Information and Physics

Serge Mijatovic, M.Sc. Computer Science
e-mail sergemij@yahoo.com

What would the Universe look like if information processing was at its very core? What is the most likely and optimum way of this information use? This paper explores a fundamental scenario of a possible connection between information and physics. Since the approach taken is antecedent to first principles of physics, we will rely only on axiomatic notions of information use, and deliberately ignore the body of physics to avoid conclusions isomorphic to it. Essential relativistic, gravitational and quantum phenomena are derived as limiting cases, including time dilation effect of Einstein's relativity theories as a special limiting case.

1. Introduction

Fundamental physical laws are presently considered the first-order concept in explaining physical phenomena. This paper examines the alternative possibility that physical laws emerge as a second order consequence of use of information. A new kind of information is introduced on a level prior to that of observable one. Use of this *fundamental information* by physical constituents is given to be the sole cause of physical patterns we perceive as physical laws. To avoid dangers of circular reasoning, virtually no premises, postulates or conclusions of the body of physics will be used. This paper will focus on the important aspect of information use: its throughput. The results in limiting cases will be apparently what is known as relativistic, gravitational and quantum effects, with some interesting predictions.

We do not know of any form of decision-making without the use of information. Generally, a decision of any kind can lead to a specific action only if information was used, and conversely any decision made without the use of information can only lead to an arbitrary action.

We have no reason to believe that Nature can escape this conundrum of *having to use information to function in a non-arbitrary way*. The goal of this paper is to explore the effect of throughput of information on the behavior of basic constituents of physical reality. These explorations begin and end before the law-based physics even starts with its core postulates and principles.

The theory here developed is henceforth called the Fundamental Information Theory, or FIT for short. Because FIT precedes existing frameworks, we start with a simple setup of a flat 3D Euclidean space and not a 4-D Minkowski space. For the same reason, time is considered constant, addition of velocities is Galilean and relativistic conclusions do not apply. Surprisingly, this setup does not contradict the body of evidence related to what is known as relativistic effects. ***FIT is an information theory alone with only aforementioned physical assumptions.***

Special Relativity Theory (SRT) and General Relativity Theory (GRT) essentially derive from two approaches: one is a heuristic approach of accepting certain non-trivial premises as universal principles (*i.e.* related to the speed of light, relativity & equivalence principle), and the other is the development of conceptually brand new aspects of reality (*i.e.* space-time, curvature). Quantum phenomena generally start with the heuristics of its own (i.e. principle of uncertainty).

FIT derives from the well-known generic concept of information and the requirements for its use. It does not need any heuristic postulates.

2. Types of Information

Information is often said to describe something. In physics, the concept of information is related to facts that are (or could be) obtained by observation. We will call this *discernible information*.

When a charged particle moves in an electric field, we could say that, to an observer, its observed location and velocity is information. What is largely left out of this picture is the (seemingly nonsensical) question of *what is the information to an electron?* How does an electron know how to behave? Simply saying that it follows a 'law' is a tautology (or circular reasoning), even if the pattern of such behavior can be explained by adherence to a law. The premise here is that an electron moves because of the information that is available to it at every location in space. Only having and using this new kind of information can provide the basis for physical effects. This idea applies to any physical process, an electron movement being just an analogy.

The difference between concepts of a field moving electrons (on one hand) and an electron moving itself by means of information use (on the other hand) is fundamental. If an electron moves because all it has is the information available to it wherever it is, then this movement is purely the result of using this information and nothing else. The information itself comes from all the constituents comprising the Universe, including the electron itself. The information thus changes at all times, coinciding with movements of all such constituents. *That new kind of information* is what drives all physical processes and we call it *fundamental information*. Physical processes are thus supposed not to be driven by *the adherence to a law per se (which in many ways is an anthropomorphic concept borrowed from a notion of social government), even though patterns of information use may produce an appearance of one.* The discernible information is just a small subset of ever-changing *fundamental information* that is used constantly.

The notion of fundamental information and its usage does not relieve the explanation for behavioral patterns from the eventual use of new laws that would be applied to this idea. The goal is not to eliminate a notion of a predetermined law. The goal is to suggest a deeper explanation for foundational patterns in physical systems, as opposed to an answer by a proclamation of a law (regardless of its appeal or predictive value).

The concept of fundamental information will be used going forward, and it will be referred to simply as 'information'.

3. A Model for Information Use

We will present the most likely generic model of usage of information without presupposing *how it works*. The absence of *how* and the necessity of *what* such model should describe is why such model can be considered *generic*. The model rests on axiomatic principles of information science, simplicity and the elementary requirements of information use. The term 'use' as employed here is a broad one and is intentionally independent of any particular method of use or its physical form.

Information is a set of facts. To use information there has to be more than one set of facts. In a simplest scenario, there are two sets of facts (with A and B facts). In this case, *use of information* means combining facts from both sets (to account for all the facts). So the number of fact pairs being used is $A \times B$. We will assume that all fact pairs take the same finite time to process.

3.1 The Physical Embodiment of Information

We assume that there is a physical entity that can take facts as an input, store them, and then use them. We call this entity *a computer,* and we call the usage of information *processing (or computing)* with *a clear disclaimer* that this notation does not imply any particular physical embodiment, encoding, or method.

A computer possesses some *inherent* information that describes it (i.e. some facts about it that do not change in time). Computer also has facts that describe its *state* (i.e. some facts about it that may change in time). Inherent and state facts together comprise *self-information* that fully describes a computer.

We will assume a computer to have limited information storage just as we assumed that the processing of facts takes finite time. Every computer is given the same storage for information. We assume that the number of computers is unchanging (i.e. they are neither created nor destroyed).

Imagine a physical representation of a computer to be a tiny finite sphere, so small it can be considered a dot. This sphere contains self-information and we assume this information to be on the surface of the sphere.

