On an entropic universal Turing machine isomorphic to physics

Alexandre Harvey-Tremblay

aht@protonmail.ch

Published: August 13, 2017

According to the second law of thermodynamics, a physical system will tend to increase its entropy over time. In this paper, I investigate a universal Turing machine (UTM) running multiple programs in parallel according to a scheduler. I found that if, over the course of the computation, the scheduler adjusts the work done on programs so as to maximize the entropy in the calculation of the halting probability Ω, the system will follow the laws of physics. Specifically, I show that the computation will obey algorithmic information theory (AIT) analogues to general relativity, entropic dark energy, the Schrödinger equation, a maximum computation speed analogous to the speed of light, the Lorentz's transformation, light cone, the Dirac equation for relativistic quantum mechanics, spins, polarization, etc. As the universe follows the second law of thermodynamics, these results would seem to suggest an affinity between an "entropic UTM" and the laws of physics.

Contents

1 *Universal Turing Machine*

A universal Turing machine (UTM) is a Turing machine (TM) that has the property that can correctly simulate every TM for every program x. Said differently, for any UTM this must hold;

$$\forall x \forall TM[UTM(\langle TM \rangle . \langle x \rangle) = \langle TM(x) \rangle] \quad (1.1)$$

The bracket notation $\langle TM \rangle$, $\langle x \rangle$ and $\langle TM(x) \rangle$ simply indicates that the input is encoded in the language of the UTM and the period symbol "." indicates the concatenation. As each TM outputs a list of symbols, the equal sign means that for any program x the output of the UTM must equal the encoded output of the TM.

We introduce a prefix-free encoding to avoid the possibility of the UTM confusing two programs. For example, such confusion can occur with the program 110 and program 1101. If the UTM were to halt immediately after reading the first zero, we could not trace which of these two programs was its actual input. A popular and simple prefix-free encoding is the unitary encoding. It is the following set of programs,

$$\text{Unitary Encoding} := \{0, 10, 110, 1110, \dots\} \quad (1.2)$$

By limiting ourselves to a maximum of a single zero per encoded program at the very end, we guarantee that no program is the prefix of another.

Gregory Chaitin proved that any prefix-free UTM can be associated with an halting probability Ω.

Definition 1.3. *Ω is the halting probability of a prefix free UTM*[1]. *Ω is a normal, non-computable, algorithmically random and transcendental real number.*

For unitary encoded programs, Ω is obtained via the following sum,

$$\Omega = \sum_{x=1}^{\infty} 2^{-E(x)-x} \quad (1.4)$$

where x is a program and where $E(x)$ is the halting-event function and is defined as:

$$E(x) = \begin{cases} 0 & x \text{ halts} \\ \infty & \text{otherwise} \end{cases} \quad (1.5)$$

[1] Gregory J. Chaitin. An algebraic equation for the halting probability. https://www.cs.auckland.ac.nz/~chaitin/berlin.pdf, 1988; Gregory J. Chaitin. How real are real numbers? http://www.worldscientific.com/doi/abs/10.1142/S0218127406015726 and https://www.cs.auckland.ac.nz/~chaitin/olympia.pdf, 2006; Gregory Chaitin. Mathematics, complexity and philosophy. https://www.academia.edu/31320410/Mathematics_Complexity_and_Philosophy_full_bilingual_text_, 2010; and Ming Li and Paul Vitányi. An introduction to kolmogorov complexity and its applications. Springer, 1997

Note that as $E(x)$ is the bearer of non-halting information, it is connected to the halting problem of computer science. As a result, it is a non-computable function.

Expanding the sum into binary and using example values for $E(x)$, we get

$$\Omega = \sum_{x=1}^{\infty} 2^{-E(x)} 2^{-x} \tag{1.6}$$
$$= 2^{-\infty}2^{-1} + 2^{-0}2^{-2} + 2^{-0}2^{-3} + 2^{-0}2^{-4} + 2^{-\infty}2^{-5} + ... \tag{1.7}$$
$$= 0_b + 0.01_b + 0.001_b + 0.0001_b + 0_b + ... \tag{1.8}$$
$$= 0.01110..._b \tag{1.9}$$

We obtain a number Ω where its bits are in a one-to-one correspondence with the unitary encoded programs running of the UTM. If the x^{th} program halts, then the x^{th} bit of Ω is 1, otherwise it is 0. The reason why $E(x)$ uses 0 and ∞ as its two states is to remove the terms associated with non-halting programs from the sum by making them vanish to 0.

Knowing the bits of Ω is enough to determine if any program of the UTM will halt or not. As knowing Ω would solve the halting problem and that such is unsolvable in the general case, Ω is necessarily non-computable.

1.1 Entropic UTM

The entropy is defined as

Gibb's entropy	$S = -k_b \sum p_i \ln p_i$	(1.10)
Shannon's entropy	$S = -\sum p_i \log_2 p_i$	(1.11)

What is the entropy of Ω? We consider the case of an unspecified UTM. In this case, Ω can be represented as a concatenation of bits such that;

$$\Omega = 0.\omega_1\omega_2\omega_3\omega_4 \dots \tag{1.12}$$

where every ω_i is either 0 or 1, but as the UTM is unspecified its actual value is unknown to us. There are infinitely many such bits and as such, we can define a non-divergent entropy for at most $N \in \mathbb{N}$ bits of Ω. Furthermore, as Ω is a normal random real number, p_i is $1/2$ for each ω_i. Therefore, the entropy for N bits of Ω is;

$$S = k_b N \ln 2 \tag{1.13}$$

This result applies to any UTM. As a set, it applies each element of Λ ;

$$\Lambda := \{U | isUTM(U)\} \quad (1.14)$$

where $isUTM(U)$ is a function returning true if U is a UTM and false otherwise. Of course *isUTM* is a non-computable function as it is not possible to prove that an arbitrary TM is a UTM. Nonetheless, defining Λ will prove to be useful in future sections.

1.2 *Why work with the set Λ of all UTM?*

We note four reasons;

1. By requiring that the results of this paper apply to all elements of Λ, we will be able to connect the laws of physics to any UTM - so long as its programs are scheduled so as to maximize the entropy in the calculation of Ω. This makes the results more general.
2. If we were to consider a specific UTM for us to study, the chances that it would actually be the one describing the universe are next to NIL.
3. We consider the experience of an observer inside the universe but lacking complete knowledge of its laws or initial conditions. As a result, the observer cannot replicate on paper the UTM running the universe. Therefore, from the entropic UTM perspective, his understanding of the universe is limited to that which applies to all possible UTM.
4. In the section on quantum measurement (section 7.1), we will see how this formulation will allow us to derive a definition and mechanism for the collapse. Indeed, we will see that each measurement of the bits of Ω will eliminate from Λ the UTMs that are incompatible with the value of the measurement. It will be argued that this is isomorphic to the quantum measurement and collapse.

2 *Statistical physics*

We note the similarities between 1.4 and the Gibb's ensemble of thermodynamics. In fact, these similarities have been noted by other authors before [2]. Simple replacements (changing the name of the variables) are enough to switch back and forth between the two representations. The Gibb's ensemble compares to the hating probability as;

[2] K. Tadaki. A statistical mechanical interpretation of algorithmic information theory. https://arxiv.org/pdf/0801.4194.pdf, 2008; John C. Baez and Mike Stay. Algorithmic thermodynamics. arXiv:1010.2067 [math-ph], 2010; and Ming Li and Paul Vitányi. An introduction to kolmogorov complexity and its applications. Springer, 1997

Gibb's ensemble	Halting probability	
$Z = \sum_x e^{-\beta(E+pV+Fx)}$	$\Omega = \sum_{x=1}^{\infty} 2^{-E(x)-x}$	(2.1)

Let us do a quick recall of statistical physics, then we will give a thermodynamic interpretation of the halting probability.

2.1 *Recall of statistical physics*

In statistical physics, we are interested in the distribution that maximizes entropy

$$S = -k_b \sum_{x \in X} p(x) \ln p(x) \tag{2.2}$$

subject to the fixed macroscopic observables. The solution is the Gibbs ensemble. As an example we take Table 1 as the observables.

Observable	Conjugate variable
Energy E	Temperature $\beta = 1/(k_b T)$
Volume V	Pressure $\gamma = p/(k_b T)$
Number of particles N	Chemical potential $\delta = -\mu/(k_b T)$

Table 1: Typical observables of statistical mechanics.

then the partition function becomes

$$Z = \sum_{x \in X} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \tag{2.3}$$

The probability of occupation of a micro-state is

$$p(x) = \frac{1}{Z} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \tag{2.4}$$

the average values and their variance for the observables are

$$\overline{E} = \sum_{x \in X} p(x) E(x) \qquad \overline{E} = \frac{-\partial \ln Z}{\partial \beta} \qquad \overline{(\Delta E)^2} = \frac{\partial^2 \ln Z}{\partial \beta^2} \tag{2.5}$$

$$\overline{V} = \sum_{x \in X} p(x) V(x) \qquad \overline{V} = \frac{-\partial \ln Z}{\partial \gamma} \qquad \overline{(\Delta V)^2} = \frac{\partial^2 \ln Z}{\partial \gamma^2} \tag{2.6}$$

$$\overline{N} = \sum_{x \in X} p(x) N(x) \qquad \overline{N} = \frac{-\partial \ln Z}{\partial \delta} \qquad \overline{(\Delta N)^2} = \frac{\partial^2 \ln Z}{\partial \delta^2} \tag{2.7}$$

The laws of thermodynamics can be recovered from the partition function by taking the derivatives

$$\left.\frac{\partial S}{\partial E}\right|_{V,N} = \frac{1}{T} \qquad \left.\frac{\partial S}{\partial V}\right|_{E,N} = \frac{p}{T} \qquad \left.\frac{\partial S}{\partial N}\right|_{E,V} = -\frac{\mu}{T} \tag{2.8}$$

and summarizing them as

$$dE = TdS - pdV + \mu dN \tag{2.9}$$

which is known as the state equation of the thermodynamic system.

