mathrm{d}x \tag{102}$$

These are the same relations as those obtained by the formalism of Hilbert spaces in ordinary quantum mechanics, but here resulting from universal statistical physics entirely.

Let us now add a statistical physics observable to (96). Let $\mathbf{O}$ be a diagonal matrix with real entries, and let $O[\mathbb{P}]$ be a real-valued function, then:

$$\overline{O} = \langle\psi|\,\mathbf{O}\,|\psi\rangle = O[\mathbb{P}_1]\big\|\mathbf{Z}[\mathbb{P}_1]\big\| + \dots + O[\mathbb{P}_n]\big\|\mathbf{Z}[\mathbb{P}_n]\big\| \tag{103}$$

The pre-existing requirement that observables of statistical physics be real valued, allows us to meet the requirement that $\mathbf{O}$ be Hermitian.

Now, lets make this interesting by kicking it up a notch:

6.4 *The generalized probability measure of 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 probabilities to matrices (including non-diagonal matrices):

Definition 12 (Extended Born rule)**.** *Let $\mathbb{P}$ be a countable set and let $\mathbf{M}[p]$ be a $n \times n$ matrix. We define the extended Born rule as follows:*

$$\|\mathbf{Z}\| = \det \sum_{p \in \mathbb{P}} \exp \mathbf{M}[p] \tag{104}$$

In the case where $\mathbf{M}[p]$ is diagonal, we recover definition (11).

Theorem 4 (Quantum probabilities)**.** *In the case where $\mathbf{M}[p]$ is the matrix representation of the complex numbers, one recovers the familiar probability amplitude and Born rule of quantum mechanics, even if $\mathbf{M}[p]$ is non-diagonal.*

Proof. 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:

$$\mathbf{M}[p] = \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix} \tag{105}$$

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

$$\|\mathbf{Z}\| = \det \sum_{p \in \mathbb{P}} \exp \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix} \tag{106}$$

$$\tag{107}$$

The matrix exponential reduces to the following expression:

$$\exp \begin{pmatrix} r[p] & \theta[p] \\ -\theta[p] & r[p] \end{pmatrix} = \begin{pmatrix} \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \exp \begin{pmatrix} r[p] - i\theta[p] & 0 \\ 0 & r[p] + i\theta[p] \end{pmatrix} \begin{pmatrix} \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{108}$$

$$= \begin{pmatrix} e^{r[p]} \cos[\theta[p]] & e^{r[p]} \sin[\theta[p]] \\ -e^{r[p]} \sin[\theta[p]] & e^{r[p]} \cos[\theta[p]] \end{pmatrix} \tag{109}$$

If we now take a two state ensemble $\mathbb{P} := \{p_1, p_2\}$, we get:

$$\|\mathbf{Z}\| = \det\left[\begin{pmatrix} e^{r[p_1]}\cos[\theta[p_1]] & e^{r[p_1]}\sin[\theta[p_1]] \\ -e^{r[p_1]}\sin[\theta[p_1]] & e^{r[p_1]}\cos[\theta[p_1]] \end{pmatrix} + \begin{pmatrix} e^{r[p_2]}\cos[\theta[p_2]] & e^{r[p_2]}\sin[\theta[p_2]] \\ -e^{r[p_2]}\sin[\theta[p_2]] & e^{r[p_2]}\cos[\theta[p_2]] \end{pmatrix}\right] \tag{110}$$

With straightforward algebraic manipulation and simplifications, we get:

$$= (e^{r[p_1]})^2 + (e^{r[p_1]})^2 + 2e^{r[p_1]}e^{r[p_2]}\cos[\theta[p_2] - \theta[p_1]] \tag{111}$$

□

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 $\mathbf{M}[p]$ is the matrix representation of a complex number. However, it is also possible to interpret $\mathbf{M}[p]$ as the matrix representation of the even sub-algebra of $\mathcal{G}_2(\mathbb{R})$ (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 $\mathcal{G}_2(\mathbb{R})$, an element of the even sub-algebra is:

$$\mathbf{v} = r + \theta\mathbf{I} \tag{112}$$

which is group isomorphic with the complex. For $\mathcal{G}_{3,1}(\mathbb{R})$, an element of the even sub-algebra is:

$$\mathbf{v} = r + \mathbf{F} + \theta\mathbf{I} \tag{113}$$

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[45]:

[45] David Hestenes. Spacetime physics with geometric algebra. *American Journal of Physics*, 71(7):691–714, 2003

$$\psi = \exp\left(\frac{1}{2}(r + \mathbf{F} + \theta\mathbf{I})\right) = R\exp\left(\frac{1}{2}(r + \theta\mathbf{I})\right) \tag{114}$$

where $R = \exp\mathbf{F}/2$ is a rotor.

Theorem 5 (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:

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{115}$$

$$\sigma_{01} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \quad \sigma_{02} = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix} \quad \sigma_{03} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad \sigma_{12} = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix} \quad \sigma_{31} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

$$\sigma_{23} = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \end{pmatrix} \tag{116}$$

$$\mathbf{I} = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ -i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} \tag{117}$$

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

$$\mathbf{M}[p] = r[p]I + F_{01}[p]\sigma_{01} + F_{02}[p]\sigma_{02} + F_{03}[p]\sigma_{03} + F_{23}[p]\sigma_{23} + F_{13}[p]\sigma_{13} + F_{12}[p]\sigma_{12} + \theta[p]\mathbf{I} \tag{118}$$

$$= \begin{pmatrix} r[p]+iF_{12}[p] & F_{13}[p]+iF_{23}[p] & -F_{03}[p]-i\theta[p] & -F_{01}[p]+iF_{02}[p] \\ -F_{13}[p]+iF_{23}[p] & r[p]-iF_{12}[p] & -F_{01}[p]-iF_{02}[p] & F_{03}[p]-i\theta[p] \\ -F_{03}[p]-i\theta[p] & -F_{01}[p]+iF_{02}[p] & r[p]+iF_{12}[p] & F_{13}[p]+iF_{23}[p] \\ -F_{01}[p]-iF_{02}[p] & F_{03}[p]-i\theta[p] & -F_{13}[p]+iF_{23}[p] & r[p]-iF_{12}[p] \end{pmatrix} \tag{119}$$

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

$$\det \exp \frac{1}{2}\mathbf{M}[p] = \exp \operatorname{Tr} \frac{1}{2}\mathbf{M}[p] = \exp 2r[p] \tag{120}$$

Which yields the same result as in the complex case. □

The extended Born rule then automatically cancels out the rotor (via the relation $R\widetilde{R} = 1$) as well as the complex part (via the relation the square modulus) and maps ψ directly to a real probability value. The scope of this cancellation effect 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 $\mathbb{M}(n,\mathbb{C})$. The effect is to reduce the microscopic domain to the set of all general linear matrices via the well-known correspondence:

$$\exp : \mathbb{M}(n,\mathbb{C}) \to \mathrm{GL}(n,\mathbb{C}) \tag{121}$$

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 $\exp \mathbf{M} \exp -\mathbf{M} = I$. 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 $\mathcal{G}_4(\mathbb{R})$ 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 multivector simply extends the rotor component to the group of versors which will now accounts for all possible Lorentz transformation, including reflections and inversions.

Theorem 6 (Extended Born rule for a complete multi-vector of $\mathcal{G}_4(\mathbb{R})$)**.**

Proof. The inclusion of the odd part to the microscopic element allows us to express all Lorentz transformations, extending the rotors made available by the even sub-algebra to now include those of space and time inversions and reflections. 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 $\mathcal{G}_4(\mathbb{C})$ as the representation for $\mathbb{M}(4,\mathbb{C})$ matrices. With it we can associate the determinant of a general 4×4 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 $\mathcal{G}_4(\mathbb{C})$ comprises the full set of $\mathbb{M}(4,\mathbb{C})$; the set of 4×4 matrix with complex entries.