3.2 Multiple Computers

Two of such computers are at some distance. In this scenario, presumably each computer would somehow obtain the self-information from the other in order for any meaningful information processing to take place. To understand what this method of obtaining information must be, consider the same scenario on a much smaller scale. The computers now *look like* large spheres and the distance between them has similarly increased. This is still the same system, only observed on a different scale.

Now consider the original system altered so that spheres *are indeed* much larger, and the distance between them is (equally) larger as well. The observer of this system has not changed scale. The system is really larger. Or is it? The difference is apparent only from the perspective of the observer however it is not from the perspective of the computers. The two scenarios are equivalent for the computers. We can imagine a point in space far from a computer and then enlarge the whole system so this point is now on the surface of the sphere that contains the computer's information. The consequence is that the same information has to exist on virtually any sphere around the computer (i.e. *computers have no preference for scale*). It follows that the density of this information is declining the same way the surface of the sphere does, i.e. with the *square of distance*. We will assume that information has no preference for direction either, so the information at some distance R from a computer is:

$$i_R = a \times i / R^2$$

i_R is the average number of facts at distance R from the computer in some small volume of space; R is the distance from the computer; i is the self-information of the computer; a is a dimensional constant. For simplicity, we henceforth omit a and use only the dimensionless value R in further text. Since computer's own information has to be available to itself, the minimum value for R has to be the unit length.

The method of communicating information between computers is thus a trivial one. A computer *is centered in a specific point in space but its self- information occupies all of the space, its density declining with the square of distance.* This distribution of information due to a single computer is called *information field.* A computer *simply uses* the information available from information field due to all computers.

3.3 Time and Space

Since we assume the throughput of information use to be finite there is a delay between the results of computation. Thus whatever the result of computations is, such result is a discrete event in time.

The *information field has to change over time with some frequency.* This is because self-facts are distributed randomly on any given sphere around a computer and if any particular distribution of facts remained fixed it would be a preferred distribution of facts in space (i.e. there would be preferred directions in space). This is called *reloading of information field.*

We will assume that *if* information available to a computer is larger than its storage, then any fact has equal chance to be used by a computer.

Use of information consists of gathering available information and then processing it. We will not assume any parallelism in processing of a single computer as it is not simple and would not be characteristic of an *elementary* information processing entity (separate computers use information entirely in parallel).

We will, however, assume that gathering of local information and processing occur in parallel. The two are trivial to synchronize because both begin and end at the same time. Therefore the time to take in the available information is exactly the time it takes to process the available information collected one computation before. We call this a *processing cycle.* We will assume that a computer will attempt to use all the available information during the processing cycle.

At the beginning of the processing cycle, the information taken in fills the storage of a computer. In a Universe made of a single computer, its self-information should fill its own storage. If it didn't, the excess storage would be wasted, and at the same time it should be large enough at least for its own information. We will assume that the information storage of a computer is determined by its self-information.

3.4 Computer Memory (Persistence of Information)

A computer must have a memory. Without memory, anything that happens now would not in any way depend on what came before. The simplest memory is in the form of *previous and current information.* The previous information is the current one from the previous processing cycle. The two sets of facts are combined and this constitutes a computation. The storage for the two sets should be the same, as they are conceptually the same information as seen by a computer in two points in time.

After initial take-in of information, the available information may change during the processing cycle. It may change for two reasons. The first reason is because information field changes over time, i.e. due to reloading of information field so a computer sees new facts that were not there at the beginning of the cycle. We call this additional information the *R-information*. The second reason is because relative movement can expose a computer to additional information that is not present at its current location but *is* present at a different nearby location through which computer moves. Apparently this additional information is proportional to the relative speed of computers because moving twice as fast lets a computer visit twice as many locations and obtain twice as information from the information field. We will call this additional information the *M-information*. The *R* and *M* information are called the *additional information*.

A computer must have limited temporary information storage for additional information. This storage cannot be used for processing of information because it would only make the two processing sets larger to a point where there are no limits.

Since the current set gets filled (in the initial information collecting), due to limited storage, it is possible to use the additional information only if some information in current or previous set is lost. For the purpose of FIT, we will assume the additional information would be stored in the previous set and not the current (either choice leads to the same conclusions in this paper). We will assume that the previous information that is about to be lost will be processed to minimize its loss (i.e. compressed into remaining storage). Similarly as before, the cost to process this previous information of size A would be $A \times A$, meaning that it is essentially combined with the information of equal size. The information is lost in this process, *i.e.* it is a compression. The current set will now be larger than the previous set due to the influx of additional information.

3.5 Information Throughput

If a computer has A facts, then by using previous and current set of facts, it combines $A \times A$ pairs of facts. The actual throughput of current facts is A, because that is the influx of information during the processing cycle. This means that the throughput of facts is a square root of the number of fact pairs processed by a computer (*i.e.* $T_A = \sqrt{A \times A}$). So if a computer had two sets of different sizes (A and B as in the case when there is additional information), then the throughput is $T_{AB} = \sqrt{A \times B}$. This is so because there is no reason to believe that pairs of facts produced carry any information about *how* they were produced. A set of 36 facts can be produced by pairing two sets of sizes 6 and 6, but it can also be produced by pairing two sets of sizes 2 and 18. *Two identical fact sets cannot be distinguished from one another.* It means that they must produce the same information throughput, the same one when both sets are equal (i.e. symmetrical in the size of information content).

Repeated use of the same two sets of facts produces exactly the same result. This generally does not necessitate deterministic computing *per se* because there may be a loss of information due to the additional information and limited storage.

It is conceivable that the loss of previous information could go as far as losing it all (due to enough of additional information). We will assume that no computer would let lose all of its previous information because that would mean that the previous set of facts would vanish and the computation would stop. We assume the computation is a ubiquitous quality. This means a computer would perform an action to prevent the loss of the whole previous set.