2.2 *Algorithmic thermodynamics and related work*

In their paper [3], John C. Baez and Mike Stay suggest an interpretation of algorithmic information theory based on thermodynamics, where the characteristics of programs are considered to be observables. Starting from Gregory Chaitin's Ω number, the halting probability

[3] John C. Baez and Mike Stay. Algorithmic thermodynamics. arXiv:1010.2067 [math-ph], 2010

$$\Omega = \sum_{p \text{ halts}} 2^{-|p|} \tag{2.10}$$

is extended with algorithmic observables to obtain

$$\Omega' = \sum_{x \in X} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \tag{2.11}$$

Noting the similarity between equation 2.3 and 2.11, they suggest an interpretation where E is the expected value of the logarithm of the program's runtime, V is the expected value of the length of the program and N 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 tradeoff between runtime and length. Roughly speaking, it counts how much you need to decrease the mean length to increase the mean log runtime by a specified amount, while holding the number of programs in the ensemble and their mean output fixed.

3. $\mu = -\delta/\beta$ is the *algorithmic potential* (analogous to chemical potential). Roughly speaking, this counts how much the mean log runtime increases when you increase the mean output while holding

the number of programs in the ensemble and their mean length fixed.

From equation 2.11, they derive analogues of Maxwell's relations and they consider thermodynamic cycles such as the Carnot cycle or Stoddard cycle. For this they introduce the concepts of *algorithmic heat* and *algorithmic work*.

The authors then claim that the choice of correspondence between thermodynamic observables and algorithmic observables is somewhat arbitrary and reference other authors [4] who have used completely different correspondences.

[4] Ming Li and Paul Vitányi. An introduction to kolmogorov complexity and its applications. Springer, 1997; and K. Tadaki. A statistical mechanical interpretation of algorithmic information theory. https://arxiv.org/pdf/0801.4194.pdf, 2008

2.3 *A physical interpretation of the halting probability*

Our preferred choice of correspondence will be one that recovers the conventional language of physics. We will map the program-observables to physical-observables as follows.

- The program-runtime is the number of *Iterations* a UTM needs to perform until a program halts. It is therefore natural to associate it with the physical *Time* in *seconds*. Indeed, a program requiring more iterations to halt will also require more time to terminate. If a system performs iterations at a faster or slower rate, the conjugate variable to time, the *Power* in *Watts*, can be adjusted to account for this variation.

- Its inverse, the algorithmic-frequency, is associated with the reverse of the second, s^{-1}, and its conjugate variable is the *Action* in *Joules-seconds*.

- The program-size is expressed in number of *bits*. Writing the bits one after the other on any medium (paper, disk drive, etc.) will require a certain physical size for each bit. As the line is the lowest dimensional geometry to spread bits, the program-size is naturally associated with the physical *length* as its simplest case. Furthermore, if an encoding medium would allow greater or lesser "packing-tightness" of the bits, it can be modelled with its conjugate variable the *Force* in *Newtons* pushing the bits together or pulling them apart. If one wishes instead to investigate geometries of higher dimensions, one can use different units. For the $3D$ case, the program-size can be mapped to a *Volume* in m^3 and its conjugate variable will be the *Pressure* in N/m^2. For the $2D$ case, it can be mapped to an *Area* in m^2 and its conjugate variable will be the *Surface tension* in N/m. In the section on the spin, we will see that problems occurs in dimensions higher than 3. As a result the volume case will be our upper bound.

- Only the halting event remains. As it is the only quantity with *no units*, it is natural to map it to the *Energy* in *Joules*. Indeed, in the Gibb's ensemble, the energy is the only observable not multiplied by a conjugate variable. Adding extra units to the halting event only to have them cancelled out by a conjugate variable would be futile.

Summarizing the points above, we obtain Table 2 as our mapping of choice between *algorithmic thermodynamics* and *physical thermodynamics*.

Observable	Variable	Units	Conjugate	Variable	Units
Halting event	E	J	Temperature	T	K
Program-size (length)	x	m	Force	F	N
Program-size (area)	A	m^2	Stiffness	γ	N/m
Program-size (volume)	V	m^3	Pressure	p	N/m^2
Program-frequency	τ	$1/s$	Action	$\mathcal{S}$	$J \times s$
Program-runtime	t	s	Power	P	W

Table 2: The preferred correspondence between *algorithmic thermodynamics* and *statistical physics*.

We combine these observables to construct a thermodynamic equation for the halting probability (equation 1.4),

$$\Omega^Z = \sum_x e^{-(\ln 2)\beta(E+Fx+Pt+\gamma A+pV+...)} \tag{2.12}$$

and its state equation is

$$dE = TdS - Pdt - Fdx - \gamma dA - pdV - \dots \tag{2.13}$$

where the triple dots represent other possible observables. We interpret the program x as a micro-state of the set of all prefix-free programs that are run on the UTM. It is easy to see that the function for Ω^Z is the partition function of the Gibbs ensemble of thermodynamics.

Both the running frequency and the runtime are associated with time and are the converse of each other. As a result we only need to select one of them as our conjugate-observable pair whenever we want to take into account the effects of time. In this work, we will select $\mathcal{S}\tau$ over Pt (and vice-versa) whenever it leads to conceptually simpler results.

3 *A maximally entropic calculation of* Ω

As Ω is formulated as a Gibb's ensemble, it is necessarily maximally entropic. Indeed, for an unspecified UTM, each bit ω_i of the normal number Ω is equally likely to be a 1 as it is to be a 0. However, during the calculation itself and at intermediary steps this will not necessarily be the case. To understand why, we must first understand how progress towards the calculation of Ω can be made.

Attempting to run programs on a UTM for the purposes of calculating Ω will have two difficulties. First, if we start one program and wait for it to terminate before starting another one, the UTM will hang at the first non-halting program. Second, if we start each program in parallel, since there are infinitely many such programs, the UTM will never return to work on the first one. The solution is to dovetail programs.

Definition 3.1 (Dovetailing)**.** *Dovetailing is a program execution strategy for a Turing machine to guarantee that progress will be made on arbitrarily-many programs even in the presence of non-halting programs.*

Definition 3.2 (Standard dovetailing)**.** *Consider the case of standard dovetailing. First, we start the shortest program and perform one iteration. Then, we start the second program and perform one iteration on the first and second program. Then, we start the third program and perform one iteration on the first, second and third program. And so on. Using dovetailing, progress will eventually be made on every program and no program will cause the TM to hang.*

The problem with standard dovetailing is that there is no guarantee that the system is maximally entropic during the calculation. To understand why, consider that the entropy of the bits of Ω computed via dovetailing. The bits are indeed compressible to a short algorithm - the dovetailing algorithm itself! This algorithm includes both the code to run the dovetailing and an encoding of the UTM. Its total size (in bits) is the upper bound for the entropy of the calculated bits of Ω.

To produce a maximally entropic calculation of the bits of Ω, we must make adjustments to standard dovetailing. How can we do that? Recall that the halting probability is a Gibb's ensemble which maximizes the entropy of the system subject to its program observables.

To adjust dovetailing so as to maximize the entropy throughout the calculation, it suffices to create a dovetailing algorithm as a Gibb's ensemble. Specifically, we must add program-observables to the halting probability so as to 1) produce a dovetailing-type algorithm and 2) eventually obtain Ω when $t \to \infty$.

To do so and as a result of the similarities between the halting probability and the Gibb's ensemble, we import the notions of thermodynamics into AIT to augment Ω with additional program observables, analogous to thermodynamic observables.

First, we augment the rightmost term of the exponential, x, with a conjugate variable. Lets call it F. Second we multiply the terms of the exponential by β. As $E(x)$ is either 0 or ∞, it absorbs β. With these additions, we obtain a Tadaki D-random number [5]. Here, we take the liberty to use the symbol F to define Tadaki's number as we will later connect this variable to the force in the physical interpretation.

[5] K. Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. http://arxiv.org/abs/nlin/0212001, 2002; and K. Tadaki. A statistical mechanical interpretation of algorithmic information theory. https://arxiv.org/pdf/0801.4194.pdf, 2008

$$\Omega^F = \sum_{x=1}^{\infty} 2^{-\beta[E(x)-Fx]} \tag{3.3}$$

Tadaki has shown that Gregory Chaitin's constant can be extended to include a compression term $F \in \mathbb{R}$ on 2^{-x}, which becomes 2^{-Fx} such that the Takadi constant Ω^F remains non-computable F-random. Furthermore, he goes to show that Ω^F's first n bits contain $|n - F|$ halting bits. For example, take the case where $F = 2$, then expanding the sum into binary we get

$$\begin{aligned}
\Omega^F &= \sum_{x=1}^{\infty} 2^{-Fx} && (3.4)\\
&= 2^{-2\times1} + 2^{-2\times2} + 2^{-2\times3} + 2^{-2\times4} + 2^{-2\times5} + ... && (3.5)\\
&= 2^{-2} + 2^{-4} + 2^{-6} + 2^{-8} + 2^{-10}... && (3.6)\\
&= 0.01 + 0.0001 + 0.000001 + 0.000000001 + ... && (3.7)\\
&= 0.0101010101... && (3.8)
\end{aligned}$$

The compression factor F "decompresses" the information by inserting some 0 in between the bits. It does not erase data. For the full proof, refer to Takadi's paper.

Third, we augment Takadi's number with the conjugate-observable pair of program-action to algorithmic-frequency, $\mathcal{S}\tau$. We obtain a partition function of algorithmic information theory (AIT).

$$\Omega^Z = \sum_{x=1}^{\infty} 2^{-\beta[E(x)-\mathcal{S}\tau-Fx]} \tag{3.9}$$

This partition function relates halting-event to program-size and to program-frequency. The program-observables are conjugated to the program-action $\mathcal{S}$ and the compressibility F, respectively. It has the following state equation,

$$dE = TdS - \mathcal{S}d\tau - Fdx \tag{3.10}$$

Let us now prove the two requirements for this algorithm; 1) It is a dovetailing algorithm and 2) it recovers Ω when $t \to \infty$.