First, let us note that the Dirac matrices form the generators of the basis of $\mathbb{M}(4,\mathbb{C})$. 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 $v_\mu = \gamma_5\gamma_\mu$

(where $\gamma_a\gamma_b$ is the usual matrix multiplication.)
Explicitly, the 16 matrices are:

$$I = \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix} \tag{122}$$

$$\gamma_0 = \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&-1&0\\ 0&0&0&-1 \end{pmatrix} \quad \gamma_1 = \begin{pmatrix} 0&0&0&1\\ 0&0&1&0\\ 0&-1&0&0\\ -1&0&0&0 \end{pmatrix} \quad \gamma_2 = \begin{pmatrix} 0&0&0&-i\\ 0&0&i&0\\ 0&i&0&0\\ -i&0&0&0 \end{pmatrix} \quad \gamma_3 = \begin{pmatrix} 0&0&1&0\\ 0&0&0&-1\\ -1&0&0&0\\ 0&1&0&0 \end{pmatrix} \tag{123}$$

$$\sigma_{01} = \begin{pmatrix} 0&0&0&-1\\ 0&0&-1&0\\ 0&-1&0&0\\ -1&0&0&0 \end{pmatrix} \quad \sigma_{02} = \begin{pmatrix} 0&0&0&i\\ 0&0&-i&0\\ 0&i&0&0\\ -i&0&0&0 \end{pmatrix} \quad \sigma_{03} = \begin{pmatrix} 0&0&-1&0\\ 0&0&0&1\\ -1&0&0&0\\ 0&1&0&0 \end{pmatrix} \quad \sigma_{12} = \begin{pmatrix} i&0&0&0\\ 0&-i&0&0\\ 0&0&i&0\\ 0&0&0&-i \end{pmatrix} \quad \sigma_{31} = \begin{pmatrix} 0&1&0&0\\ -1&0&0&0\\ 0&0&0&1\\ 0&0&-1&0 \end{pmatrix}$$

$$\sigma_{23} = \begin{pmatrix} 0&i&0&0\\ i&0&0&0\\ 0&0&0&i\\ 0&0&i&0 \end{pmatrix} \tag{124}$$

$$v_0 = \begin{pmatrix} 0&0&i&0\\ 0&0&0&i\\ -i&0&0&0\\ 0&-i&0&0 \end{pmatrix} \quad v_1 = \begin{pmatrix} 0&i&0&0\\ i&0&0&0\\ 0&0&0&-i\\ 0&0&-i&0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 0&1&0&0\\ -1&0&0&0\\ 0&0&0&-1\\ 0&0&1&0 \end{pmatrix} \quad v_3 = \begin{pmatrix} i&0&0&0\\ 0&-i&0&0\\ 0&0&-i&0\\ 0&0&0&i \end{pmatrix} \tag{125}$$

$$\gamma_5 = \begin{pmatrix} 0&0&-i&0\\ 0&0&0&-i\\ -i&0&0&0\\ 0&-i&0&0 \end{pmatrix} \tag{126}$$

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):

$$\begin{aligned} \mathbf{M} = {} & (X_0)\gamma_0 + (X_1)\gamma_1 + (X_2)\gamma_2 + (X_3)\gamma_3 \\ & + (E_1)\sigma_{01} + (E_2)\sigma_{02} + (E_3)\sigma_{03} + (B_1)\sigma_{23} + (B_2)\sigma_{31} + (B_3)\sigma_{12} \\ & + (V_0)v_0 + (V_1)v_1 + (V_2)v_2 + (V_3)v_3 \\ & + (a) + (R)\gamma_5 \end{aligned} \tag{127}$$

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:

$$\begin{aligned} \mathbf{u} = {} & (X_0)\hat{\mathbf{x}}_0 + (X_1)\hat{\mathbf{x}}_1 + (X_2)\hat{\mathbf{x}}_2 + (X_3)\hat{\mathbf{x}}_3 \\ & + (E_1)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 + (E_2)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_2 + (E_3)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_3 + (B_1)\hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 + (B_2)\hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_3 + (B_3)\hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \\ & + (V_0)\hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 + (V_1)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 + (V_2)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_3 + (V_3)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \\ & + (a) + (R)\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 \end{aligned} \tag{128}$$

There exists an bijective map between the elements of $\mathcal{G}_4(\mathbb{C})$ and those of $\mathbb{M}(4, \mathbb{C})$:

$$\mathbf{M} = M[\mathbf{g}] \qquad M^{-1}[\mathbf{M}] = \mathbf{u} \tag{129}$$

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

$$\mathbf{M}^2 = M[\mathbf{u}^2] \tag{130}$$

The map is realized by replacing the basis $\hat{\mathbf{x}}_i$ by the gamma matrix γ_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 $\mathcal{G}_4(\mathbb{R})$:

$$\psi[p] = e^{\frac{1}{2}r[p]} \exp\left(\frac{1}{2}(\theta[p]\mathbf{I} + \mathbf{X} + \mathbf{F} + \mathbf{V})\right) = e^{\frac{1}{2}r[p]}\mathbf{G} \tag{131}$$

where $\mathbf{F}$ is the previously defined bivector, and where:

$$\mathbf{X} = X_0\hat{\mathbf{x}}_0 + X_1\hat{\mathbf{x}}_1 + X_2\hat{\mathbf{x}}_2 + X_3\hat{\mathbf{x}}_3 \tag{132}$$

$$\mathbf{V} = V_0\hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 + V_1\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_2 \wedge \hat{\mathbf{x}}_3 + V_2\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_3 + V_3\hat{\mathbf{x}}_0 \wedge \hat{\mathbf{x}}_1 \wedge \hat{\mathbf{x}}_2 \tag{133}$$

and finally where

$$\mathbf{G} = \exp\left(\frac{1}{2}(\theta[p]\mathbf{I} + \mathbf{X} + \mathbf{F} + \mathbf{V})\right) \tag{134}$$

Transformations of the wavefunction, such a Lorentz boots, rotation, change of frame of reference, inversions, reflections, etc, can done by sandwiching: $\psi' = L\psi L^{-1}$, where L is an element of the versor group, a subgroup of the general linear group. We can further require that $LL^{-1} = 1$ (in which case the versor group becomes the pin group), but this is may not be an absolute requirement as we normalize the partition function in any case. In the matrix representation, $\mathbf{M}[p]$ is:

$$\mathbf{M}[p] = \begin{pmatrix} a+X_0-iB_3-iV_3 & B_2-iB_1+V_2-iV_1 & -ib+X_3+E_3-iV_0 & X_1-iX_2+E_1-iE_2 \\ -B_2-iB_1-V_2-iV_1 & a+X_0+iB_3+iV_3 & X_1+iX_2+E_1+iE_2 & -ib-X_3-E_3-iV_0 \\ -ib-X_3+E_3+iV_0 & -X_1+iX_2+E_1-iE_2 & a-X_0-iB_3+iV_3 & B_2-iB_1-V_2+iV_1 \\ -X_1-iX_2+E_1+iE_2 & -ib+X_3-E_3+iV_0 & -B_2-iB_1+V_2+iV_1 & a-X_0+iB_3-iV_3 \end{pmatrix} \tag{135}$$

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

$$\det\exp\frac{1}{2}\mathbf{M}[p] = \exp\operatorname{Tr}\frac{1}{2}\mathbf{M}[p] = e^{2r[p]} \tag{136}$$

and thus reduces to a real number yielding the probability.

□

In fact, if we consider the constraint of a system of universal thermodynamics to be an arbitrary matrix $\mathbb{M}(n, \mathbb{C})$, and possibly even with the subset of matrix representations of the real geometric algebra $M[\mathcal{G}_n(\mathbb{R})] \subset \mathbb{M}(n, \mathbb{C})$, we get a no-go theorem regarding the number of dimensions the system can have:

Theorem 7 (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 $\mathbb{M}(n, \mathbb{C})$ matrix,*

then for a general/arbitrary matrix, the dimensions stops at 4×4 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 $\mathcal{G}_4(\mathbb{C})$ form a complete representation of any elements of $\mathbb{M}(4, \mathbb{C})$.

Then, it follows that the characteristic polynomial associated with the matrix representation of 4×4 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 4×4, 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 $\mathbb{M}(4,\mathbb{C})$ and $\mathcal{G}_4(\mathbb{C})$ —allowing for a purely geometric interpretation of $\mathbb{M}(4,\mathbb{C})$—, 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 $\mathcal{G}_4(\mathbb{C})$ 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 $\mathcal{G}_4(\mathbb{C})$:

$$\mathbf{v} = X_0\hat{\mathbf{x}}_0 + X_1\hat{\mathbf{x}}_1 + X_2\hat{\mathbf{x}}_2 + X_3\hat{\mathbf{x}}_3 \tag{137}$$

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

The matrix representation of $\mathbf{v}$ is:

$$M[\mathbf{v}] = \begin{pmatrix} X_0 & 0 & X_3 & X_1 - iX_2 \\ 0 & X_0 & X_1 + iX_2 & -X_3 \\ -X_3 & -X_1 + iX_2 & -X_0 & 0 \\ -X_1 - iX_2 & X_3 & 0 & -X_0 \end{pmatrix} \tag{138}$$

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

$$\lambda^2 = X_0^2 - X_1^2 - X_2^2 - X_3^2 \tag{139}$$

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.