3.6 Physical Evolution as Computation

For information use to have any physical meaning, the result of its computation *must have a physical manifestation*. In addition, this result *must cause change in information use*, i.e. it must provide the *feedback to computation.* Without feedback, a physical manifestation would be meaningless as it would not have any effect, and by definition would not be a physical change. The simplest effect that fits both criteria is *change in velocity, i.e. acceleration.* It provides a measurable effect (*i.e.* a physical manifestation), and at the same time recall that the additional information changes with relative velocity (*i.e.* there is a feedback to computation). By adopting the simplest scenario, we will assume that *the only physical consequence of computation is acceleration.* This includes a null result of computation, i.e. uniform motion.

We assume all physical effects to be the result of information use. Thus we have two distinct mechanisms by which reality unfolds: one of processing information (*computing the acceleration)*, and the other of actually *performing the acceleration.*

Because a computation is pairing of two finite sets of facts, its result is a number of limited precision, i.e. it is an integer. It means at fundamental level of a single computer *the acceleration is quantized.*

We can reasonably assume that both the resources for computation and acceleration are limited. However, while computation is performed at all times, acceleration is not. We will assume a computer would act to preserve the acceleration resources (*i.e.* to minimize the *acceleration cost*) as part of its computation. Because such resources are finite, if being in different nearby locations at the same velocity uses different amounts of acceleration resources, we assume that any computer would move to minimize acceleration cost.

The method of information usage can be arbitrary, so long as the set of above assumptions is satisfied. The usage of fundamental information is independent of any underlying mechanism, so none are discussed here. It is implied that all computers behave the same since the difference would imply more complexity and preferences for some computers.

4. Information Influence Decline over Space

The available information at a location of a computer k is:

$$I_k = \sum_{j=1}^{U} i_j \Big/ R_{jk}^2$$

where U is the number of all computers; i_j is the self-information of the computer j; R_{jk} is the distance to computer j. For itself, the distance R_{kk} is the unit since all of k's information is available to itself.

Storage of computer k is generally smaller than available information I_k, and any fact has equal chance to be used. So the percentage of k's storage allotted for available information of computer m is:

$$f_{mk} = \left(i_m \Big/ R_{mk}^2\right) \Big/ \sum_{j=1}^{U} i_j \Big/ R_{jk}^2$$

This percentage is called *information influence* of m at k. Apparently, the sum of all influences on a computer is 1 (or 100%):

$$\sum_{j=1}^{U} f_{jk} = 1$$

A number of distant computers grouped closely together is called a *cluster computer*. A single (non-clustered) computer is a *standalone computer*. A distant cluster can be *approximated as a standalone computer* represented by a sum of information of the cluster *for the purpose of cluster's available information*:

$$I_k \approx \frac{\sum_{j=1}^{U} i_j}{R^2}, \text{ for } R_{jk} \approx R, \forall j, j \neq k$$

Each computer from a cluster would process the information of a distant computer slower than if it were alone. For example a computer k in a cluster of U computers would process the following portion of distant computer's information i_d (assuming i_0 is the constant information of any standalone computer and r is some distance used to approximate all distances in the cluster where $r \ll R_{dk}$, R_{dk} is the distance from cluster to a distant computer d):

$$f_{dk} = \frac{i_d \Big/ R_{dk}^2}{\sum_{j=1}^{U} i_j / r_{jk}^2 + i_d \Big/ R_{dk}^2} \approx \frac{i_d \Big/ R_{dk}^2}{U \times i_0 / r^2 + i_d \Big/ R_{dk}^2}$$

In case of two different clusters (with U_1 and U_2 computers), the ratio of above influences is (as cluster's influence is much higher than of a distant computer $i_d \Big/ R_{dk}^2 \ll U \times i_0 / r^2$):

$$\frac{f_{dk1}}{f_{dk2}} \approx \frac{i_d \Big/ R_{dk}^2 \times (U_2 \times i_0 / r^2 + i_d \Big/ R_{dk}^2)}{i_d \Big/ R_{dk}^2 \times (U_1 \times i_0 / r^2 + i_d \Big/ R_{dk}^2)} \approx \frac{U_2}{U_1}$$

This means the influence of a distant computer on any member of a cluster is reversely proportional to the size of a cluster.

In N processing cycles (where N is statistically large) the number of facts used from any computer m is approximately:

$$N \times i_k \times f_{mk}$$

Since number of facts is an integer, the above number has to amount to an integer for computer m to contribute any facts. It also means that at any given time, a number of computers *will not have any direct influence at all on computer k and as such are effectively non-existent.* A group of computers that have nonzero influence is called the *constraint group* because they are the only ones that affect (constrain) the behavior of a computer.

If m alone comprises the constraint group, then:

$$f_{mk} \approx 1 \Big/ \left[1 + (i_k / i_m) \times R_{mk}^2 \Big/ R_{kk}^2\right]$$

If m is large and close, then $f_{mk} \approx 1$. When m is small and far away, then $f_{mk} \approx 0$. The small computers are *overwhelmed* by the facts from nearby large computers (and *vice versa*).

5. The Effect of Motion Through Space

If (a computer) n moves relative to m, the number of locations visited in a given period of time will be proportional to its speed, and so will be the amount of additional M-information. A nearby computer q at rest relative to m will not see this additional information.

Thus the change of n's data is proportional to its speed relative to m:

$$\Delta i_n / i_n = s \times v_{mn} \times f_{mn} \tag{1}$$

where Δi_n is the change in current set at n resulting from the movement relative to m; i_n is the size of n's current set; s is a dimensional constant of proportion; f_{mn} is the information-influence of m at n; v_{mn} is a relative speed achieved where f_{mn} can be considered constant.