Theorem 3.11. *At the limit of $t \to \infty$, we recover Ω^F*

Proof. A program x can have any value of $\mathcal{S}$ within $[0, \infty]$. If the program halts immediately, $\mathcal{S} = 0$. If it never halts, $\mathcal{S} = \infty$. If it halts after a certain time, $\mathcal{S} \in \mathbb{N}$. A program that never halts will not contribute to the halting partition. This will be the case if $\mathcal{S} = \infty$. As a result we obtain,

$$\lim_{\tau \to 0^+} \tau S_x = \lim_{t \to \infty} \frac{\mathcal{S}_x}{t} = \begin{cases} 0 & \text{x halts} \\ \infty & \text{otherwise} \end{cases} \tag{3.12}$$

As this is the definition of $E(x)$ (see 1.5), we obtain

$$\lim_{t \to \infty} \frac{\mathcal{S}_x}{t} = E(x) \tag{3.13}$$

Lemma 3.14. $E(x) + E(x) = E(x)$

Proof. $E(x)$ is either 0 or ∞. Since $0 + 0 = 0$ and $\infty + \infty = \infty$, the lemma holds. □

Therefore,

$$\lim_{t \to \infty} \Omega^Z = \lim_{t \to \infty} \left(\sum_{x=1}^{\infty} e^{-(\ln 2)\beta[E(x) + \mathcal{S}\tau + Fx]} \right) \tag{3.15}$$

$$= \sum_{x=1}^{\infty} e^{-(\ln 2)\beta[E(x) + E(x) + Fx]} \tag{3.16}$$

$$= \sum_{x=1}^{\infty} e^{-(\ln 2)\beta[E(x) + Fx]} \tag{3.17}$$

$$= \Omega^F \tag{3.18}$$

The question then is, can we recover Ω knowing Ω^F. The answer is of course yes as we just need to remove the zero-valued bits inserted in between the bits of Ω. □

Theorem 3.19. *To show that equation 3.9 dovetails programs, it suffices to show the following. For $0 < t < \infty$, the partition function Z is*

$$\Omega^Z(t) = \Omega - 2^{-k(t)}$$

where $2^{-k(t)}$ is an error rate that is monotonically decreasing to 0 as $t \to \infty$. As a result of increasing time, the calculation of Ω^Z produces an ever more precise estimation of Ω. The bits of Ω are found from left to right.

Proof. [6].

Definition 3.20. *For any $k \geq 0$ and time $t \geq 0$, let $k(t)$ be the location of the first zero bit after position k in the estimation of Ω.*

[6] Here, we have reproduced the definition of $k(t)$ and the proof provided by John C. Baez and Mike Stay in their paper on *algorithmic thermodynamics*.

John C. Baez and Mike Stay argue as follows:

The term $2^{-\mathcal{S}\tau}$ exponentially suppresses long program runtimes. Then because $-\frac{\mathcal{S}_x}{t}$ is a monotonically decreasing function of the running frequency and decreases faster than $k(t)$, there will be a time step where the total contribution of all the programs that have not halted yet is less than $2^{-k(t)}$.

□

For example, say

$$\Omega = 0.0111100\dots \tag{3.21}$$

To keep it simple we consider, in isolation, a single program and assume that all other programs have long halted (at $t \to 0^+$). Let us take the values $x = 5$ and $\mathcal{S}_x = 50$ for this program. We obtain,

$$Z_x(t) = 2^{-x}2^{-\frac{\mathcal{S}_x}{t}} \tag{3.22}$$

$$Z_5(t) = 2^{-5}2^{-\frac{50}{t}} \tag{3.23}$$

$$= 0.00001 \times 2^{-\frac{50}{t}} \tag{3.24}$$

The halting probability Ω is,

$$\Omega = 0.0111000 \cdots + Z_5(t) \tag{3.25}$$

Let us look at what happens as we vary t.

1. If $t \to 0^+$, then $Z_5(0^+) = 0$. Z differs from Ω by the maximum uncertainty of 2^{-5}. Therefore $\Omega - Z_5(0^+)$ is accurate only in its first 5 bits.

2. As $t \to \infty$, then $Z_5(\infty) = 0.00001$.

3. Between 0 and ∞, $Z_5(t)$ varies from 2^{-5} at $t = 0$ to 0 at $t \to \infty$. Since $-(\mathcal{S}_5/t)$ is monotonically decreasing, the uncertainty $2^{-k(t)}$ must decrease monotonically to 0 as t increases.

4. At distances further than $2^{-k(t)}$, the partition function contains bits of programs that have yet to halt. So, in a sort, a reversal of time occurs where halting information is available before the time t is long enough for the program to have halted.

Remark 3.26. *In this construction, $\mathcal{S}$ is the bearer of non-halting information and is non-computable. As a result, the entropy of the dovetail algorithm is bound by the entropy of $\mathcal{S}$ which is equal to Ω. The dovetailing algorithm is able to grow an entropy equal to that of Ω.*

Part I

The laws of physics

We can refer to the Gibb's ensemble of the halting probability either as its physical interpretation or as its AIT interpretation. The two interpretations are equivalent and differ only in the nomenclature. In what follows, we will use the physical interpretation during the mathematical derivation of the laws of physics and will briefly discuss the AIT interpretation when appropriate. We compare the two interpretations in Table 3.

Variable	AIT interpretation	Physical interpretation
E	Halting event	Energy
x	Program-size	Length
A	Program-size	Area
V	Program-size	Volume
F	Compressibility	Force
γ	Compressibility	Stiffness
p	Compressibility	Pressure
t	Runtime	Time
P	Compute-power	Power
τ	Program-frequency	Frequency
$\mathcal{S}$	Program-action	Action
T	Randomness	Temperature

Table 3: The *AIT-interpretation* compared to the *physical-interpretation*.

4 *Entropic spacetime*

We will derive a maximum speed, light-cone and the Lorentz's transformation from the physical interpretation, then we will discuss its AIT interpretation.

4.1 *Speed of light*

We now investigate the halting partition with the Power to time pair ($P \times t$). To obtain it, we replace τ by t^{-1}.

Theorem 4.1. *An object travelling at more than c, will violate the second law of thermodynamics.*

Proof.

$$
\begin{aligned}
dE &= TdS - \mathcal{S}d\tau - Fdx - \gamma dA - pdV && \text{(State equation)} \\
0 &= TdS - \mathcal{S}d\tau - Fdx && \text{(Posing } dE, dA \text{ and } dV \text{ to 0)} \\
TdS &= Fdx + \mathcal{S}d\tau && \text{(Addition by } Fdx + \mathcal{S}d\tau) \\
TdS &= Fdx - \mathcal{S}t^{-2}dt && (d\tau = -t^{-2}dt) \\
TdS &= Fdx - Pdt && \text{(Posing } P = \mathcal{S}t^{-2}) \\
\frac{T}{F}\frac{dS}{dt} &= \frac{dx}{dt} - \frac{P}{F} && (4.2)
\end{aligned}
$$

Note that the units for each term are meters per second. The equation therefore relates a speed to a change of entropy.

Let us look at three cases:

1. If $\frac{dx}{dt} - \frac{P}{F} < 0$, then $\frac{dS}{dt} < 0$. The entropy decreases with time.
2. If $\frac{dx}{dt} - \frac{P}{F} > 0$, then $\frac{dS}{dt} > 0$. The entropy increases with time.
3. If $\frac{dx}{dt} - \frac{P}{F} = 0$, then $\frac{dS}{dt} = 0$. The entropy remains constant.

To understand why this implies a speed barrier at P/F, we must ask how can a UTM decrease the entropy of Ω. To do so it must erase the value of a bit of Ω, which violates the conservation of information. Hence any system which conserves information will have a characteristic power and a characteristic force which limits the speed of the system.

□

Taking P to be the characteristic Planck power, and F to be the characteristic Planck force of the universe, we do in fact recover the speed of light.

$$P\left(\frac{1}{F}\right) = \frac{c^5}{G}\left(\frac{G}{c^4}\right) = c \tag{4.3}$$

4.2 *Light-cone*

We look at the thermodynamic cycle of the system transiting through time and space starting at A_0 to A_t to A_{xt} and back to A_0 as illustrated on Figure 1. During the transitions and to keep the energy constant, tradeoffs must be made between time, distance and entropy. This cycle is reminiscent of other thermodynamic cycles such as those involving pressure and volume, etc. The cycle presented here is reminiscent of relativistic light cones.

We work in the quasi static approximation

$$\Delta E = T\Delta S - F\Delta x + P\Delta t \tag{4.4}$$

and we pose that $\Delta E = 0$ throughout the cycle

$$T\Delta S = F\Delta x - P\Delta t \tag{4.5}$$

A_t to A_{xt}: As we translate A_t closer in space to A_{xt} while keeping the time fixed, the entropy must increase to compensate. This situation occurs when $\Delta x > 0$ and when $\Delta t = 0$.

$$(T\Delta S = F\Delta x - P\Delta t|_{\Delta t=0} \tag{4.6}$$

$$\implies \Delta S = \frac{F}{T}\Delta x \tag{4.7}$$

From the equation above, we note that ΔS is positive when $\Delta x > 0$. For a UTM calculating Ω, its entropy will be higher the less programs have halted. Hence a position far away from A_0 will appear quite young having benefited from very few iterations of the scheduler.