6.5 *Space-time interference pattern*

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

Theorem 8 (Hyperbolic interference (1D))**.** *Now we will apply the geometric properties of the matrix representation of $\mathcal{G}_2(\mathbb{R})$ to the universal Born rule:*

$$\mathbf{I} \cong \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \qquad \hat{\mathbf{x}}_1 \cong \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \qquad \hat{\mathbf{x}}_2 \cong \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \qquad I \cong \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \tag{140}$$

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

Proof. To start, we define two maps:

$$r : \mathbb{P} \to \mathbb{R} \tag{141}$$
$$x : \mathbb{P} \to \mathbb{R} \tag{142}$$

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

$$\mathbf{M}[p] = r[p]\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x[p]\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} r[p] + x[p] & 0 \\ 0 & r[p] - x[p] \end{pmatrix} \tag{143}$$

The universal Born rule is:

$$\begin{aligned} \|\mathbf{Z}\| &= \det \sum_{p\in\mathbb{P}} \begin{pmatrix} \exp\big(r[p] + x[p]\big) & 0 \\ 0 & \exp\big(r[p] - x[p]\big) \end{pmatrix} &(144)\\ &= \left(\sum_{p\in\mathbb{P}} \exp\big(r[p] + x[p]\big)\right)\left(\sum_{p\in\mathbb{P}} \exp\big(r[p] - x[p]\big)\right) &(145) \end{aligned}$$

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

$$\begin{aligned} &= \Big(\exp\big(r[p_1] + x[p_1]\big) + \exp\big(r[p_2] + x[p_2]\big)\Big)\Big(\exp\big(r[p_1] - x[p_1]\big) + \exp\big(r[p_2] - x[p_2]\big)\Big) &(146)\\ &= \exp\big(r[p_1] + x[p_1]\big)\exp\big(r[p_1] - x[p_1]\big) + \exp\big(r[p_2] + x[p_2]\big)\exp\big(r[p_2] - x[p_2]\big) \\ &\qquad + \exp\big(r[p_1] + x[p_1]\big)\exp\big(r[p_2] - x[p_2]\big) + \exp\big(r[p_2] + x[p_2]\big)\exp\big(r[p_1] - x[p_1]\big) &(147)\\ &= (e^{r[p_1]})^2 + (e^{r[p_2]})^2 + 2e^{r[p_1]}e^{r[p_2]}\cosh\big(x[p_1] - x[p_2]\big) &(148) \end{aligned}$$

Here, we obtain a hyperbolic interference term in lieu of the cosine term normally present in quantum mechanics. □

Theorem 9 (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:

$$r : \mathbb{P} \to \mathbb{R} \tag{149}$$
$$x : \mathbb{P} \to \mathbb{R} \tag{150}$$
$$y : \mathbb{P} \to \mathbb{R} \tag{151}$$

The resulting matrix is obtained by summing:

$$\mathbf{M}[p] = r[p]\mathbf{I} + x[p]\hat{\mathbf{x}}_1 + y[p]\hat{\mathbf{x}}_2 \tag{152}$$

we get:

$$\mathbf{M}[p] = \begin{pmatrix} r[p] + x[p] & y[p] \\ y[p] & r[p] - x[p] \end{pmatrix} \tag{153}$$

The universal Born rule is:

$$\|\mathbf{Z}\| = \det \sum_{p \in \mathbb{P}} \exp \begin{pmatrix} r[p] + x[p] & y[p] \\ y[p] & r[p] - x[p] \end{pmatrix} \tag{154}$$

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

$$\|\mathbf{Z}\| = \det \sum_{p \in \mathbb{P}} e^{r[p]} \begin{pmatrix} \cosh \sqrt{x[p]^2 + y[p]^2} + \frac{x[p] \sinh \sqrt{x[p]^2+y[p]^2}}{\sqrt{x[p]^2+y[p]^2}} & \frac{y[p] \sinh \sqrt{x[p]^2+y[p]^2}}{\sqrt{x[p]^2+y[p]^2}} \\ \frac{y[p] \sinh \sqrt{x[p]^2+y[p]^2}}{\sqrt{x[p]^2+y[p]^2}} & \cosh \sqrt{x[p]^2 + y[p]^2} - \frac{x[p] \sinh \sqrt{x[p]^2+y[p]^2}}{\sqrt{x[p]^2+y[p]^2}} \end{pmatrix} \tag{155}$$

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

$$\|\mathbf{Z}\| = |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2| \left(\cosh \sqrt{x_1^2 + y_1^2} \cosh \sqrt{x_2^2 + y_2^2} - \frac{(x_1 x_2 + y_1 y_2) \sinh \sqrt{x_1^2 + y_1^2} \sinh \sqrt{x_2^2 + y_2^2}}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}} \right) \tag{156}$$

where $x_1 x_2 + y_1 y_2$ is an inner product between the two states.

□

Then, finally, to a space-time event.

Definition 13 (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:*

$$\mathbf{I}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\quad\gamma_0=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}\quad\gamma_1=\begin{pmatrix}0&0&0&1\\0&0&1&0\\0&-1&0&0\\-1&0&0&0\end{pmatrix}\quad\gamma_2=\begin{pmatrix}0&0&0&-i\\0&0&i&0\\0&i&0&0\\-i&0&0&0\end{pmatrix}\quad\gamma_3=\begin{pmatrix}0&0&1&0\\0&0&0&-1\\-1&0&0&0\\0&1&0&0\end{pmatrix}\tag{157}$$

One introduces fours maps:

$$r:\mathbb{P}\to\mathbb{R}\tag{158}$$
$$X_0:\mathbb{P}\to\mathbb{R}\tag{159}$$
$$X_1:\mathbb{P}\to\mathbb{R}\tag{160}$$
$$X_2:\mathbb{P}\to\mathbb{R}\tag{161}$$
$$X_3:\mathbb{P}\to\mathbb{R}\tag{162}$$

Then, the matrix $\mathbf{M}[p]$ *is:*

$$\mathbf{M}[p]=r[p]\mathbf{I}+X_0[p]\gamma_0+X_1[p]\gamma_1+X_2[p]\gamma_2+X_3[p]\gamma_3\tag{163}$$

and its representation is:

$$\mathbf{M}[p]=\begin{pmatrix}r+X_0&0&X_3&X_1-iX_2\\0&r+X_0&X_1+iX_2&-X_3\\-X_3&-X_1+iX_2&r-X_0&0\\-X_1-iX_2&X_3&0&r-X-0\end{pmatrix}\tag{164}$$

Using the matrix representation leads to a substantially verbose proof. Instead, we will remain in the language of geometric algebra. We can write $\exp\mathbf{M}[p]$ *as:*

$$\begin{aligned}\exp\mathbf{M}[p]&=e^{r[p]}\exp\bigl(X_0[p]\hat{\mathbf{x}}_0+X_1[p]\hat{\mathbf{x}}_1+X_2[p]\hat{\mathbf{x}}_2+X_3[p]\hat{\mathbf{x}}_3\bigr)&(165)\\&=e^{r[p]}\exp\bigl(\mathbf{X}[p]\bigr)&(166)\end{aligned}$$

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

$$\|\mathbf{Z}\|=\det\Bigl(e^{r[p_1]}\exp\bigl(\mathbf{X}[p_1]\bigr)+e^{r[p_2]}\exp\bigl(\mathbf{X}[p_2]\bigr)\Bigr)\tag{167}$$

We note three observations:

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

$$\begin{aligned}\|\mathbf{v}\|&=\sqrt{(r+\mathbf{x})(r-\mathbf{x})}&(168)\\&=\sqrt{r^2+\mathbf{x}^2}&(169)\end{aligned}$$

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

$$\exp \mathbf{x} = \cosh\|\mathbf{x}\| + \frac{\mathbf{x}}{\|\mathbf{x}\|}\sinh\|\mathbf{x}\| \tag{170}$$

3. *We note that the Clifford conjugate of* $\exp \mathbf{x}$ *is:*

$$(\exp \mathbf{x})^{\square} = \cosh\|\mathbf{x}\| - \frac{\mathbf{x}}{\|\mathbf{x}\|}\sinh\|\mathbf{x}\| \tag{171}$$

4. *Finally, we adopt the notation* $\mathbf{X}[p_i] = \mathbf{x}_i$ *and* $r[p_i] = r_i$*.*