The exact equation for Δi_n would include all the computers. When adding the additional information from all computers, the velocities can be added as vectors in Cartesian coordinate system with the exception that vector components are added as absolute values (because M-information depends on relative speed only). These components in the direction of x, y and z axis are:

$$V_n^x = \sum_{j=1}^{U} \left|v_{jn}^x\right| \times f_{jn}$$

$$V_n^y = \sum_{j=1}^{U} \left|v_{jn}^y\right| \times f_{jn}$$

$$V_n^z = \sum_{j=1}^{U} \left|v_{jn}^z\right| \times f_{jn}$$

where $\left|v_{jn}^x\right|$, $\left|v_{jn}^y\right|$ and $\left|v_{jn}^z\right|$ are the absolute values of relative velocities between computers j and n in the direction of x, y and z coordinate axis, respectively. A vector described by (V_n^x, V_n^y, V_n^z) is the *M- information vector of computer* n.

The exact equation for M-information is then:

$$\Delta i_n = s \times i_n \times V_n \quad (2)$$

where $V_n = \sqrt{(V_n^x)^2 + (V_n^y)^2 + (V_n^z)^2}$ is the M-*information speed* (*i.e.* absolute value of M- information vector). Note that Δi_n cannot become greater than i_n due to limited information storage:

$$\Delta i_n \le i_n$$

6. The Effect of Time Passing

6.1 On the Storage of Information

The previous and current information sets are combined together to compute the result. The previous set is the current set from the previous use of information:

$$i_{\text{previous}}(t) = i_{\text{current}}(t - \Delta t)$$

Between the two, they can fill a fixed total storage capacity:

$$i_{\text{previous}} + i_{\text{current}} = \text{constant}$$

Δi denotes additional information to fit into storage:

$$(i_{\text{previous}} - \Delta i) + (i_{\text{current}} + \Delta i) = \text{constant}$$

6.2 On Processing Requirements

The duration of computing previous and current sets in the lossless case (when there is no additional information; *i.e.* $\Delta i = 0$) is proportional to:

$$H_0 = i \times i = i^2$$

In a case when there is additional information ($\Delta i \ne 0$), H is:

$$H = (i - \Delta i) \times (i + \Delta i) = i^2 - (\Delta i)^2$$

Apparently it is $H < H_0$. When $\Delta i > 0$, there is more information than there is storage. In this case, to minimize the loss of information, the previous set will be compressed to occupy less storage. This compression is denoted as a lossy transformation Q. The time spent for this compression is not used in computing the change – it only produces a lossy version of the same information, thus it can be considered 'not useful'. For the reasons outlined in this paragraph, the quantities H and H_0 will be referred to as 'useful'.

The compression Q effectively computes information Δi with equal information, so its duration is proportional to:

$$H_Q = (\Delta i) \times (\Delta i) = (\Delta i)^2$$

The total is:

$$H_1 = H + H_Q = (i - \Delta i) \times (i + \Delta i) + (\Delta i)^2 = i^2$$

And we have:

$$H_0 = H_1$$

This result means that the total duration is the same both in lossless and lossy case. *The processing cycle always takes the same time.* However in the case of information loss, the amount of useful computing is less.

6.3 On Processing Throughput

The computational throughput T is the amount of useful information produced per unit of time. In the lossless case,

$$T_{\text{lossless}} = i / \Delta t$$

where Δt is the time needed to process information i.

The useful throughput of processing information in the lossy case would be quantified with:

$$T_{\text{lossy}} = \sqrt{(i - \Delta i) \times (i + \Delta i)} \big/ \Delta t = \sqrt{i^2 - (\Delta i)^2} \big/ \Delta t$$

or:

$$T_{\text{lossy}} = T_{\text{lossless}} \times \sqrt{1 - (\Delta i)^2 / i^2}$$

What T means is that, for example, if the size of both information sets together is 20, and the size of each set is 10 (so $10 + 10 = 20$), the amount of useful information processed would be a square root of 10×10, or 10 per unit of time. Due to additional information, if the current set size is 11 and the previous size is 9 (so $11 + 9 = 20$), then the amount of useful information processed would be a square root of 11×9, or approximately 9.95 per unit of time. When the amount of available information increases, the lower throughput of useful information is the direct consequence of the limited information storage. When the amount of available information decreases, the throughput is higher, with the highest throughput achieved when a computer is far enough from other computers, in which case its own data is all that's available to it. In further text, the term 'information' will have a connotation of useful information.

Physical interpretation: Rate of physical processes changes the same way as the throughput of information. This rate falls to nearly zero when additional information accumulated during processing cycle approaches the storage limit. This is probably the most notable conclusion of FIT. Physical processes *are solely* *based on information use* and throughput of information use *is the same as* the rate of physical processes.

7. Local Speed Limit

7.1 A Speed Limit Near Large Computers

The dimensional constant s has a meaning beyond just being a constant of proportion. In dimensional analysis it has to be an inverse of speed. Consider when speed v_c is such that the throughput of computation is nearly 0:

$$T_{\text{lossy}} = \sqrt{i_n^2 - (\Delta i_n)^2} \Big/ \Delta t \approx 0 / \Delta t = 0$$

i.e. it must be $i_n \approx \Delta i_n$. This speed then locally becomes the highest attainable relative speed of a computer. In a system of two isolated computers m and n this is equivalent to (from Eq. (1)):

$$s = 1 \big/ (v_{mn} \times f_{mn})$$

If m is much larger then n, then $f_{mn} \approx 1$. We will denote this maximum local speed v_{mn} near large isolated computer simply as c. From above two equations, in this particular case:

$$s = 1 / c \tag{3}$$

7.2 Speed Limit in the General Case

For any computer n, there is an exact value for maximum relative speeds c_{jn} (relative to every other computer j from its constraint group) in a given location at a given time. From Equations (1) and (2)] (when $i_n \approx \Delta i_n$) we can find the maximum M-information speed V_n from:

$$\sqrt{(V_n^x)^2 + (V_n^y)^2 + (V_n^z)^2} = c \tag{4}$$

This means that *maximum M-information speed is constant and equal c.* However, the maximum (spatial) speeds (relative to other computers) can vary and depend on information influences.