A_0 to A_t: As we translate A_0 backward in time to A_t while keeping the distance fixed, the entropy must increase to compensate. This situation occurs when $\Delta t < 0$ and when $\Delta x = 0$.

$$(T\Delta S = F\Delta x - P\Delta t|_{\Delta x=0} \tag{4.8}$$

$$\implies \Delta S = -\frac{P}{T}\Delta t \tag{4.9}$$

From the equation above, we note that ΔS is negative when $\Delta t < 0$. To travel backward in time, the system must erase halting bits from its pool of information so as to increase its entropy. As this would violate the conservation of information, an irreversible arrow of time is guaranteed. We will discuss this point in more detail in the section on entropic time (section 7).

A_{xt} to A_0: As we translate A_{xt} forward in time and backward in space to A_0 keeping the entropic constant ($\Delta S = 0$), we have movement at the speed c.

$$(T\Delta S = F\Delta x - P\Delta t|_{\Delta S=0} \tag{4.10}$$

$$\implies \frac{\Delta x}{\Delta t} = \frac{P}{F} = c \tag{4.11}$$

From the equation above, an object travelling at speed c is neither encouraged nor discouraged by entropic considerations. Hence it experiences no change in its perception of time.

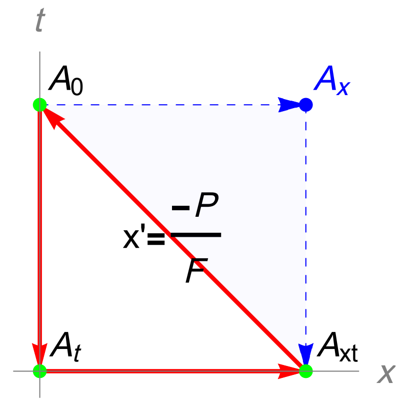

Figure 1: A thermodynamic cycle through space, time and entropy as observables.

From the results of this section, we conclude that the halting entropy guarantees that positions further away will appear quite young and that travelling backward in time is impossible unless halting information is erased.

4.3 Lorentz's transformation

To recover the Lorentz's factor γ, it suffices to take the triangle produced by A_0 to A_{xt} to A_t to A_0. The longest segment is posed to be of length ct, then the other two are posed to be ct' and vt. Using the Pythagorean rule,

$$\left(\overline{A_{xt}A_t}\right)^2 + \left(\overline{A_tA_0}\right)^2 = \left(\overline{A_0A_{xt}}\right)^2 \tag{4.12}$$

$$(ct')^2 + (vt)^2 = (ct)^2 \tag{4.13}$$

Then solving for t/t', we recover the Lorentz's factor γ

$$c^2t'^2 + v^2t^2 = c^2t^2 \tag{4.14}$$

$$t'^2 + \frac{v^2}{c^2}t^2 = t^2 \tag{4.15}$$

$$t'^2 = t^2 - \frac{v^2}{c^2}t^2 \tag{4.16}$$

$$t' = \sqrt{t^2 - \frac{v^2}{c^2}t^2} \tag{4.17}$$

$$t' = t\sqrt{1 - \frac{v^2}{c^2}} \tag{4.18}$$

$$\frac{t'}{t} = \sqrt{1 - \frac{v^2}{c^2}} \tag{4.19}$$

$$\frac{t}{t'} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \gamma \tag{4.20}$$

From the AIT perspective, special relativity can be interpreted as follows. Shorter programs have benefited from longer execution time such that their lead running time is related to longer program-size via P/F. This dynamic on an entropic UTM is governed as per special relativity. Programs exceeding an execution speed of P/F would lower the entropy.

5 Entropic space

In this section we investigate the following three reductions of the state equation, each respectively corresponding to a 1,2 or 3 dimensional encoding of program sizes.

$$dE = dt = dA = dV = 0 \implies \quad TdS = Fdx \qquad \text{(Length)}$$
$$dE = dt = dx = dV = 0 \implies \quad TdS = \gamma dA \qquad \text{(Area)}$$
$$dE = dt = dx = dA = 0 \implies \quad TdS = pdV \qquad \text{(Volume)}$$

Remark 5.1. *As described in section 6.3, three dimensions appear to be an upper bound for physical encoding of program-sizes. As a result we will limit ourselves to three dimensions.*

We suggest an interpretation such that linear entropy is dominant at short distances until it is overtaken by area entropy which is itself eventually overtaken by volumetric entropy. We will show that linear entropy produces the law of inertia, area entropy produces general relativity and volumetric entropy suggests dark energy. These correspond to the three characteristic scales of the universe, where Newton's law dominates at local scales, general relativity dominate at galactic scales and dark energy dominate at cosmological scales.

To recover the laws, we will need to identify a constant temperature applicable to each relation. In the case of dx, the constant temperature will be the Unruh temperature as experienced by an accelerating body. In the case of dA, the constant temperature will be the radially symmetric red shift temperature. In the case of dV and the dark energy we will cite a suggestion from the literature.

For each of the three relations, we consider that the entropy of the system is related to the halting bits of the UTM. Hence, the entropy is proportional to the number of bits and the following holds

$$TdS = k_b TdN \tag{5.2}$$

From the AIT interpretation, each relation corresponds to a different dimension according to which the scheduler can assign an execution priority.

5.1 *Exfoliation of spacetime*

As we have noticed in the section on the speed of light, as time progresses more program halts and the halting entropy of the UTM is decreased. Such a decrease in entropy over time would imply a violation of the second law of thermodynamics unless an entropy sink was available to compensate. We suggest that the other thermodynamic observables such as dx, dA and dV act as such a sink. We will call the compensatory increase in entropy of the non-halting information of the system; the exfoliation of spacetime.

5.2 *Law of inertia*

To recover the law of inertia, we consider the linear case where $Fdx = TdS$ such that the entropy S is proportional to the distance x. Such a correspondance was previously investigated by Erik Verlinde[7]. In his paper he argued that the temperature of the position-encoding bits should be the Unruh temperature, the temperature experienced by an accelerating object.

[7] Erik Verlinde. On the origin of gravity and the laws of newton. arXiv:1001.0785v1 [hep-th], 2010

$$T = \frac{ha}{k_b c} \qquad \text{(Unruh temperature)} \qquad (5.3)$$

Theorem 5.4. *The force is related to the acceleration via the mass, such that*

$$F = ma$$

Proof.

$$\begin{aligned}
dE &= TdS - Pdt - Fdx - \gamma dA - pdV && \text{(State equation)} \\
0 &= TdS - Fdx && \text{(Posing } dE, dt, dA, dV \text{ to 0)} \\
Fdx &= TdS && \text{(Addition by } Fdx\text{)} \\
F &= T\frac{dS}{dx} && \text{(Division by } dx\text{)} \\
F &= k_b T\frac{dN}{dx} && \text{(Binary entropy)}
\end{aligned}$$

To link this equation to an acceleration, we consider the case where the bits are at a uniform Unruh temperature.

$$\begin{aligned}
F &= k_b\left(\frac{ha}{k_b c}\right)\frac{dN}{dx} && \text{(Unruh temperature)} \\
F &= \frac{h}{c}\frac{dN}{dx}a && \text{(Clean up)}
\end{aligned}$$

What is the term multiplying the acceleration? By unit inspection, it must have the units of the mass. Let us indeed pose it to be equal to m so as to recover $F = ma$ then solve for dN/dx, we obtain

$$m = \frac{h}{c}\frac{dN}{dx} \qquad (5.5)$$

$$\implies \frac{dx}{dN} = \frac{h}{mc} = \lambda \qquad (5.6)$$

, the Compton wavelength! Since every massive object has a Compton wavelength, the law of inertia can be recovered in all cases. The AIT interpretation is now clear. The UTM uses an algorithm that multiples a length scale λ to a program length in order to encode linear spacial positions... and the Compton wavelength is the physical equivalent of this scale!

□

5.3 *General Relativity*

To recover general relativity, we will proceed in a manner similar to the previous case, except that we will use the area to entropy relation of $\gamma dA = TdS$ as a starting point. This situation describes a scenario where bits of programs are spread over a surface, e.g. a disk drive or a holographic screen, etc.

We must justify a temperature for the system such that the temperature is the same for every bit. In the case of the surface of a sphere (e.g. a holographic screen), the temperature will need to have a radial symmetry. Erik Verlinde[8] in the same paper argues that this temperature should be the red shift temperature. Indeed, it is a good candidate as the red shift temperature is the same in all directions for a given distance. As per Erik Verlinde's paper, the red shift temperature for this system is

[8] Erik Verlinde. On the origin of gravity and the laws of newton. arXiv:1001.0785v1 [hep-th], 2010

$$T = \frac{\hbar}{2\pi c} e^{\phi} N^b \nabla_b \phi \tag{5.7}$$

Furthermore, we consider the case where the bits on the surface of the sphere are maximally compressed such that each occupies the Planck area.