With these observations, we can now find the expression of $\|\mathbf{Z}\|$*:*

$$\begin{aligned}
\sqrt{\|\mathbf{Z}\|} = {} & \left(e^{r_1}\cosh\|\mathbf{x}_1\| + e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| + e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right) \\
& \left(e^{r_1}\cosh\|\mathbf{x}_1\| - e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| - e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right)
\end{aligned} \tag{172}$$

$$\begin{aligned}
= {} & e^{r_1}\cosh\|\mathbf{x}_1\|\left(e^{r_1}\cosh\|\mathbf{x}_1\| - e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| - e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right) \\
& + e^{r_2}\cosh\|\mathbf{x}_2\|\left(e^{r_1}\cosh\|\mathbf{x}_1\| - e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| - e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right) \\
& + e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\|\left(e^{r_1}\cosh\|\mathbf{x}_1\| - e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| - e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right) \\
& + e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\left(e^{r_1}\cosh\|\mathbf{x}_1\| - e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\| + e^{r_2}\cosh\|\mathbf{x}_2\| - e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\right)
\end{aligned} \tag{173}$$

$$\begin{aligned}
= {} & e^{2r_1}\cosh^2\|\mathbf{x}_1\| - e^{2r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\cosh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_1\| + e^{r_1}e^{r_2}\cosh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_2\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\cosh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_2\| \\
& + e^{r_2}e^{r_1}\cosh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_1\| - e^{r_2}e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\cosh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_1\| + e^{2r_2}\cosh^2\|\mathbf{x}_2\| - e^{2r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\cosh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_2\| \\
& + e^{2r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_1\| - e^{2r_1}\frac{\mathbf{x}_1^2}{\|\mathbf{x}_1\|^2}\sinh^2\|\mathbf{x}_1\| \\
& \quad + e^{r_1}e^{r_2}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_2\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_1\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_2\| \\
& + e^{r_1}e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_1\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_1\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_1\| \\
& \quad + e^{2r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_2\| - e^{2r_2}\frac{\mathbf{x}_2^2}{\|\mathbf{x}_2\|^2}\sinh^2\|\mathbf{x}_2\|
\end{aligned} \tag{174}$$

To simplify this expression, we note the following observations:

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

$$\cosh^2 x - \sinh^2 x = 1 \tag{175}$$

2. *we also note that:*

$$\frac{\mathbf{x}^2}{\|\mathbf{x}\|^2} = 1 \tag{176}$$

Proof. Let $\mathbf{x} = X_0\hat{\mathbf{x}}_0 + X_1\hat{\mathbf{x}}_1 + X_2\hat{\mathbf{x}}_2 + X_3\hat{\mathbf{x}}_3$. Then the geometric product is $\mathbf{x}^2 = X_0^2 - X_1^2 - X_2^2 - X_3^2$ which is the same as the square of the norm. □

We proceed with our simplifications as follows:

$$\begin{aligned}
&= e^{2r_1} + e^{2r_2} \\
&\quad - e^{2r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\cosh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_1\| + e^{r_1}e^{r_2}\cosh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_2\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\cosh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_2\| \\
&\quad + e^{r_2}e^{r_1}\cosh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_1\| - e^{r_2}e^{r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\cosh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_1\| - e^{2r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\cosh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_2\| \\
&\quad + e^{2r_1}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_1\| \\
&\qquad + e^{r_1}e^{r_2}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\sinh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_2\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_1\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_2\| \\
&\quad + e^{r_1}e^{r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_1\| - e^{r_1}e^{r_2}\frac{\mathbf{x}_1\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\sinh\|\mathbf{x}_1\| \\
&\qquad + e^{2r_2}\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_2\|\cosh\|\mathbf{x}_2\| && (177) \\
&= e^{2r_1} + e^{2r_2} + 2e^{r_1}e^{r_2}\cosh\|\mathbf{x}_1\|\cosh\|\mathbf{x}_2\| - 2e^{r_1}e^{r_2}\frac{\mathbf{x}_1\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\sinh\|\mathbf{x}_1\|\sinh\|\mathbf{x}_2\| && (178)
\end{aligned}$$

Here, the pseudo-inner-product (interval of special relativity) is recovered as the geometric product $\mathbf{x}_1\mathbf{x}_2$.

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

6.6 *The Universal Norm*

Definition 14 (Universal Norm)**.** *We take the norm of the geometric algebra $\mathcal{G}_4(\mathbb{R})$ to be a function $\|\cdot\| : \mathcal{G}_4(\mathbb{R}) \to \mathbb{R}$ with the requirement that it's output be the same as that of the determinant of its matrix representation (If we include the complex $\mathcal{G}_4(\mathbb{C})$, the norm remains the same but its domain is now $\|\cdot\| : \mathcal{G}_4(\mathbb{C}) \to \mathbb{C}$. It attributes no new geometry to the complexification of the pre-factors and they simply "pass-through" the norm.):*

$$\|\mathbf{u}\| := \sqrt[4]{\det M[\mathbf{u}]} \tag{179}$$

We can equivalently define this norm[46] *fully in the language of geometric algebra, as follows:*

[46] Douglas Lundholm and Lars Svensson. Clifford algebra, geometric algebra, and applications. *arXiv preprint arXiv:0907.5356*, 2009

$$\|\mathbf{u}\| = \sqrt[4]{\lfloor(\mathbf{u}^{\square})\mathbf{u}\rfloor_{\{3,4\}}(\mathbf{u}^{\square})\mathbf{u}} \tag{180}$$

Let us explain the notation. First, $\mathbf{u}^{\square}$ *is the geometric conjugate (also called the Clifford conjugate) of a multivector defined, in* $\mathcal{G}_4$*, as follows:*

$$\mathbf{u}^{\square} = \langle\mathbf{u}\rangle_0 - \langle\mathbf{u}\rangle_1 - \langle\mathbf{u}\rangle_2 + \langle\mathbf{u}\rangle_3 + \langle\mathbf{u}\rangle_4 \tag{181}$$

Furthermore, the notation $\lfloor(\mathbf{u}^{\square})\mathbf{u}\rfloor_{\{3,4\}}$ *represents the* $\{3,4\}$*-grade conjugate. It is defined as follows:*

$$\lfloor\mathbf{u}\rfloor_{\{3,4\}} = \langle\mathbf{u}\rangle_0 + \langle\mathbf{u}\rangle_1 + \langle\mathbf{u}\rangle_2 - \langle\mathbf{u}\rangle_3 - \langle\mathbf{u}\rangle_4 \tag{182}$$

For reference, the grades of $\mathbf{u}$ are:

$$\langle\mathbf{u}\rangle_0 = r \tag{183}$$

$$\langle\mathbf{u}\rangle_1 = (X_0)\hat{\mathbf{x}}_0 + (X_1)\hat{\mathbf{x}}_1 + (X_2)\hat{\mathbf{x}}_2 + (X_3)\hat{\mathbf{x}}_3 \tag{184}$$

$$\langle\mathbf{u}\rangle_2 = (E_1)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1 + (E_2)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_2 + (E_3)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_3 + (B_1)\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3 + (B_2)\hat{\mathbf{x}}_3\wedge\hat{\mathbf{x}}_1 + (B_3)\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2) \tag{185}$$

$$\langle\mathbf{u}\rangle_3 = (V_0)\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3 + (V_1)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3 + (V_2)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_3 + (V_3)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2) \tag{186}$$

$$\langle\mathbf{u}\rangle_4 = (\theta)\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3 \tag{187}$$

Now, let us write out the "half-product" of the norm: $(\mathbf{u}^{\square})\mathbf{u} \in \mathcal{G}_0 \oplus \mathcal{G}_3 \oplus \mathcal{G}_4$:

$$\begin{aligned}
(\mathbf{u}^{\square})\mathbf{u} = \\
& r^2 - \theta^2 + B_1^2 + B_2^2 + B_3^2 - E_1^2 - E_2^2 - E_3^2 + V_0^2 - V_1^2 - V_2^2 - V_3^2 - X_0^2 + X_1^2 + X_2^2 + X_3^2 \\
& + \hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2(2E_3V_0 + 2B_2V_1 - 2B_1V_2 + 2rV_3 - 2B_3X_0 + 2E_2X_1 - 2E_1X_2 - 2\theta X_3) \\
& + \hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_3(-2E_2V_0 + 2B_3V_1 - 2rV_2 - 2B_1V_3 + 2B_2X_0 + 2E_3X_1 + 2\theta X_2 - 2E_1X_3) \\
& + \hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3(2E_1V_0 + 2rV_1 + 2B_3V_2 - 2B_2V_3 - 2B_1X_0 - 2\theta X_1 + 2E_3X_2 - 2E_2X_3) \\
& + \hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3(2rV_0 + 2E_1V_1 + 2E_2V_2 + 2E_3V_3 - 2\theta X_0 - 2B_1X_1 - 2B_2X_2 - 2B_3X_3) \\
& + \hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3(2r\theta - 2B_1E_1 - 2B_2E_2 - 2B_3E_3 - 2V_0X_0 + 2V_1X_1 + 2V_2X_2 + 2V_3X_3)
\end{aligned} \tag{188}$$