In a limiting case, in a system of two isolated computers (the other computer being a large computer m) [from Eq. (1)], above reduces to:

$$c_{mn} = 1 / (s \times f_{mn})$$

As f_{mn} can vary between 0 and 1, the maximum speed c_{mn} can vary too, depending on the location:

$$c = 1/(s \times 1) < c_{mn} < 1/(s \times 0) \to \infty$$

Near large computer m, maximum relative speed of n (*i.e.* c_{mn}) has its minimum value of c, while sufficiently far away from m it can approach infinity. In addition, the larger the computer n is (which means the smaller the influence f_{mn}), the higher the speed limit for it will be.

Clearly if a computer moves at its maximum speed relative to nearby large computer m (where its information influence its practically 1), it will *always have the same speed limit* c_{mn} *relative to* m, *regardless of what its initial speed was.*

In order to always use both its current and previous set, a computer must change its velocity relative to other computers if it enters a locale where it is moving beyond its local maximum velocity. A special kind of a computer is one with a purpose of always moving near the maximum local speed. Such computer (we will call it *a 'ping'*) would be useful to allow propagation of information influence from one location to another (be it as a standalone computer moving from one point to another or in a medium comprised of many computers where there is the propagation of a wave). Another use for a ping is to remove information influence from a given location in a most expedient manner possible. A ping always moves at the maximum local speed, meaning its *speed generally changes depending on the local speed limit, per Eq. (4).* Exchanging a ping can be either in form of a direct exchange or in form of a wave.

A computer must be physically a sphere. If it were not, the additional information would be different when moving in different directions in a uniform information field. A computer would reach the speed c when the additional information becomes equal to its storage. In terms of speed that would happen when computer moves so fast that at the end of its processing cycle it no longer occupies any space it occupied at the beginning of processing cycle (i.e. during processing cycle a computer has virtually doubled the input). If the physical length of a computer is L_c and the duration of processing cycle is t_c, then:

$$c = L_c / t_c$$

In other words, the speed limit is a ratio of the physical length of a computer and the time it takes to complete a processing cycle in a given locale. Speed limit has nothing to do with an imposed maximum speed of relative movement but rather with the physical characteristic of a computer. The maximum local speed is *an indirect result of these physical characteristics of a computer and not a postulated property.*

Physical interpretation: While SRT *postulates* the notion of constancy of light speed from which there is a speed limit, FIT *derives* that there has to be one.

8. Information Throughput and Time

8.1 Information Throughput and Motion

Let us say there is an *application* running on a computer (meaning the information processing that can lead to change). A computer always spends the same time on a processing cycle. An application though will work faster or slower, depending on the throughput of processing useful information. Let us formalize how much will this throughput change in general.

We will call the time measured by an application the *application time*. The actual time for which computer will run will be the *real time*.

The throughput of computation T in real–time can vary:

$$T_1(t) = \sqrt{i^2 - (\Delta i_1)^2} \Big/ t \quad , \quad T_2(t) = \sqrt{i^2 - (\Delta i_2)^2} \Big/ t \quad ,$$
$$T_1(t) \neq T_2(t) \quad .$$

Let dt_1 and dt_2 be small increments of application time measured by an application at two different moments in real-time. The application throughput measured in terms of application-time must be the same:

$$T_1(dt_1) = T_2(dt_2)$$

We have

$$\sqrt{i^2 - (\Delta i_1)^2} \Big/ dt_1 = \sqrt{i^2 - (\Delta i_2)^2} \Big/ dt_2$$

and

$$dt_1 = dt_2 \times \sqrt{i^2 - (\Delta i_1)^2} \Big/ \sqrt{i^2 - (\Delta i_2)^2}$$

From this and Eq. (2) we have

$$dt_1 / dt_2 = \sqrt{\frac{1 - s^2 \times V_1^2}{1 - s^2 \times V_2^2}} = \sqrt{\frac{1 - V_1^2 / c^2}{1 - V_2^2 / c^2}} \tag{5}$$

This represents the general transformation of application-time, where: dt_1 is the small application time interval when M-information speed is V_1, dt_2 is the small application time interval with M-information speed of V_2.

The conclusion is that application-times of computers differ when in motion relative to other computers.

Physical interpretation: The transformation of application times (i.e. times as measured in a moving system) has a form that is reminiscent of SRT. The transformation in FIT however describes the *slowdown of rate of physical processes and not the time dilation*.

8.2 Limiting Case for Velocity

Let us consider a situation *of a small moving computer* n *near large isolated computer* m. The information-influence f_{mn} is nearly unity, and the information-influence of all other computers is nearly zero.

Let us have t_1 and t_2 such that the two computers are at rest ($v_1 = 0$) for a unit of application-time t_1, and the relative speed is uniform v_2 for a unit of application-time t_2:

$$f_{mn} \approx 1 \,,\, f_{jn} \approx 0, \forall j, j \neq m, v_1 = 0, v_2 = v \neq 0$$

For a small computer, from Eq. (5):

$$t_1 \approx t_2 \times \sqrt{\frac{1 - 0^2 / c^2}{1 - (v+0)^2 / c^2}} = t_2 \Big/ \sqrt{1 - v^2 / c^2} \quad . \tag{6}$$

Application time runs slower for a small computer n when moving at speed v. We call this a *performance hit* due to the *motion effect.*

For a large computer m, we will have:

$$f_{nm} \approx 0,\, f_{jm} \approx 0,\, \forall j,\, j \neq m,\, v_1 = 0,\, v_2 \neq 0$$

For a large computer M:

$$t_1 \approx t_2 \times \sqrt{(1 - 0^2 / c^2) \big/ (1 - 0^2 / c^2)} = t_2$$

A small computer will run slower, but a large computer will practically not slow-down.