Theorem 5.8 (General relativity)**.** *The state equation $TdS = \gamma dA$ implies general relativity.*

Proof.

$$dE = TdS - Pdt - Fdx - \gamma dA - pdV \qquad \text{(State equation)}$$
$$0 = TdS - \gamma dA \qquad \text{(Posing } dE, dt, dx, dV \text{ to 0)}$$
$$TdS = \gamma dA \qquad \text{(Addition by } \gamma dA\text{)}$$
$$k_b TdN = \gamma dA \qquad \text{(Binary entropy)}$$
$$TdN = \frac{\gamma}{k_b} dA \qquad \text{(Division by } k_b\text{)}$$
$$\int TdN = \oiint_A \frac{\gamma}{k_b} dA \qquad \text{(Integration)}$$
$$\frac{1}{2}\int TdN = \frac{1}{2}\oiint_A \frac{\gamma}{k_b} dA \qquad \text{(Division by 2)}$$
$$E = \frac{1}{2}\oiint_A \frac{\gamma}{k_b} dA \qquad \text{(Equipartition theorem)}$$
$$M = \frac{1}{2c^2}\oiint_A \frac{\gamma}{k_b} dA \qquad \text{(Division by } c^2\text{)}$$
$$M = \frac{1}{2c^2}\oiint_A T\frac{1}{L^2} dA \qquad \text{(Unit shift)}$$
$$M = \frac{1}{2c^2}\oiint_A \left(\frac{\hbar}{2\pi c} e^{\phi} N^b \nabla_b \phi\right) \frac{1}{L^2} dA \qquad \text{(Red shift temperature)}$$
$$M = \frac{1}{2c^2}\oiint_A \left(\frac{\hbar}{2\pi c} e^{\phi} N^b \nabla_b \phi\right) \left(\frac{c^3 dA}{G\hbar}\right) \qquad \text{(Planck area)}$$
$$M = \frac{1}{4\pi G}\oiint_A \left(e^{\phi} N^b \nabla_b \phi\right) dA \qquad \text{(Clean up)}$$

We obtain the generalization of Gauss' law to general relativity, which is enough to recover general relativity[9]. □

[9] Erik Verlinde. On the origin of gravity and the laws of newton. arXiv:1001.0785v1 [hep-th], 2010

How do we interpret this result from the universal Turing machine perspective? As programs halt, the corresponding bit of Ω flips from 0 to 1. As those flips are random, regions of higher or lower concentration of flipping may emerge over time. As a result, a scheduler can improve its entropy production by reducing the work done in regions of high bit flipping and instead focus on regions of low bit flipping. To do this, the scheduler will reduce the number of iterations made available to programs in regions of high bit flipping, so as to slow the halting progress in those regions.

This reduction of iterations slows the passage of time. And if such is related to the bit entropy on the surface of a sphere, say a holographic screen, the entropy is maximized when the scheduler recovers the law of general relativity.

5.4 Dark energy

In this section, we suggest that the third and final relation $TdS = pdV$ is related to dark energy.

$$dE = TdS - Pdt - Fdx - \gamma dA - pdV \qquad \text{(State equation)}$$
$$0 = TdS - pdV \qquad \text{(Posing } dE, dt, dx, dA \text{ to 0)}$$
$$TdS = pdV \qquad (5.9)$$

A derivation of dark energy is outside the scope of this paper. Therefore we will refer to another paper[10] by Erik Verlinde. In it, he makes a compelling argument that a volumetric entropy can account for the observed dark energy. He links a volumetric entropy overtaking other forms of entropy over large distances to the theory of elasticity, and recovers numerous experimental results.

[10] Erik Verlinde. Emergent gravity and the dark universe. https://arxiv.org/pdf/1611.02269.pdf, 2016

6 Quantum mechanics

Thus far, we have investigated what occurs when the values of the bits of Ω are unspecified. We have recovered relations between entropy and the program observables such that they correspond to laws of physics: law of inertia, special relativity, general relativity and possibly dark energy. In this section, we return to our UTM roots and consider the impact of non-computability on these relations. We will see that it is the non-computability of Ω that is responsible for quantum mechanical effects "within the UTM". With this, we will be able to derive spins, photons, the Schrödinger equation and the Dirac equation.

6.1 Spin

As we have proven the existence of a maximum speed in the universe and further presented a thermodynamic cycle as a light-cone, it follows that the Lorentz group is a credible representation to continue our investigation with. Let us investigate this connection in more detail.

The Lorentz group is represented by the Lie Group $O(1,3)$. The Lorentz group embeds multiple other representations as subgroups. Some of them are:

1. The subgroup of transformations that preserves the direction of time is called **orthochronous** and is represented by $O^+(1,3)$
2. The subgroup of transformations that preserves the orientation, having a determinant of $+1$, is called **proper** and is represented by $SO(1,3)$.

3. The subgroup which includes transformations that are both proper and orthochronous is denoted by $SO^+(1,3)$.

4. The set of all rotations also forms a subset and is denoted by $SO(3)$.

We will now investigate how the halting partition behaves under transformation of these subgroups. Let us inject the simplest of the subgroups listed above, the rotations $SO(3)$, and investigate the results.

Theorem 6.1. *Injecting the $SO(3)$ rotation group into the partition function produces a quantum partition function where each halting bit is a spin.*

Remark 6.2. *We will represent the $SO(3)$ group via a unit quaternion of $\vec{r} = u\vec{i} + v\vec{j} + w\vec{k}$ where $|\vec{r}| = 1$. Then we will inject it as a conjugate to the halting event observable.*

Remark 6.3. *We recall the existence of an exponential map between a Lie Group and its corresponding Lie Algebra such that,*

$$SL(2,\mathbb{C}) = exp(\mathfrak{sl}(2,\mathbb{C})) \tag{6.4}$$
$$SO(3) = exp(\mathfrak{so}(3)) \tag{6.5}$$
$$SU(2) = exp(\mathfrak{su}(2)) \tag{6.6}$$
$$\text{etc.}$$

The existence of such a correspondence allows us to inject the Lie algebra $\mathfrak{so}(3)$ in an exponential function, such as the Gibb's ensemble, and still recover a workable Lie Group after the exponentiation is executed.

Proof.

$$\Omega^{\beta} = \sum_{x}^{\infty} e^{-\beta(\ln 2)E} \qquad \text{(Tadaki D-random number)}$$
$$Z = \sum_{x}^{\infty} e^{-\beta(\ln 2)(u\vec{i}+v\vec{j}+w\vec{k})E} \qquad \text{(Remark 6.2)}$$
$$= e^{-\beta(\ln 2)} SU(2)_1 \omega_1 + e^{-\beta(\ln 2)} SU(2)_2 \omega_2 + \dots \qquad \text{(Remark 6.3)}$$

The $SU(2)$ matrices are defined as

$$SU(2) = \left\{ \begin{bmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{bmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1 \right\} \tag{6.7}$$

We recall that the values ω_i are not computable. Furthermore, since they are multiplied by a 2×2 matrix, it is best to represent them as 2×1 matrix. This representation acknowledges that each bit can have two possible distinct values such that they are orthogonal when injected with a rotational observable.

$$= e^{-\beta(\ln 2)} SU(2)_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + e^{-\beta(\ln 2)} SU(2)_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \dots \tag{6.8}$$

We recall the quantum mechanical definition of a partition function.

$$Z = \operatorname{tr}\left\{ e^{-\beta \hat{H}} \right\} \tag{6.9}$$

The trace of the eigenvalues of the orthogonal states of the hamiltonian is equal to the partition injected with the unit quaternion observable. We have therefore recovered a thermodynamic quantum system comprised of a mixture of spins. □

This is enough to recover the usual spin, but with significant improvements. Let us compare the state pre-measurement to the state post-measurement to notice the differences on Ω^Z.

Pre-measurement, the value of ω_i is unknown. Hence it takes the form of a 2×1 matrix. It is therefore possible to define a matrix of the $SU(2)$ group and to multiply it by ω. Macroscopically the $SU(2)$ matrix is perceived as a thermodynamic observable related to the rotation of the system. This description corresponds to the usual spin.

Post-measurement, ω_i ceases to be a matrix and is instead fixed to a specific bit. The $SU(2)$ matrix can no longer be multiplied with ω, now a simple scalar, and the macroscopic observable of rotation must be eliminated from the description along with the $SU(2)$ matrix. This system is no longer a spin, but a classical bit.

This is where we notice the improvement over the usual description of the spin. Indeed, the conventional theory of spins is unable to predict the value of the quantum measurement and only state that it is random as per experimental evidence. Here however, the value of a spin measurement is well defined and even deterministic but nonetheless provably non-computable. It is the value of the corresponding bit of the halting probability. Although its value is deterministic and reproducible, it can be shown to be related to the non-halting problem in such a way as to be non-computable and algorithmically random to any observer.

6.2 *Photons*

Theorem 6.10. *Injecting the $SO(2)$ rotation group into the partition function produces a quantum partition function where each halting bit is a* *photon.*

Proof. It suffices to repeat the proof of theorem 6.1, but to use the $SO(2)$ rotation group instead of $SO(3)$. We recover an $U(1)$ matrix multiplying the two-state bits. We obtain a mixture of photons. □

6.3 *Other dimensions*

Remark 6.11. *Injecting $SO(1)$ produces a multiplication by* 1 *and has no impact on the partition function.*

Remark 6.12. *If we were to inject a rotation for a dimension higher than $SO(3)$, the size of the matrix would be too high to allow its multiplication with a two-bit system. This could suggest that rotations in dimensions higher than three cannot be defined as a* <u>*macroscopic*</u> *thermodynamic variable.*

Remark 6.13. *A derivation of the $SU(3)$ representation group from the UTM description has not been investigated by the author and could be an interesting area of future research.*

6.4 *Schrödinger equation*

In a previous section, we have used the program-size to entropy relation $TdS = Fdx$ to recover $F = ma$. In this section we use the same relation but we extend it with the non-computable features of the UTM. Doing so will allow us to recover the Schrödinger equation.

We recall that a UTM encodes position via a scale λ multiplied by a program-size. As a result, the UTM can only express a position if the program with the corresponding size is part of its partition function (e.i. it halts). In this section, we will argue that the missing non-halting programs are responsible for a universal Brownian motion in space applicable to the dx variable. This will be enough to recover the Schrödinger's equation.

Theorem 6.14. *A position described with missing program-sizes will evolve in time according to Schrödinger's equation.*

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \left[\frac{-\hbar^2}{2m}\nabla^2 + V(x,t)\right]\psi(x,t)$$

The proof is slightly more involved than the preceding theorems. First, here is a sketch of the proof.

1. We will show that non-halting programs leave holes in space such that a position cannot be expressed.

2. We will show that these holes are causing a Brownian motion of the encoded position.

3. We will derive its diffusion coefficient to be $\hbar/(2m)$.

4. We will consider that the presence of any external field is experienced as acceleration via $F = ma$.

5. Using the well known Brownian motion equations of Langevin, we show that the above reproduces Schrödinger's equation exactly.