If we now complete the full product we end up with the following norm applicable to a general multivector of $\mathcal{G}_4(\mathbb{C})$, which we call the universal norm:

$$
\begin{aligned}
&\left(\lfloor(\mathbf{u}^{\square})\mathbf{u}\rfloor_{\{3,4\}}(\mathbf{u}^{\square})\mathbf{u}\right)^4 = \\
&\quad (r^2-\theta^2+B_1^2+B_2^2+B_3^2-E_1^2-E_2^2-E_3^2+V_0^2-V_1^2-V_2^2-V_3^2-X_0^2+X_1^2+X_2^2+X_3^2)^2 \\
&\quad +(2E_3V_0+2B_2V_1-2B_1V_2+2rV_3-2B_3X_0+2E_2X_1-2E_1X_2-2\theta X_3)^2 \\
&\quad +(-2E_2V_0+2B_3V_1-2rV_2-2B_1V_3+2B_2X_0+2E_3X_1+2\theta X_2-2E_1X_3)^2 \\
&\quad +(2E_1V_0+2rV_1+2B_3V_2-2B_2V_3-2B_1X_0-2\theta X_1+2E_3X_2-2E_2X_3)^2 \\
&\quad -(2rV_0+2E_1V_1+2E_2V_2+2E_3V_3-2\theta X_0-2B_1X_1-2B_2X_2-2B_3X_3)^2 \\
&\quad +4(r\theta-B_1E_1-B_2E_2-B_3E_3-V_0X_0+V_1X_1+V_2X_2+V_3X_3)^2
\end{aligned}
\tag{189}
$$

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

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

$$\mathbf{v} := r \tag{190}$$

$$\implies \|\mathbf{v}\| = \sqrt[4]{\lfloor(r^{\square})r\rfloor_{\{3,4\}}(r^{\square})r} = r \tag{191}$$

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

$$\mathbf{v} := a + b\mathbf{I} \tag{192}$$

$$\implies \|\mathbf{v}\| = \sqrt{a^2+b^2} \tag{193}$$

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

$$\mathbf{v} := X_1\hat{\mathbf{x}}_1 + X_2\hat{\mathbf{x}}_2 + X_3\hat{\mathbf{x}}_3 \tag{194}$$

$$\implies \|\mathbf{v}\| = \sqrt{X_1^2+X_2^2+X_3^2} \tag{195}$$

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

$$\mathbf{v} := X_0\hat{\mathbf{x}}_0 + X_1\hat{\mathbf{x}}_1 + X_2\hat{\mathbf{x}}_2 + X_3\hat{\mathbf{x}}_3 \tag{196}$$

$$\implies \|\mathbf{v}\| = \sqrt{-X_0^2+X_1^2+X_2^2+X_3^2} \tag{197}$$

Example 5 (Universal norm applied to a 2-vector (Faraday tensor))**.**

$$\mathbf{v} := E_1\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_1 + E_2\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2 + E_3\hat{\mathbf{x}}_0\wedge\hat{\mathbf{x}}_3 + B_1\hat{\mathbf{x}}_2\wedge\hat{\mathbf{x}}_3 + B_2\hat{\mathbf{x}}_3\wedge\hat{\mathbf{x}}_1 + B_3\hat{\mathbf{x}}_1\wedge\hat{\mathbf{x}}_2 \tag{198}$$

$$\implies \|\mathbf{v}\| = \sqrt[4]{(E_1^2+E_2^2+E_3^2-B_1^2-B_2^2-B_3^2)^2+4(E_1B_1+E_2B_2+E_3B_3)^2} \tag{199}$$

We note that the quantities $E_1^2+E_2^2+E_3^2-B_1^2-B_2^2-B_3^2$ and $E_1B_1+E_2B_2+E_3B_3$ 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.

6.7 Geometric interference (Falsifiable)

The language of geometric algebra allows us to simplify the two-state universal interference pattern substantially and reveals considerable insight. We start with two sub-wavefuntions:

$$\psi_1 = e^{\frac{1}{2}r_1} \exp\left(\frac{1}{2}(\theta_1 \mathbf{I} + \mathbf{X}_1 + \mathbf{F}_1 + \mathbf{V}_1)\right) = e^{\frac{1}{2}r_1}\mathbf{G}_1 \tag{200}$$

$$\psi_2 = e^{\frac{1}{2}r_2} \exp\left(\frac{1}{2}(\theta_2 \mathbf{I} + \mathbf{X}_2 + \mathbf{F}_2 + \mathbf{V}_2)\right) = e^{\frac{1}{2}r_2}\mathbf{G}_2 \tag{201}$$

and we define a wavefunction ψ as the sum of the two:

$$\psi = \psi_1 + \psi_2 \tag{202}$$

We can the inject ψ into the definition of the universal norm:

$$\begin{aligned}
N[\psi] &= \lfloor \psi^{\square}\psi \rfloor_{3,4} \psi^{\square}\psi & (203)\\
&= \lfloor (\psi_1+\psi_2)^{\square}(\psi_1+\psi_2) \rfloor_{3,4} (\psi_1+\psi_2)^{\square}(\psi_1+\psi_2) & (204)\\
&= \lfloor (\psi_1^{\square}+\psi_2^{\square})(\psi_1+\psi_2) \rfloor_{3,4} (\psi_1^{\square}+\psi_2^{\square})(\psi_1+\psi_2) & (205)\\
&= \lfloor \psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2 \rfloor_{3,4} (\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2) & (206)\\
&= (\lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4} + \lfloor \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 \rfloor_{3,4} + \lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4})(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2) & (207)\\
&= \lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2)\\
&\quad + \lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4}(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2)\\
&\quad + \lfloor \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2) & (208)\\
&= \lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1) + \lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4}(\psi_2^{\square}\psi_2)\\
&\quad + \lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2)\\
&\quad + \lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4}(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1)\\
&\quad + \lfloor \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1 + \psi_1^{\square}\psi_2 + \psi_2^{\square}\psi_1 + \psi_2^{\square}\psi_2) & (209)
\end{aligned}$$

Since $\lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1)$ and $\lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4}(\psi_2^{\square}\psi_2)$ are simply the determinant of the matrix representation of the multivector, they reduce to:

$$\lfloor \psi_1^{\square}\psi_1 \rfloor_{3,4}(\psi_1^{\square}\psi_1) = (e^{r_1})^2 \tag{210}$$

$$\lfloor \psi_2^{\square}\psi_2 \rfloor_{3,4}(\psi_2^{\square}\psi_2) = (e^{r_2})^2 \tag{211}$$

We get:

$$\begin{aligned} N[\psi] = {} & (e^{r_1})^2 + (e^{r_2})^2 \\ & + \lfloor \psi_1^{\square} \psi_1 \rfloor_{3,4} (\psi_1^{\square} \psi_2 + \psi_2^{\square} \psi_1 + \psi_2^{\square} \psi_2) \\ & + \lfloor \psi_2^{\square} \psi_2 \rfloor_{3,4} (\psi_1^{\square} \psi_1 + \psi_1^{\square} \psi_2 + \psi_2^{\square} \psi_1) \\ & + \lfloor \psi_1^{\square} \psi_2 + \psi_2^{\square} \psi_1 \rfloor_{3,4} (\psi_1^{\square} \psi_1 + \psi_1^{\square} \psi_2 + \psi_2^{\square} \psi_1 + \psi_2^{\square} \psi_2) \end{aligned} \tag{212}$$

Thus, in the general case, an interference pattern, far exceeding the complex interference pattern producible by ordinary quantum mechanics, is predicted by this model.

I note that this paper[47] has also predicted an extended interference pattern in the geometric formulation of the wavefunction, as an even algebra subset of the above and suggesting a variation of the Aharonov-Bohm Effect experiment as a possibly way to falsify these predictions,

[47] I Bohdan et al. Wave function as geometric entity. *Journal of Modern Physics*, 2012, 2012

6.8 Extended Path integral

The Feynman path integral is:

$$K = \int Dx[t] \exp\left(i \int dt L[x, \dot{x}, t]\right) = \int Dx[t] \exp\big(iS[x[t]]\big) \tag{213}$$

and the associated probability distribution is obtained by application of the Born rule:

$$P[x] = \left| \int Dx[t] \exp\left(i \int dt L[x, \dot{x}, t]\right) \right|^2 \tag{214}$$

Theorem 10 (Extended path integral)**.** *We will now derive the extended path integral using the formalism of universal statistical physics.*

Proof. We take a complete multivector of $\mathcal{G}_4(\mathbb{R})$:

$$\mathbf{g}[p] = e^{\frac{1}{2} r[p]} \exp\left(\frac{1}{2}(\theta[p]\mathbf{I} + \mathbf{X}[p] + \mathbf{F}[p] + \mathbf{V}[p])\right) \tag{215}$$

Then we define a eigen-partition function for each eigenvalues of the matrix representation of $\mathbf{g}[p]$, which we note as the functions $\lambda_a[p], \lambda_b[p], \lambda_c[p], \lambda_d[p]$. Each lambda function is a map from $\mathbb{P}$, a countable set, to a continuous but countable set, such as the rationals or preferably the computable reals. In the infinite extension of the case shown in Figure 4, the probability amplitude of B given A, and for each eigenvalues, is given as follows:

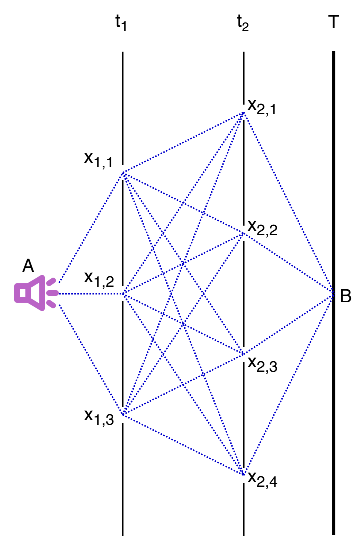

Figure 4: The Feynman path integral is obtained when the quantities m and n of slices $(t_1, t_2, \ldots, t_m)$ and $(x_{i,1}, x_{i,2}, \ldots, x_{i,n})$ are increased to infinity

$$K_a = \sum_{p_1 \in \mathbb{P}_1} \sum_{p_2 \in \mathbb{P}_2} \cdots \sum_{p_n \in \mathbb{P}_n} e^{-\tilde{k}\lambda_a[p_1]} e^{-\tilde{k}\lambda_a[p_2]} \ldots e^{-\tilde{k}\lambda_a[p_n]} \tag{216}$$

$$K_b = \sum_{p_1 \in \mathbb{P}_1} \sum_{p_2 \in \mathbb{P}_2} \cdots \sum_{p_n \in \mathbb{P}_n} e^{-\tilde{k}\lambda_b[p_1]} e^{-\tilde{k}\lambda_b[p_2]} \ldots e^{-\tilde{k}\lambda_b[p_n]} \tag{217}$$

$$K_c = \sum_{p_1 \in \mathbb{P}_1} \sum_{p_2 \in \mathbb{P}_2} \cdots \sum_{p_n \in \mathbb{P}_n} e^{-\tilde{k}\lambda_c[p_1]} e^{-\tilde{k}\lambda_c[p_2]} \ldots e^{-\tilde{k}\lambda_c[p_n]} \tag{218}$$

$$K_d = \sum_{p_1 \in \mathbb{P}_1} \sum_{p_2 \in \mathbb{P}_2} \cdots \sum_{p_n \in \mathbb{P}_n} e^{-\tilde{k}\lambda_d[p_1]} e^{-\tilde{k}\lambda_d[p_2]} \ldots e^{-\tilde{k}\lambda_d[p_n]} \tag{219}$$

where $\tilde{k}$ is the Lagrange multiplier. Then the probability is given by:

$$P = K_a K_b K_c K_d \tag{220}$$

This is a version of the Path integral which does not rely under an assumption of space-time. One can perhaps understand it as a "path integral" in a purely logical space. We can derive the path integral in the continuous form we are familiar with, but to do so we have to create a map $x : \mathbb{P} \to \mathbb{R}$ that attributes a spatial coordinates to each experiment. Since, $\mathbb{P}$ is countably infinite, we can map to a continuous countable subset of the reals such as the rationals, or preferably the computable reals. One then performs the convolution of $x[p]$ with each of the eigenvalue functions, and get these replacements:

$$\lambda_a[p] \to \lambda_a[x[p]] \tag{221}$$

$$\lambda_b[p] \to \lambda_b[x[p]] \tag{222}$$

$$\lambda_c[p] \to \lambda_c[x[p]] \tag{223}$$

$$\lambda_d[p] \to \lambda_d[x[p]] \tag{224}$$

The lambdas are now functions from the reals to the real, and represent a density evaluated at a point $x[p]$. Finally, making the dependence on p implicit, one can then define the path integral with the familiar notation:

$$K_a = \int D[\lambda_a[x]] e^{-\tilde{k} \int \lambda_a[x] dx} \tag{225}$$

$$K_b = \int D[\lambda_b[x]] e^{-\tilde{k} \int \lambda_b[x] dx} \tag{226}$$

$$K_c = \int D[\lambda_c[x]] e^{-\tilde{k} \int \lambda_c[x] dx} \tag{227}$$

$$K_d = \int D[\lambda_d[x]] e^{-\tilde{k} \int \lambda_d[x] dx} \tag{228}$$

The probability is this case is still $P = K_a K_b K_c K_d$ □

We note that in the case of the path integral, the limitation of 4 space-time dimensions produced by the no-go theorem takes all of its sense. The laws of physics normally associated to the Lagrangian here takes the form of an eigenvalue. Since the roots of polynomials of degree 5 of higher have no solutions in radicals, and their expression can vary based on the numeral value of the entries of the matrix, then the loss of structure prevents any invariance within the Lagrangian, for the general case of the extended Born rule.

6.9 Quartic Hilbert spaces

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

$$\|\cdot\| : \exp\big(\mathcal{G}_4(\mathbb{R})\big) \to \mathbb{R} \tag{229}$$

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 $\psi[\mathbf{x}]$ is:

$$\int_{[D]} \lfloor \psi^{\square}\psi \rfloor_{3,4} \psi^{\square}\psi \,\mathrm{d}\mathbf{x} < \infty \tag{230}$$

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

$$\int_{[D]} \det \exp M[\mathbf{u}[\mathbf{x}]] \,\mathrm{d}\mathbf{x} < \infty \tag{231}$$

6.10 Octic Hilbert spaces

If we now want a probability norm with map:

$$\|\cdot\| : \exp\big(\mathcal{G}_4(\mathbb{C})\big) \to \mathbb{R} \tag{232}$$

such that exponential map over the complexification of the space-time algebra, representing the general liner group of transformations, the $\mathrm{GL}(4,\mathbb{C})$ group, is obtained by taking the square modulos of the universal norm:

$$\|\psi\| = \left(\lfloor \psi^{\square}\psi \rfloor_{3,4} \psi^{\square}\psi\right)\left(\lfloor \psi^{\square}\psi \rfloor_{3,4} \psi^{\square}\psi\right)^{*} \tag{233}$$

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 $\mathcal{G}_6(\mathbb{C}) \cong \mathbb{M}(8, \mathbb{C})$. 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.

7 *Discussion*

Now that we have derived the (extended) Born rule from first principle, can we use this insight to solve both the measurement problem, and the interpretation of quantum mechanics problem? Was deriving the mathematical origin of said rule the missing ingredient? We note that we also have a mathematical description of the observer, and a definition of reality. These are new tools, not previously available, and long suspected to be key to uncover this mystery.

7.1 *Click-first, wavefunction-second interpretation*

Universal statistical physics on the one hand inherits its interpretation from ordinary statistical physics, and on the other it is a superset of quantum mechanics, therefore it bequeaths an interpretation to quantum physics consistent with statistical physics. I note that my framework shares very strong similarities to the statistical ensemble interpretation[48], but it is actually more than just an interpretation, because it derives the wavefunction, rather than simply postulating it as is the case with interpretations. It is therefore more accurate to state that it is an explanation of quantum mechanics, rather than an interpretation.

[48] Leslie E Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970

In ordinary statistical physics, one assumes the existence of measuring instruments, such as a thermometer or a barometer. Such instruments are used to gather finitely many individual measurements. One assumes that, in the limit, these measurement converges to an average value, called the constraint. Finally, under the principle of maximum entropy and under said constraints, one recovers the Gibbs measure.

In the case of quantum mechanics, a similar interpretation is found in universal statistical physics, but instead of measuring temperature or pressure, our detector registers clicks. Clicks have more structure than simple scalar quantities, but otherwise, have the same assumptions as those of ordinary statistical physics. This extra structure, along with a click-first wavefunction-second foundation, allows us to define all problematic quantum effects in a remarkably nonproblematic manner. This even includes the problematic quantum measurement problem:

For instance, a detector may measure a given electron's spin as: up, up, up, up, up, up, ... If we then consider these clicks to be the foundation of an statistical ensemble, then said ensemble is comprised of a single microstate. Maximizing the entropy of such a system using ordinary statistical physics produces a Gibbs measure with 100% probability of this single state, whereas doing the same with universal statistical physics will produce a collapsed wavefunction.