In principle, it is impossible to know exact information-influences of all other computers, because those are statistical in nature. However, near large computers, the information-influence of all other computers is very small so knowing the performance-hit becomes possible. In this case, assuming that reading of the two same application clocks near one another can be synchronized to begin with, their readings could then in practical terms be known even when they are separated.

Physical interpretation: Conventional relativistic equation for time dilation is derived as a limiting case in FIT.

8.3 Mass in Terms of Information

For a cluster computer, the information influence of a distant computer declines with cluster's size. For the same available information at the cluster's location, processing it to produce the same change in velocity (the end result of computation) will take longer if the cluster is larger, thus in simplest form:

$$\frac{\Delta v_k}{\Delta t} = \frac{F(I_{kc}, I_{kp})}{i_k}$$

where I_{kc} and I_{kp} are the current and previous sets of available information at cluster computer k from distant computers (the pairing of which *is* computing); i_k is self-information of the cluster and $F(I_{kc}, I_{kp})$ is some (generic) function of available information that via computation produces change in velocity. We will call this function $F(I_{kc}, I_{kp})$ an *information force:*

$$F(I_{kc}, I_{kp}) = i_k \times \frac{\Delta v_k}{\Delta t}$$

If a cluster is moving relative to its constraint group its information throughput will be lower and so will be the rate of the end result of its computation (which is change in velocity), i.e. the change in velocity will happen after a greater number of processing cycles. This is equivalent to the unchanging information throughput and the velocity higher in the same proportion:

$$F(I_{kc}, I_{kp}) = i_k \times \frac{d}{dt}\left(\frac{v}{\sqrt{1 - \Delta i_k^{\,2} / i_k^{\,2}}}\right)$$

Another interpretation of above equation is that the time needed to process information of a computer has increased, so we can introduce a quantity:

$$\frac{i_k}{\sqrt{1 - \Delta i_k^{\,2} / i_k^{\,2}}}$$

to be the *information mass*, and the quantity i_k to be the *rest information mass* (when there is no additional information, i.e. $\Delta i_k = 0$). Information mass is then *a measure of time needed to* *react to the same available information (i.e. to the same information force).* In other words, the slower the information throughput, the higher the information mass.

Because the information influence declines outward in a radial fashion, its decline or increase is largest in the direction from or towards the source of information field. Thus the minimum number of processing cycles needed to reach the location with the same change in information field is in the direction from or towards the source of the information field. This also represents the minimum use of acceleration resources and would be the path taken by a cluster. The change in velocity is then a vector corresponding to the information represented by $F(I_{kc}, I_{kp})$ where we use the vector notation signifying the direction between two clusters:

$$\vec{F}(I_{kc}, I_{kp}) = \frac{d}{dt}\left(\frac{i_k \times \vec{v}}{\sqrt{1-\Delta i_k^2 / i_k^2}}\right)$$

The quantity in parenthesis in above equations is *information momentum*:

$$\vec{p} = \frac{i_k \times \vec{v}}{\sqrt{1-\Delta i_k^2 / i_k^2}}$$

In limiting case of Eq. (6):

$$\vec{F}(I_{kc}, I_{kp}) = \frac{d}{dt}\left(\frac{i_k \times \vec{v}}{\sqrt{1-v^2 / c^2}}\right) \quad (7)$$

Physical interpretation: The concepts of inertial mass, momentum and force are derived in FIT. Information content takes the place of rest mass, adjusted for a constant of proportion.

8.4 The Information Reloading Effect

In the previous analysis, the information reloading was not taken into account. Let us observe a computer n at some distance from a computer m. The more information of m is present, the more is added to data of n during the uniform move away from m by dR due to reloading of information field of m. The longer it takes for n to move by dR, the more of m's information will add to n's data, hence:

$$d(\Delta i_n) = -w \times (i_m / R^2) \times i_n \times f_{mn} \times dt \quad (8)$$

where $d(\Delta i_n)$ is change in R-information of n due to reloading effect of m; w is a term of proportion that effectively describes how often the information field of m is reloaded in a unit of time; i_m is the information of m; R is the distance between m and n; f_{mn} is the information influence of m at n; i_n is the information of n, and $i_n \times f_{mn}$ is the portion of n's current set that comes from m; dt is a small time interval it takes to change distance by dR. There is no change due to motion effect because the speed is uniform and f_{mn} is practically constant.

Now consider a separate and equivalent situation when n moves away from m by a small distance dR with radial speed increasing by dv_R (here we ignore the reload effect and consider only motion effect). There will be an increase of M-information due to the motion effect. We will calculate the speed v_R at any given distance R so that the reload effect is equivalent to the motion effect. The change in additional information due to change of speed dv_R is:

$$d(\Delta i_n) = s \times i_n \times dv_R \times f_{mn} \quad (9)$$

After substituting $\Gamma = w / s$, from the previous two equations:

$$dv_R = -\Gamma \times (i_m / R^2) \times dt$$

Multiplying both sides by v_R, and with $dR = v_R \times dt$ we have (knowing that at distance of infinity the reload effect vanishes and so the speed is zero):

$$\int_v^0 v_R \times dv_R = -\int_R^\infty \Gamma \times (i_m / R^2) \times dR$$

The solution is:

$$v_R^2 = 2 \times \Gamma \times i_m / R$$

From limiting case of Eq. (6):

$$dt_1 = dt_2 \Big/ \sqrt{1 - 2 \times \Gamma \times i_m / R \times c^2} \quad (10)$$

This is the performance-hit due to the reload effect of i_m. The reload effect is apparently equivalent to a radial speed of $u_R = \sqrt{2 \times \Gamma \times i_m / R}$. We will call this radial speed the re*load velocity* $\vec{U}_R$. The additional information from the reload effect is thus:

$$\Delta i_n^R = s \times i_n \times f_{mn} \times u_R \quad (10.1)$$

Physical interpretation: A form of performance-hit that resembles Einstein's gravitational time dilation is deduced *without* the notion of gravity, as a result of changing information field.