Lemma 6.15. *A spacial encoding based on programs multiplied by λ leave holes in space corresponding to non-halting programs.*

Proof. We recall the general halting partition

$$Z = \sum_x e^{-(\ln 2)\beta(E+Fx)} \tag{6.16}$$

We have also seen that the observable x denotes program lengths. However, not all programs halt hence some lengths are missing from the sum. These missing programs are holes in space the position of which cannot be expressed by the UTM's positional algorithm. Since Ω is a normal number, we can expect the position of these holes to be algorithmically random.

□

Lemma 6.17. *A particle in space will experience Brownian motion due to the holes.*

Proof. We will calculate the average displacement $\overline{\Delta x}$ of a particle subjected to entropic positioning and space holes. Since Z is a normal number, we conclude that half of the program's lengths are available to describe position and half are not. Therefore, to describe a particle at position x, there is a 50% chance there is a halting program available to express it. And in the case where there is no program at exactly x, then there is a 50% chance that there will be one at position $x+1$, and so on. In other words, a particle at x has 50% chance of being at x, 25% chance of being at $x+1$, 12.5% chance of being at $x+2$, etc. Expressed as a sum, we obtain

$$\overline{\Delta x} = \frac{1}{2}0 + \frac{1}{4}1 + \frac{1}{8}2 + \frac{1}{16}3 + ... \tag{6.18}$$

$$= \sum_{i=0}^{\infty} \frac{i}{2^{i+1}} \tag{6.19}$$

$$= 1 \tag{6.20}$$

On average, as it moves through space, a position will shift by $\overline{\Delta x} = 1$ at each iteration of the Brownian motion. □

Lemma 6.21. *The diffusion coefficient of the described Brownian motion is*

$$D = \frac{\hbar}{2m}$$

Proof. It is well known that in general the diffusion coefficient of Brownian motion is given by

$$D = \frac{l^2}{2\tau} \tag{6.22}$$

where l is the length of the random step and τ is the frequency of the occurrence of the steps. Entropic position uses the scale factor λ for each unit of length. As we are now dealing with wave, we will use the reduced Compton wavelength. When λ is the reduced Compton wavelength, we get a scaling factor of

$$\lambda = \frac{\hbar}{mc} \tag{6.23}$$

Since entropic positioning can only express position as multiples of λ, we take it as the Brownian step of length l. The diffusion coefficient becomes

$$D = \left(\frac{\hbar}{mc}\right)^2 \frac{1}{2\tau} \tag{6.24}$$

This leaves of us with the need to define τ. For τ, we take the characteristic frequency of the wave $E = \hbar\omega$. Solving for $\tau = 1/\omega$, we obtain

$$\omega = \frac{E}{\hbar} \tag{6.25}$$

$$\omega^{-1} = \frac{\hbar}{E} = \tau \tag{6.26}$$

Replacing τ in the equation for D, we obtain

$$D = \frac{\hbar^2}{m^2c^2}\left(\frac{E}{2\hbar}\right) \tag{6.27}$$

Using $E = mc^2$, and reducing the constants, we obtain our final expression of D,

$$D = \frac{\hbar^2}{m^2c^2}\frac{(mc^2)}{2\hbar} \tag{6.28}$$

$$= \frac{\hbar}{2m} \tag{6.29}$$

□

Lemma 6.30. *The Langevin equations for Brownian motion with a diffusion coefficient of $\hbar/(2m)$ and an external inertial field $F = ma$ reproduces Schrödinger's equation.*

Proof. We recall the well known Langevin equation,

$$d\left[x(t)\right] = v(t)dt \tag{6.31}$$

$$d\left[v(t)\right] = -\frac{\gamma}{m}v(t)dt + \frac{1}{m}W(t)dt \tag{6.32}$$

where $W(t)$ is a random force and a stochastic variable giving the effect of a background noise to the motion of the particle.

From $F = ma$ and replacing the acceleration $d[v(t)]/dt$ with F/m, Edward Nelson [11] is able to show that the Langevin equation becomes,

$$\frac{1}{2}\nabla u^2 + D\nabla^2 u = \frac{1}{m}\nabla V \tag{6.33}$$

where D is the diffusion coefficient of $\hbar/(2m)$ obtained in lemma 6.21, where $F = -\nabla V$, where $u = v\nabla \ln \rho$ and ρ is the probability density of $x(t)$. For brevity, the proof of 6.33 is omitted here but can be reviewed in Nelson's paper. Eliminating the gradients on each side and simplifying the constants, we obtain

$$\frac{m}{2}u^2 + \frac{\hbar}{2}\nabla u = V - E \tag{6.34}$$

where E is the arbitrary integration constant. This equation in non-linear because of the term u^2 but it can be made linear by a change of dependant variable. To make it linear, let us pose

$$u = \frac{\hbar}{m}\frac{1}{\psi}\nabla\psi \tag{6.35}$$

and replace it into equation 6.34 , we obtain

$$\frac{m}{2}\left(\frac{\hbar}{m}\frac{1}{\psi}\nabla\psi\right)^2 + \frac{\hbar}{2}\nabla\left(\frac{\hbar}{m}\frac{1}{\psi}\nabla\psi\right) = V - E \tag{6.36}$$

taking the gradients and the exponents, we obtain

$$\frac{\hbar^2}{2m}\frac{1}{\psi^2}\nabla^2\psi + \frac{\hbar^2}{2m}\left[-\frac{1}{\psi^2}\nabla^2\psi + \frac{1}{\psi}\nabla^2\psi\right] = V - E \tag{6.37}$$

The first two terms cancel each other.

$$\frac{\hbar^2}{2m}\frac{1}{\psi}\nabla^2\psi = V - E \tag{6.38}$$

[11] Edward Nelson. Derivation of schrödinger's equation from newtonian mechanics. `http://dieumsnh.qfb.umich.mx/archivoshistoricosMQ/ModernaHist/Nelson%20a.pdf`, 1966. Physical Review, Volume 150, Number 4

Finally, it simplifies to

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V - E\right]\psi = 0 \tag{6.39}$$

which is the time independent Schrödinger's equation.

□

We are now ready to derive the time dependent Schrödinger equation and prove theorem 6.14.

Proof. We use the same proof used by Edward Nelson in the same paper. Starting from the time dependent Schrödinger equation, we show that a replacement of $\psi = e^{R+iS}$ leads to the Langevin equation of Brownian motion.

$$\frac{\partial\psi}{\partial t} = i\frac{\hbar}{2m}\nabla^2\psi - i\frac{1}{\hbar}V\psi \tag{6.40}$$

Replacing ψ with e^{R+iS}, we obtain

$$\frac{\partial\left(e^{R+iS}\right)}{\partial t} = i\frac{\hbar}{2m}\nabla^2\left(e^{R+iS}\right) - i\frac{1}{\hbar}V\left(e^{R+iS}\right) \tag{6.41}$$

Taking the derivatives and the gradients, we obtain

$$\left[\frac{\partial R}{\partial t} + i\frac{\partial S}{\partial t}\right]\left(e^{R+iS}\right) = \frac{i\hbar}{2m}\left[\nabla^2 R + i\nabla^2 S + (\nabla(R+iS))^2\right]\left(e^{R+iS}\right) - i\frac{1}{\hbar}V\left(e^{R+iS}\right) \tag{6.42}$$

Eliminating e^{R+iS} from each side and simplifying, we obtain

$$\frac{\partial R}{\partial t} + i\frac{\partial S}{\partial t} = \frac{i\hbar}{2m}\left[\nabla^2 R + i\nabla^2 S + (\nabla(R+iS))^2\right] - i\frac{1}{\hbar}V \qquad \text{(eliminating } e^{R+iS}\text{)}$$

$$\frac{\partial R}{\partial t} + i\frac{\partial S}{\partial t} = \frac{i\hbar}{2m}\left[\nabla^2 R + i\nabla^2 S + (\nabla R)^2 + 2i\nabla R\nabla S - (\nabla S)^2\right] - i\frac{1}{\hbar}V \qquad \text{(taking the product)}$$

$$\frac{\partial R}{\partial t} + i\frac{\partial S}{\partial t} = \frac{\hbar}{2m}\left[i\nabla^2 R - \nabla^2 S + i(\nabla R)^2 - 2\nabla R\nabla S - i(\nabla S)^2\right] - i\frac{1}{\hbar}V \qquad \text{(distributing the } i\text{)}$$

We obtain two equations by separating the real and the imaginary parts

$$\frac{\partial R}{\partial t} = \frac{\hbar}{2m}\left[-\nabla^2 S - 2\nabla R\nabla S\right] \tag{6.43}$$

$$\frac{\partial S}{\partial t} = \frac{\hbar}{2m}\left[\nabla^2 R + (\nabla R)^2 - (\nabla S)^2\right] - \frac{1}{\hbar}V \tag{6.44}$$

This is equivalent to the Langevin equations with some replacements

$$\frac{\partial u}{\partial t} = -\frac{\hbar}{2m}\nabla^2 v - \nabla(v\cdot u) \tag{6.45}$$

$$\frac{\partial v}{\partial t} = \frac{\hbar}{2m}\nabla^2 u + \frac{1}{2}\nabla(u^2) - \frac{1}{2}\nabla(v^2) - \frac{1}{m}\nabla V \tag{6.46}$$

Lemma 6.47. *Equation 6.43 with the replacements $\nabla R = (m/\hbar)u$ and $\nabla S = (m/\hbar)v$ produces 6.45*

Proof.