To obtain a plurality of microstates rather than just one, one will have to measure a plurality of similarly prepared electrons, such that the detector reports multiple measurement values, and finally, to maximize the entropy of such an ensemble in order to get a probability measure over multiple states. Like for ordinary statistical physics, we consider that measuring many clicks obtained under copies of a similarly prepared systems will converge towards a fixed finite value which we will call a constraint. Remarkably, the probability measure resulting from maximizing the entropy under the constraint of clicks is the relativistic wavefunction in lieu of the Gibbs measure. We restate for emphasis: The relativistic wavefunction is —simply—- the generalization of the Gibbs measure to clicks. From the click-first wavefunction-second direction, the gamut of 'unintuitive' quantum behaviour is transposed to the empirically-undeniable set of all possible sequences of clicks. Each sequence of clicks is responsible for defining an ensemble for which it is possible to derive a wavefunction of the appropriate structure. Clicks that repeat the same value implies a probability measure describing a collapsed wavefunction, and clicks that produce different random values implies a probability measure describing a multi-state non-collapsed wavefunction.

The interpretation of quantum mechanics continues within this setup by referencing the concepts of microstates and macrostates and how they relate. In ordinary statistical physics, for instance in an ideal gas, each possible distribution of air molecules is a microstate. The observer is assumed to rest somewhere out of sight and to take notes of the (macroscopic) temperature and pressure measurements perhaps using pen and paper. However, in universal statistical physics, the terms microstate and macrostate are somewhat misleading because we are not necessarily dealing with difference in sizes as we typically do in ordinary statistical physics.

In the case of universal statistical physics, the observer <u>is</u> aware of the result of a measurement and thus necessarily constitutes the microscopic description of the system, whereas the macroscopic description is defined as a particular sequence of clicks. I note that this somewhat reminds me of Maxwell's daemon allegedly aware the position of the molecule of the gas. The measurement-problem 'industrial complex' rests upon the assumption that the wavefunction, rather than the click, is the most fundamental physical object, then is baffled as to how clicks are caused. However, universal statistical physics reveals that this assumption is incorrect. Indeed, it is the wavefunction that is derived from (maximizing the entropy of) a given sequence of clicks, and thus is the least fundamental object of the two. The clicks, by themselves, defines and constrains the physical reality of the system.

7.2 *A brief note on the choice of microstate*

In statistical physics, the formalism is generally independent of the choice of microscopic element. A statistical ensemble can represent molecules, polymer chains, even black hole entropy, etc. In some cases, a given ensemble can accommodate multiple candidate for its microscopic description. For instance we believe that black holes have an entropy, but what does this entropy represent is up to debate — does it represent internal degrees of freedom, is it a representation of the entanglement between inside states and outside states, or is it any of numerous other suggestions?

So why pick manifests as microscopic states, if other choices could be conceivable — why not pick, say, a "typical" particle-wave dual construct? The short answer is that the manifest is the simplest construction which meets the fundamental assumption of science, and thankfully achieves exactly nothing more. Indeed, using the manifest as the microscopic description is the minimal and less restrictive description which restricts the wavefunction to be a computable entity, while also importing the desirable indubitable foundation of reality, the definition of the observer and guarantees, by virtue of being indubitable, that no undue physical or logical baggage is inadvertently injected into the framework. It is therefore the safest possible choice which grabs all the desirables.

Specifically, I will note that the definition of universal statistical physics is conceptually similar (to not say 'exactly the same') as that of quantum computing[49]. This is quite interesting and I would argue very revealing. The premise of quantum computing is to produce a series of unitary transformations such that sufficient expressive power is recovered to be Turing complete. Then, repeat measurements over copies or re-runs of the computation allows one to build a statistical average used to establish the result of the computation. Quantum computer are usually described using non-relativistic quantum mechanics for practical and applicability reason, but the concepts must hold in the relativistic regime also since it is a superset

[49] Richard J Lipton and Kenneth W Regan. Quantum algorithms via linear algebra

of the non-relativistic regime. Universal statistical physics, although it goes straight to the relativistic regime, can easily be seen along the same line, but formulated in the reverse: we start with a sequence of clicks, then derive the probability measure, as a unitary evolving wavefunction, under the additional requirement that it represents the mathematical work used to verify the reference manifest, itself a set of programs constructed from a Turing complete set of statements. The requirement that experiments, defined as programs, be verified under the constraint of nature, produces within the framework of universal statistical physics a unitary evolving wavefunction capable of general quantum computation, such that the reference manifest remains verified throughout the evolution of the system. Furthermore, since the elements of a manifest are computable, one can predict any experiment by pre-calculation and then compare it to a real experiment later; thus making the whole of reality subject to falsification. Without using manifest as the microstates, a subset of physics would be non-falsifiable by virtue of been non-computable, and thus the fundamental assumption of science would fail. Attributing the manifest as the microstate of the ensemble means that we meet the definition of quantum computing (in the relativistic regime, and with the Born rule extensions), but that we come at it from the other side.

7.3 *Making reality maximally informative*

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 $\mathring{\mathbb{M}}$. 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 to the observer by satisfying the requirement of the Shannon definition of information regarding randomly selecting an element from a larger set. Nonetheless, the laws of physics do require this process lest they cannot be derived as a predictive theory. This is why the laws of physics are a theorem of science applied to reality (and 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).

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. If we restrict the domain of a theory to that of reality, then we obtain a set of Turing machine — which we call 'manifest theories'. 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 a magician". How credible is that claim? Supposedly, we may admit that being a magician violates the laws of physics, whereas being a doctor doesn't. Do we then want our laws of physics to rule out the magician, but not the doctor, even though in reality we got the physicist? A manifest theory rules out the doctor, but remarkably physics doesn't. 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. If manifest theories (whose domain is that of reality) are 'mathematically-ideal' scientific theories, then physics (whose domain is science and its subject matter is the random selection of reality from said domain) is revealed to be a meta-scientific theory.

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 are unverified.

7.4 *Why does the universe exist?*

In the present formalism, we make the distinction between reality and the universe. Reality, defined as the reference manifest, is the axiomatic foundation from which the existence of the universe can be derived from. Specifically, the universe is derivable, but reality isn't. Reality is comprised exclusively of indubitable statements and as such it comprises an indubitable 'mathematical rock bottom'. Consequently, building a more fundamental basis for it is impossible. It must be accepted as-is. The follow-up question: why do we exist as beings capable of doing this verification... the framework cannot answer. You were born, you opened your eyes, both you and reality were there — its the given. Then, by inspecting reality, you were able over time to develop the concepts of object persistence, and further an understanding that objects are distributed in space and so on, as patterns that stands out from your experience of reality. Eventually, you reached the limit of your evolutionary-produced senses, but you were able to complement them with precision instruments which revealed new microscopic patterns. But, why is it these patterns that stand out - what's special about them?

The whole point of my framework is to formalise this "experience" to the highest possible degree, such that physics stands out as the ultimate pattern. First, let me note that the set of all indubitable statements is mathematically universal, consequently it is necessarily sufficiently expressive to describe reality to any level of detail required and is on par with any other Turing complete theory or language, which is the highest level of completeness known to mathematics. Then the key is that we can claim, by referencing the syllogisms of this manuscript, that the mathematical work required to verify reality also exists indubitably. Essentially, the indubitable property of the elements of a manifest cascades to everything it is contingent upon. We find, by the syllogisms, that reality is contingent on computing work.

Physics is a meta-scientific theory that survives any transformations or rearrangements of the reference manifest constrained by the contingent mathematical work which we call nature. Contrary to a manifest theory which enumerates a subset of the reference manifest, physics assigns a probability measure to each manifest. Using universal statistical physics, we are able to the derive relativistic wavefunction as the analog to the Gibbs measure, a process self-limited to four dimensions. We discover that re-arrangement of the reference manifest, such that the computational work requirements are changed, corresponds to a sequence of click, and that the computational work is performed by unitary evolution akin to quantum computing. It turns out that this probability measure is simply the one with the richest structure allowed by universal statistical physics and this is the fundamental structure, along with its computational entailment with the reference manifest, of what we perceive as the material universe. The universe is the way it is simply because this happens to be the richest informational structure one can formulate — it is an informational maximum. Consequently, those whose goal it is to make reality maximally informative ought to eventually converge towards such a description.

References

[1] Roger Penrose. *The road to reality: A complete guide to the laws of the universe*. Random house, 2006.

[2] D Budgen. Software design . harlow, uk, 2003.

[3] Scott W Ambler. *Process patterns: building large-scale systems using object technology*. Cambridge university press, 1998.

[4] Seth Lloyd. Computational capacity of the universe. *Physical Review Letters*, 88(23):237901, 2002.

[5] Alan M Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1):230–265, 1937.

[6] Gregory J Chaitin. Meta math! the quest for omega. *arXiv preprint math/0404335*, 2004.

[7] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975.

[8] John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012.

[9] Gert-Jan Lokhorst. Descartes and the pineal gland. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.

[10] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961.

[11] Rolf Landauer et al. Information is physical. *Physics Today*, 44(5):23–29, 1991.

[12] John Archibald Wheeler. Frontiers of time. In *Problems in the Foundations of Physics*. North-Holland for the Societa italiana di fisica, 1978.