8.5 Motion and Reload Effects Combined

The reload effect exists even when computers are at rest (the R-information is always there). Therefore, the total additional information due to motion with pre-existing reload effect is a sum of R-information (Δi_n^R), the M-information ($s \times i_n \times f_{mn} \times v_{mn}$) and the additional information due to combined R and M information ($s \times \Delta i_n^R \times v_{Rmn}$ because R-information is just added information at a given location) :

$$\Delta i_n = \Delta i_n^R + s \times i_n \times f_{mn} \times v_{mn} + s \times \Delta i_n^R \times v_{Rmn} \quad (11)$$

where v_{Rmn} is the radial speed between m and n. From above, the effective radial speed is (by expanding Δi_n^R):

$$v_R = v_{Rmn} \times (1 + u_{Rmn}) + u_{Rmn}$$

In spherical coordinates (radial, polar and azimuth), the information speed is:

$$V_n = \sqrt{(V_r)^2 + (V_\theta)^2 + (V_\varphi)^2}$$

The performance hit in general case is then:

$$\frac{dt_1}{dt_2}=\sqrt{\frac{1-\left[(V_{r1})^2+(V_{\theta 1})^2+(V_{\varphi 1})^2\right]/c^2}{1-\left[(V_{r2})^2+(V_{\theta 2})^2+(V_{\varphi 2})^2\right]/c^2}}$$

For a case of isolated large computer, after using above equation for radial speed, and ignoring higher order additions for speed, we have:

$$\frac{dt_1}{dt_2}=\sqrt{1-\frac{2\times\Gamma\times i_m}{R\times c^2}\left[1+(v_R{}^2/c)^2\right]-v^2/c^2}$$

This equation is derived under limiting case of Eq. (6) and does not include the information influences f_{mn} that decline with distance. The above performance hit dissipates with distance due to declining information influence, in addition to the first-order dependence on distance R in above equation.

Physical interpretation: A form similar to Schwarzschild metric emerges from a pure information theory, as a limiting case. In FIT, the performance-hit effects decline additionally with distance (due to declining information influence). The constant Γ is still just an undetermined constant in FIT and will be later shown to be the gravitational constant.

9. Relative Speeds

9.1 Nearby Large Isolated Computer

Let us have two small computers C_1 and C_2 nearby large isolated computer, as in Fig. 1. This represents a limiting case and Eq. (6) holds. C_1 and C_2 exchange a ping traveling at a maximum local speed c relative to a large computer, as in (a), regardless of direction (arrows indicate movement of a ping). When moving at some speed v relative to a large computer, as in (b), the speeds of a ping relative to it are $c-v$, $c+v$ and $\sqrt{c^2-v^2}$ respectively as self-evident in Fig. (1) because the speed of a ping *remains* c *relative to a large isolated computer.*

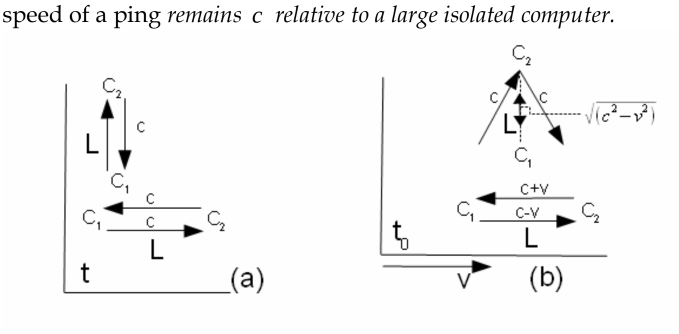

Figure 1.

The difference in times needed for ping to travel L in opposite directions in (b) is:

$$\Delta t=\frac{L}{c-v}-\frac{L}{c+v}=\frac{2\times L\times v}{c^2-v^2}\approx\frac{2\times L\times v}{c^2}$$

Note that a ping *will move slower when it travels in the direction of movement* relative to a large computer. Above result is the same when movement is circular.

Physical interpretation: In FIT the speed limit in a given locale determines the speed of light. Light emitted from a moving platform will always move at a speed limit relative to its local constraint group (a large object such as Earth in a limiting case) – meaning *light never has a priori speed relative to an emitter or a receiver.* Sagnac effect was derived (the result is the same in both linear or circular setup).

9.2 The Two-Computer System

Let us have two large isolated computers (1) and (2), as in Fig. (2). They orbit one another at fairly high speeds v_1 and v_2. Since this is not the case of a single large isolated computer, the limiting case of Eq. (6) does not hold. We will calculate information influences at (1) and (2) to obtain maximum speeds c_1 and c_2 of pings m and q (respectively relative to (1) and (2)) toward a distant observer (on the right in Fig. (2)).