$$\frac{\partial R}{\partial t} = \frac{\hbar}{2m}\left[-\nabla^2 S - 2\nabla R \nabla S\right] \qquad \text{(equation 6.43)}$$

$$\nabla \frac{\partial R}{\partial t} = \nabla \frac{\hbar}{2m}\left[-\nabla^2 S - 2\nabla R \nabla S\right] \qquad \text{(multiplying by } \nabla)$$

$$\frac{\partial \nabla R}{\partial t} = \nabla \frac{\hbar}{2m}\left[-\nabla \nabla S - 2\nabla R \nabla S\right] \qquad (6.48)$$

$$\frac{m}{\hbar}\frac{\partial u}{\partial t} = \nabla \frac{\hbar}{2m}\left[-\nabla\left(\frac{m}{\hbar}v\right) - 2\left(\frac{m}{\hbar}u\right)\left(\frac{m}{\hbar}v\right)\right] \qquad \text{(replacing } \nabla R \text{ and } \nabla S)$$

$$\frac{\partial u}{\partial t} = \nabla \frac{\hbar}{2m}\left[-\nabla v - 2\frac{m}{\hbar}u \cdot v\right] \qquad \text{(eliminating } m/\hbar)$$

$$\frac{\partial u}{\partial t} = -\frac{\hbar}{2m}\nabla^2 v - \nabla(u \cdot v) \qquad \text{(equation 6.45)}$$

□

Lemma 6.49. *Equation 6.44 with the replacements $\nabla R = (m/\hbar)u$ and $\nabla S = (m/\hbar)v$ produces 6.46*

Proof.

$$\frac{\partial S}{\partial t} = \frac{\hbar}{2m}\left[\nabla^2 R + (\nabla R)^2 - (\nabla S)^2\right] - \frac{1}{\hbar}V \qquad \text{(equation 6.44)}$$

$$\nabla \frac{\partial S}{\partial t} = \nabla \frac{\hbar}{2m}\left[\nabla \nabla R + (\nabla R)^2 - (\nabla S)^2\right] - \frac{1}{\hbar}\nabla V \qquad \text{(multiplying by } \nabla)$$

$$\frac{m}{\hbar}\frac{\partial v}{\partial t} = \nabla \frac{\hbar}{2m}\left[\nabla\left(\frac{m}{\hbar}u\right) + \left(\frac{m}{\hbar}u\right)^2 - \left(\frac{m}{\hbar}v\right)^2\right] - \frac{1}{\hbar}\nabla V \qquad \text{(replacing } \nabla R \text{ and } \nabla S)$$

$$\frac{\partial v}{\partial t} = \nabla \frac{\hbar}{2m}\left[\nabla u + \frac{m}{\hbar}u^2 - \frac{m}{\hbar}v^2\right] - \frac{1}{m}\nabla V \qquad \text{(eliminating } m/\hbar)$$

$$\frac{\partial v}{\partial t} = \frac{\hbar}{2m}\nabla^2 u + \frac{1}{2}\nabla(u^2) - \frac{1}{2}\nabla(v^2) - \frac{1}{m}\nabla V \qquad \text{(equation 6.46)}$$

□

This completes the proof of theorem 6.14. □

6.5 *Dirac equation*

In a previous section, we have used $TdS = Fdx$ to recover $F = ma$. In another section, we have used $TdS = Pdt + Fdx$ to recover special relativity. We have then used a random walk on dx to recover the Schrödinger equation which is the quantum analogue to $F = ma$. Of course, the natural question to ask is, will using $TdS = Pdt + Fdx$ and applying a random walk to both dt and dx be enough to recover the Dirac equation, the quantum analogue to special relativity? The answer is yes!

In this section, we will see that applying a random walk to both the dt and the dx variables is enough to recover the Dirac equation for relativistic quantum mechanics. Let us begin by answering why would there be a random walk on dt.

First we consider that, as is the case with program length, program runtime varies from one UTM to the next. Programs that are difficult to solve on one UTM are likely to be difficult to solve on other UTMs. For example the travelling salesman problem is hard to solve on every UTM. If we take a specific problem, then each UTM of Λ will have a corresponding program to encode it (by virtue of being universal). The runtime of these programs will be randomly distributed and centred around a mean runtime.

Second, we consider an analogous argument to the one used to justify a random walk on dx, but applied to dt. On some UTM a program of size x might have halted and on others it might not have. Therefore a particle can be defined to be at a time t only if a program halting at time t is in the partition function. If there is no such available halting program at time t, then the particle will be a time $t \pm \Delta t$, the runtime of the next available halting program. Since the halting problem is algorithmically random and non-computable, we consider this behaviour as a random walk in time.

A connection between a random walk in time and space and the telegraphic equation has been linked to the Dirac equation before.[12] . D. G. C. McKeon and G. N. Ord proposes a random walk model in space and in time. Starting from the equation for a random walk in space, we have

[12] G. N. Ord D. G. C. McKeon. Time reversal in stochastic processes and the dirac equation. https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.69.3, 1992; and G. N. Ord D. G. C. McKeon. On how the (1 + 1)-dimensional dirac equation arises in classical physics. https://link.springer.com/article/10.1007/BF02190048, 1995

$$P_{\pm}(x, t+\Delta t) = (1 - a\Delta t)P_{\pm}(x \mp \Delta x, t) + a\Delta t P_{\mp}(x \pm \Delta x, t) \tag{6.50}$$

then, D. G. C. McKeon and G. N. Ord extend this equation with a random walk in time. They obtain

$$F_{\pm}(x, t) = (1 - a_L\Delta t - a_R\Delta t)F_{\pm}(x \mp \Delta x, t - \Delta t) + a_{L,R}\Delta t B_{\pm}(x \mp \Delta x, t + \Delta t) + a_{R,L}\Delta t F_{\mp}(x \pm \Delta x, t - \Delta t) \tag{6.51}$$

where $F_{\pm}(x, t)$ is the probability distribution to go forward in time and $B_{\pm}(x, t)$, backward in time. They then introduce a causality condition such that $F_{\pm}(x, t)$ and $B_{\pm}(x, t)$ only depends on probabilities from the past.

$$F_{\pm}(x, t) = B_{\mp}(x \pm \Delta x, t + \Delta t) \tag{6.52}$$

From equation 6.5 and 6.52, they get

$$B_\pm(x,t) = (1 - a_L\Delta t - a_R\Delta t)B_\pm(x \mp \Delta x, t + \Delta t) + a_{L,R}\Delta t B_\mp(x \pm \Delta x, t + \Delta t) + a_{R,L}\Delta t F_\pm(x \mp \Delta x, t - \Delta t) \tag{6.53}$$

In the limit $\Delta x, \Delta t \to 0$ and with $\Delta x = v\Delta t$, they get,

$$\pm v\frac{\partial F_\pm}{\partial x} + \frac{\partial F_\pm}{\partial t} = a_{L,R}(-F_\pm + B_\pm) + a_{R,L}(-F_\pm + F_\mp) \tag{6.54}$$

$$\pm v\frac{\partial B_\mp}{\partial x} + \frac{\partial B_\mp}{\partial t} = a_{L,R}(-B_\mp + F_\mp) + a_{R,L}(-B_\mp + B_\pm) \tag{6.55}$$

Posing these changes of variables,

$$A_\pm = (F_\pm - B_\mp)\exp[(a_L + a_R)t] \tag{6.56}$$

$$\lambda := -a_L + a_R \tag{6.57}$$

then 6.55 becomes

$$v\frac{\partial A_\pm}{\partial x} \pm \frac{\partial A_\pm}{\partial t} = \lambda A_\mp \tag{6.58}$$

Finally, they pose $v = c$, $\lambda = mc^2/\hbar$ and $\psi = F(A_+, A_-)$, they get

$$i\hbar\frac{\partial \psi}{\partial t} = mc^2\sigma_y\psi - ic\hbar\sigma_z\frac{\partial \psi}{\partial x} \tag{6.59}$$

which is the Dirac equation in 1+1 spacetime.

7 *Entropic time*

So far, we have connected an entropic UTM to physics such that:

- Classical physics is obtained by first relating entropy to various program-observables, then solving them with the use of a justified constant temperature.
- Quantum physics is also recoverable provided that we introduce the concept of an observer with limited knowledge of Ω. If we further suppose that his knowledge of Ω is strictly acquired by measurements over some bits of Ω, then we obtain a definition of the quantum measurement and subsequent collapse. Specifically, a measurement happens when the information accessible to the observer is such that a bit becomes computable and a collapse occurs when the UTMs incompatible with the value of the measurement are removed from Λ.

We are now ready to define a directional time. To simplify, we suppose a case where an observer immediately measures all the bits of Ω has soon as their corresponding program halts.

Definition 7.1 (Directional Time)**.** *As t varies, more programs halt. Hence the value of Ω^Z changes as bits flip from 0 to 1. Each value of Ω^Z corresponds to a time slice which is distinct from the others. The halting information of Ω^Z is valid up to the error rate of $2^{-k(t)}$ which monotonically decreases with time. Future time slices contain the halting information of all past slices, but the reverse is not true.*

For example, consider the following time slices, each corresponding to a different value of Ω^Z. As we move down along the different Ω_i^Z, more and more bits are flipped from 0 to 1. Each Ω_i^Z corresponds to the state of the universe at a different time of its history. [13]

[13] Note that this example contains a slight simplification. In the calculation of Ω^Z, halting bits are shifted leftwards along Ω^Z as time increases hence there is a possibility of shifting bits left to right. However, this detail can be ignored for this example.

$$Z_0 = 00000000000000000000000\dots \tag{7.2}$$
$$Z_1 = 00000000000010000000000\dots \tag{7.3}$$
$$Z_2 = 00000010000010000100000\dots \tag{7.4}$$
$$Z_3 = 00000110000010000100000\dots \tag{7.5}$$
$$Z_4 = 00010110010010000100000\dots \tag{7.6}$$
$$Z_5 = 00010110010010001100100\dots \tag{7.7}$$
$$Z_6 = 01110110010010001100110\dots \tag{7.8}$$
$$Z_7 = 11110110010010001100110\dots \tag{7.9}$$
$$\vdots$$

In the case of an observer lacking complete knowledge of Ω^Z, he will also see time slices provided that at least some measurements are made.

How reasonable is the assumption that an observer sees non-computable bits? We will now define rigorously what we mean by an observer.