[13] John A Wheeler. Information, physics, quantum: The search for links. *Complexity, entropy, and the physics of information*, 8, 1990.

[14] N Aghanim, Y Akrami, M Ashdown, J Aumont, C Baccigalupi, M Ballardini, AJ Banday, RB Barreiro, N Bartolo, S Basak, et al. Planck 2018 results. vi. cosmological parameters. *arXiv preprint arXiv:1807.06209*, 2018.

[15] Gregory Chaitin. Doing mathematics differently. `https://inference-review.com/article/doing-mathematics-differently`, 02 2019. Accessed: 2019-12-04.

[16] Gregory Chaitin. *Proving Darwin: making biology mathematical*. Vintage, 2012.

[17] Jos Uffink. Compendium of the foundations of classical statistical physics. *Philosophy of Physics*, 03 2006.

[18] Victor S. Batista. Postulates of statistical mechanics. `http://xbeams.chem.yale.edu/~batista/vaa/node20.html`. Accessed: 2019-12-17.

[19] C. H. Bennett, P. Gacs, Ming Li, P. M. B. Vitanyi, and W. H. Zurek. Information distance. *IEEE Transactions on Information Theory*, 44(4):1407–1423, July 1998.

[20] Edward Fredkin and Tommaso Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3):219–253, Apr 1982.

[21] Andrei Nikolaevich Kolmogorov. Three approaches to the definition of the concept "quantity of information". *Problemy peredachi informatsii*, 1(1):3–11, 1965.

[22] A K Zvonkin and L A Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83, 1970.

[23] R.J. Solomonoff. A formal theory of inductive inference. part i. *Information and Control*, 7(1):1 – 22, 1964.

[24] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science*, 9(4):301–310, 1964.

[25] Kohtaro Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002.

[26] Kohtaro Tadaki. A statistical mechanical interpretation of algorithmic information theory. In *Local Proceedings of the Computability in Europe 2008 (CiE 2008)*, pages 425–434. University of Athens, Greece, Jun 2008.

[27] Ming Li and Paul M.B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008.

[28] Cristian S. Calude and Michael A. Stay. Natural halting probabilities, partial randomness, and zeta functions. *Inf. Comput.*, 204(11):1718–1739, November 2006.

[29] L. G. Kraft. A device for quanitizing, grouping and coding amplitude modulated pulses. Master's thesis, Mater's Thesis, Department of Electrical Engineering, MIT, Cambridge, MA, 1949.

[30] Cristian S Calude, Michael J Dinneen, Chi-Kou Shu, et al. Computing a glimpse of randomness. *Experimental Mathematics*, 11(3):361–370, 2002.

[31] Stephen W Hawking. Black hole explosions? *Nature*, 248(5443):30, 1974.

[32] Damien A. Easson, Paul H. Frampton, and George F. Smoot. Entropic inflation. *International Journal of Modern Physics A*, 27(12):1250066, 2012.

[33] Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, 1995.

[34] Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333, 1973.

[35] Jacob D Bekenstein. Generalized second law of thermodynamics in black-hole physics. *Physical Review D*, 9(12):3292, 1974.

[36] Jacob D Bekenstein. Black-hole thermodynamics. *Physics Today*, 33(1):24–31, 1980.

[37] Ted Jacobson. Thermodynamics of spacetime: The einstein equation of state. *Phys. Rev. Lett.*, 75:1260–1263, Aug 1995.

[38] Erik P. Verlinde. On the origin of gravity and the laws of newton. *Journal of High Energy Physics*, 2011(4):29, Apr 2011.

[39] Matt Visser. Conservative entropic forces. *Journal of High Energy Physics*, 2011(10):140, 2011.

[40] Lee Smolin. Newtonian gravity in loop quantum gravity. *arXiv preprint arXiv:1001.3668*, 2010.

[41] Tower Wang. Coulomb force as an entropic force. *Physical Review D*, 81(10):104045, 2010.

[42] Peter GO Freund. Emergent gauge fields. *arXiv preprint arXiv:1008.4147*, 2010.

[43] Rong-Gen Cai, Li-Ming Cao, and Nobuyoshi Ohta. Friedmann equations from entropic force. *Physical Review D*, 81(6):061501, 2010.

[44] Miao Li and Yi Wang. Quantum uv/ir relations and holographic dark energy from entropic force. *Physics Letters B*, 687(2-3):243–247, 2010.

[45] Damien A Easson, Paul H Frampton, and George F Smoot. Entropic accelerating universe. *Physics Letters B*, 696(3):273–277, 2011.

[46] Sabine Hossenfelder. Comments on and comments on comments on verlinde's paper" on the origin of gravity and the laws of newton". *arXiv preprint arXiv:1003.1015*, 2010.

[47] Archil Kobakhidze. Gravity is not an entropic force. *Physical Review D*, 83(2):021502, 2011.

[48] Shan Gao. Is gravity an entropic force? *Entropy*, 13(5):936–948, 2011.

[49] BL Hu. Gravity and nonequilibrium thermodynamics of classical matter. *International Journal of Modern Physics D*, 20(05):697–716, 2011.

[50] Archil Kobakhidze. Once more: gravity is not an entropic force. *arXiv preprint arXiv:1108.4161*, 2011.

[51] Brian Swingle. Entanglement renormalization and holography. *Physical Review D*, 86(6):065007, 2012.

[52] Mark Van Raamsdonk. Building up spacetime with quantum entanglement. *General Relativity and Gravitation*, 42(10):2323–2329, 2010.

[53] Glen Evenbly and Guifré Vidal. Tensor network states and geometry. *Journal of Statistical Physics*, 145(4):891–918, 2011.

[54] Thomas Faulkner, Monica Guica, Thomas Hartman, Robert C Myers, and Mark Van Raamsdonk. Gravitation from entanglement in holographic cfts. *Journal of High Energy Physics*, 2014(3):51, 2014.

[55] Thomas Faulkner, Felix M Haehl, Eliot Hijano, Onkar Parrikar, Charles Rabideau, and Mark Van Raamsdonk. Nonlinear gravity from entanglement in conformal field theories. *Journal of High Energy Physics*, 2017(8):57, 2017.

[56] Bartlomiej Czech, Lampros Lamprou, Samuel McCandlish, and James Sully. Tensor networks from kinematic space. *Journal of High Energy Physics*, 2016(7):100, 2016.

[57] Juan Maldacena. The large-n limit of superconformal field theories and supergravity. *International journal of theoretical physics*, 38(4):1113–1133, 1999.

[58] Brian Swingle and Mark Van Raamsdonk. Universality of gravity from entanglement. *arXiv preprint arXiv:1405.2933*, 2014.

[59] Juan Maldacena and Leonard Susskind. Cool horizons for entangled black holes. *Fortschritte der Physik*, 61(9):781–811, 2013.

[60] Fabio Sanches and Sean J Weinberg. Holographic entanglement entropy conjecture for general spacetimes. *Physical Review D*, 94(8):084034, 2016.

[61] Leonard Susskind. Entanglement is not enough. *Fortschritte der Physik*, 64(1):49–71, 2016.

[62] ChunJun Cao, Sean M Carroll, and Spyridon Michalakis. Space from hilbert space: recovering geometry from bulk entanglement. *Physical Review D*, 95(2):024031, 2017.

[63] Ning Bao, ChunJun Cao, Sean M Carroll, and Liam McAllister. Quantum circuit cosmology: The expansion of the universe since the first qubit. *arXiv preprint arXiv:1702.06959*, 2017.

[64] Ning Bao, ChunJun Cao, Sean M Carroll, and Aidan Chatwin-Davies. de sitter space as a tensor network: Cosmic no-hair, complementarity, and complexity. *Physical Review D*, 96(12):123536, 2017.

[65] Leonard Susskind. Dear qubitzers, gr= qm. *arXiv preprint arXiv:1708.03040*, 2017.

[66] Thanu Padmanabhan. Thermodynamical aspects of gravity: new insights. *Reports on Progress in Physics*, 73(4):046901, 2010.

[67] Ted Jacobson. Entanglement equilibrium and the einstein equation. *Physical review letters*, 116(20):201101, 2016.

[68] Sean M Carroll and Grant N Remmen. What is the entropy in entropic gravity? *Physical Review D*, 93(12):124052, 2016.

[69] David Hestenes. Spacetime physics with geometric algebra. *American Journal of Physics*, 71(7):691–714, 2003.

[70] Douglas Lundholm and Lars Svensson. Clifford algebra, geometric algebra, and applications. *arXiv preprint arXiv:0907.5356*, 2009.

[71] I Bohdan et al. Wave function as geometric entity. *Journal of Modern Physics*, 2012, 2012.

[72] Leslie E Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970.

[73] Richard J Lipton and Kenneth W Regan. Quantum algorithms via linear algebra.