Definition 7.10 (Observer)**.** *An observer is defined as a series of measurement-sets. The sets represent the partial knowledge of Ω known to the observer. The elements of the set can change with time. An observer that exists over a period of time must define a measurement set for each time slice. As an example, an observer could be defined as*

$$O_{(t=1)} = \{\omega_1 = 0\}$$
$$O_{(t=2)} = \{\omega_1 = 0, \omega_2 = 1\}$$
$$O_{(t=3)} = \{\omega_1 = 1, \omega_3 = 1\}$$

A special case occurs when an observer knows all the bits of Ω for all time slices. We recover the UTM corresponding to the universe. We will call this observer the universal observer. Unless otherwise stated, when we use the word observer, we exclude the universal observer.

Theorem 7.11. *The Kolmogorov complexity of an observer at a given time t is less than or equal to the number of elements in the set of its measurements.*

$$\mathrm{Kov}(O_t) \leq |O_t|$$

, where the double vertical bars indicate the size of the set.

Proof. The bits identified by O_t are produced by the partition function of the halting probability whose definition requires knowledge of the halting probability of a UTM. Hence it is a non-computable function. Therefore, in the worse case scenario, the bits identified by O_t are non-computable. □

Theorem 7.12. *For all non-universal observers, the future is non-computable.*

Proof. The partition function dovetails programs. As a result, it is possible to imagine a simple algorithm -the dovetailing algorithm- that, knowing the set of axioms k of the universe, is able to calculate future time slices before they occur.

To calculate future time slices as an experiment, the observer must encode k in a finite state machine and run the program until it calculates a time slice newer than his.

But in the case of a non-universal observer, this cannot be done. Indeed, k represents the halting probability of the UTM, its Kolmogorov complexity will always be higher than the Kolmogorov complexity of an observer, defined as a subset of k. Therefore, the observer has insufficient bits to express k, required to compute the future. □

Theorem 7.13 (Arrow of time)**.** *For a non-universal observer that does not erase bits, directional time is asymmetric.*

$$\forall t \forall t' [(t < t') \rightarrow \mathrm{Kov}(O_t) \leq \mathrm{Kov}(O_{t'})]$$

Proof. If no bits are erased, future $O_{t'}$ will contain all the bits of past O_t and then some more. Hence an observer can compute O_t from $O_{t'}$. □

Remark 7.14. *The physical observer can compute approximations of his future. For example, it can determine that the probability of raining tomorrow is 10%. This is not computing the future, it is computing a probability.*

Theorem 7.15 (Fixed past)**.** *The entropy of measured bits, post-measurement, is always* 0*. Hence, there is no possibility of changing the past.*

Proof. The entropy is defined as

$$S = k_b \sum_i p_i \ln p_i \tag{7.16}$$

Post measurement, there is only one state defining the observer, the set O_t. Hence $p_i = 1$ and $S = 0$. The past time-slice experienced by an observer (who is not erasing measurement bits) is always fixed to a single possibility. □

Theorem 7.17. *The measurement of a bit of* Ω *by an observer must be accompanied by an increase in entropy in the rest of the system.*

Proof. Each measurement of a bit of Ω decreases the entropy of the past of the observer by $k_b T \ln 2$. As a result, a quantum measurement must produce the same amount of entropy in the system, such as in the measuring apparatus. □

7.1 *Quantum measurement*

We suggest an interpretation of the quantum measurement such that it is connected to the non-computability of the bits of Ω. A measurement occurs when a bit of Ω becomes computable (or known) by the observer. When it happens, any UTM incompatible with the measured value are removed from Λ - this is the collapse.

To investigate this further, let us look at what happens when a bit of Ω becomes computable by the observer. As an example, let us take the case where $\omega_3 = 1$.

First, Ω becomes $\Omega_{(\omega_3=1)}$,

$$\Omega_{(\omega_3=1)} = 0.\omega_1\omega_2 1\omega_4 \ldots \tag{7.18}$$

Second, the entropy decreases from $S = k_b N \ln 2$ to

$$S = k_b(N-1)\ln 2 \tag{7.19}$$

Third, incompatible UTMs are removed from Λ and it becomes $\Lambda_{(\omega_3=1)}$,

$$\Lambda_{(\omega_3=1)} := \{U | isUTM(U) \wedge (\omega_3 = 1)\} \tag{7.20}$$

Pre-measurement, ω_3 is non-computable for the observer. Hence its post-measurement value will appear random as per the normality of Ω. But, we have claimed that ω_3 is measured by an observer once and only if its value becomes computable. How do we reconcile? Information sufficient as to deduce the value of ω_3, post-facto, must be made available to the system simultaneously as the measurement is done. The transition from unmeasured to measured occurs when the bit goes from non-computable to computable, for the observer. Physically, this information could correspond to a macroscopic measurement system absorbing the entropy associated with the measurement.

Borrowing terms from physics, we could say that we are performing a measurement on ω_i and, as a result of this measurement, Λ collapses to $\Lambda_{(\omega_3=1)}$. We note a similarity with what is described here and the quantum measurement.

Remark 7.21. *If we had defined a specific UTM to study, this would not have been possible. Rather than being simply an observer, we could have pre-calculated the values of Ω by running the UTM independently. This would have been enough to pre-calculate ω_i and avoid performing a measurement. In that case, the similarity with the quantum measurement we have just investigated would not have been possible.*

8 Conclusion

We note an affinity between an entropic UTM and the laws of physics. Understanding physics from the perspective of the entropic UTM holds several conceptual advantages. For one, the laws of physics are unified in a single equation; the halting partition (equation 3.9). The affinity occurs when we consider a UTM calculating its Ω number in a manner so as to maximize the entropy throughout the calculation (theorem 3.11 and 3.19).

From this halting partition, we recover special relativity (speed of light: 4.2, light-cones 4.2 and the Lorentz's factor 4.20). Going further, we are hit by the first hesitation. Does the universe physically encode program bits in one dimension, two dimensions or three dimensions? It is not immediately clear which one the universe prefers, if any. However, when we consider that all three possibility do occur, we recover in one dimension the law of inertia, in two dimensions general relativity and in three dimensions possibly dark energy. This corresponds to three characteristic scales of the universe each dominated by a specific entropy distribution. The local scale is dominated by inertia (theorem 5.4), the galactic scale is dominated by gravity (theorem 5.8) and the cosmological scale is dominated by dark energy (equation 5.9).

As for quantum mechanics, it can be recovered simply by extending the non-computable effects of the UTM to the dx and dt variables. Considering a random walk in dx yields the Schrödinger equation (theorem 6.14) and extending this random walk to dt yields the Dirac equation (equation 6.59). Furthermore, the spin (theorem 6.1) and the polarization (theorem 6.10) can both be recovered as a thermodynamic ensemble respectively by injecting the $SO(3)$ rotation and the $SO(2)$ rotation as macroscopic observable acting on E. We note that the operation is limited to a maximum of three dimensions so as to permit the matrix multiplication with a two-state system.

For entropic time, we have shown that the entropic UTM perspective allows us to define time in terms of slices of progressively increasing halting information (definition 7.1). This allows us to prove that, unlike past time slices, future time slices are non-computable (7.12). Furthermore, the entropy of past time slices is 0 which suggests that the past cannot be changed, while the entropy of future time slices is greater than 0. This would suggest an arrow of time (7.13 and 7.15).

The quantum measurement is linked to an observer with a Kolmogorov complexity too small to run the calculation of Ω. His knowledge of Ω is acquired by performing measurements over the bits Ω. Each value acquired reduces the space of possible UTM compatible with this knowledge. This was connected to the quantum mechanical collapse in section 7.1. As the bits of Ω are algorithmically random, the value of the collapse is non-computable.

As a reference, I have previously suggested in another paper[14] that the halting probability augmented with thermodynamic-like observables can recover most of the laws of physics.

[14] Alexandre Harvey-Tremblay. An axiomless derivation of the theory of everything. https://www.academia.edu/33079029/An_axiomless_derivation_of_the_Theory_of_Everything, 2017

References

John C. Baez and Mike Stay. Algorithmic thermodynamics. arXiv:1010.2067 [math-ph], 2010.

Gregory Chaitin. Mathematics, complexity and philosophy. https://www.academia.edu/31320410/Mathematics_Complexity_and_Philosophy_full_bilingual_text_, 2010.

Gregory J. Chaitin. An algebraic equation for the halting probability. https://www.cs.auckland.ac.nz/~chaitin/berlin.pdf, 1988.

Gregory J. Chaitin. How real are real numbers? http://www.worldscientific.com/doi/abs/10.1142/S0218127406015726 and https://www.cs.auckland.ac.nz/~chaitin/olympia.pdf, 2006.

G. N. Ord D. G. C. McKeon. Time reversal in stochastic processes and the dirac equation. https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.69.3, 1992.

G. N. Ord D. G. C. McKeon. On how the (1 + 1)-dimensional dirac equation arises in classical physics. https://link.springer.com/article/10.1007/BF02190048, 1995.

Alexandre Harvey-Tremblay. An axiomless derivation of the theory of everything. https://www.academia.edu/33079029/An_axiomless_derivation_of_the_Theory_of_Everything, 2017.

Ming Li and Paul Vitányi. An introduction to kolmogorov complexity and its applications. Springer, 1997.

Edward Nelson. Derivation of schrödinger's equation from newtonian mechanics. http://dieumsnh.qfb.umich.mx/archivoshistoricosMQ/ModernaHist/Nelson%20a.pdf, 1966. Physical Review, Volume 150, Number 4.

K. Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. http://arxiv.org/abs/nlin/0212001, 2002.

K. Tadaki. A statistical mechanical interpretation of algorithmic information theory. https://arxiv.org/pdf/0801.4194.pdf, 2008.

Erik Verlinde. On the origin of gravity and the laws of newton. arXiv:1001.0785v1 [hep-th], 2010.

Erik Verlinde. Emergent gravity and the dark universe. https://arxiv.org/pdf/1611.02269.pdf, 2016.