Speeds in Fig. (2) are shown relative to a distant observer. For a ping m emitted toward distant observer, Eq. (2) gives assuming axis x is in the direction toward the observer:

$$(c_1+v_1)\times f_{1m}+(c_1-v_2)\times f_{2m}=1/s$$

Similarly for a small computer q:

$$(c_2+v_1)\times f_{1q}+(c_2-v_2)\times f_{2q}=1/s$$

where f_{1m} is information influence of (1) on m, f_{2m} is information influence of (2) on m, f_{1q} is information influence of (1) on q, f_{2q} is information influence of (2) on q, c_1+v_1 is the relative speed between (1) and m, c_1-v_2 is the relative speed between (2) and m, c_2+v_1 is the relative speed between (1) and q, c_2-v_2 is the relative speed between (2) and q.

In this case it is $f_{1m}\approx 1, f_{1q}\approx 0$ and $f_{2m}\approx 0, f_{2q}\approx 1$:

$$c_1=1/s-v_1=c-v_1$$

$$c_2=1/s+v_2=c+v_2$$

In the *vicinity of (1) and (2) the speed of a ping is dependent on the speed of (1) and (2).*

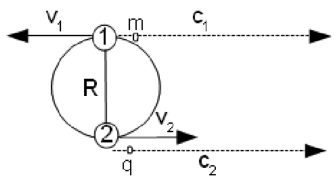

Figure 2.

When the distances d_1 and d_2 of both pings from (1) and (2) are sufficiently larger than R we have (assuming $d_1 \approx d_2$, $i_1 \approx i_2$, $i_q = i_m \ll i_1$, $v_1 \approx v_2$)

$$f_{1m} \approx \frac{i_1 / d_1^2}{i_m + i_1 / d_1^2 + i_2 / d_2^2} = f_{1q} = f_{2m} = f_{2q} \approx 0.5$$

$$c_1 = c_2 = c + 0.5 \times (v_2 - v_1) \approx c$$

The ping velocities away from (1) and (2) *become equal*. Thus pings emitted from (1) and (2) will effectively take the same time to reach a distant observer. If (1) and (2) move away from a distant observer *faster than* c, a ping may still reach distant observer because away from (1) and (2) *their information influence diminishes and the maximum local speed will be higher than* c *relative to (1) and (2)* according to Eq. (4).

Physical interpretation: De Sitter effect is explained in FIT *without the postulate about constancy of speed of light*. The speed of light changes depending on the locale. Light from distant objects moving faster than c relative to us may still reach us eventually.

9.3 Speed Limit When Approaching

If two computers approach each other at speeds higher than the speed limit (relative to their constraint groups), both computers will slow down. Each will slow down according to Eq. (4). From that, in a case of a small and a large computer, a large computer will therefore slow down very little and a small computer will slow down considerably.

Physical interpretation: A collision faster than c never happens even if the relative speed far exceeds c. *Due to increasing information influence, the relative speed must change to become* c *just before the collision*.

9.4 Length contraction

Consider a cluster of computers as in Fig. 3 under the limiting case of Eq. (6). We will assume that computers exchange pings, which are used as a method of interaction that keeps the cluster bound. The assumption is that the delay between pings (as measured by cluster computers) is used to maintain the form of the cluster. The shape of the cluster does not matter.

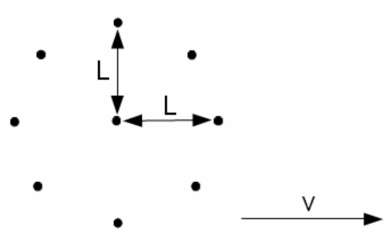

Figure 3.

When the cluster is stationary ($v = 0$), the delay between pings is, in any direction:

$$\Delta t = \frac{L}{c} + \frac{L}{c} = \frac{2 \times L}{c}$$

When moving, all computers in the cluster will experience a performance hit that will slow down the processing of pings. In case of pings traveling perpendicular to the direction of motion, from Fig. 1(b), the delay between pings, as measured by cluster computers is the same as when the group is stationary because when we account for the decreased information throughput and the change in delay between pings:

$$\Delta t = \frac{2 \times L}{\sqrt{c^2 - v^2}} \times \sqrt{1 - v^2 / c^2} = \frac{2 \times L}{c}$$

. In case of pings traveling parallel to the direction of motion, from Fig. 1(a), the delay between pings is, as measured by cluster computers:

$$\Delta t = \left(\frac{L}{c - v} + \frac{L}{c + v}\right) \times \sqrt{1 - v^2 / c^2} = \frac{2 \times L}{c} \times \frac{1}{\sqrt{1 - v^2 / c^2}}$$

The delay in the direction of motion is longer. If the delay between pings is used to maintain the shape, the computers in the direction parallel to the motion will move so that the delay remains the same and the length of the group will become:

$$L' = L \times \sqrt{1 - v^2 / c^2}$$

In more complex cases, such as when moving relative to two or more large nearby computers, the exact deformation of the group can be found by using Eq. (4) for maximum speeds.

Physical interpretation: In FIT, length contraction is a real phenomenon that however *applies to bound groups* of physical objects that relies on communication at the maximum local speed to maintain their position. The meaning of length contraction is only in decreasing distance between the bound objects. For instance, there is no length contraction in a system of two moving bodies that do not maintain structure this way.

10. Constant Γ

In order to preserve acceleration resources, a computer changes velocity to arrive to a nearby location where acceleration cost is permanently lower. The very act of changing velocity to arrive at this location will incur acceleration cost. This *acceleration cost* C is the change of information throughput T due to this change of speed:

$$dC = \frac{\partial T}{\partial v_R} \times dv_R$$

We used radial speed v_R in case of two isolated computers where the lowest acceleration cost is always in radial direction. When a lower-cost location is reached there is a *gain of acceleration resources* Y due to this change of location:

$$dY = \frac{\partial T}{\partial R} \times dR$$

Consider an accelerating computer moving to a nearby location where the usage of resources is lower. The goal is to reach the location with minimum usage of resources, in addition to achieving the permanent lower usage of resources. If acceleration is too high, the cost of moving to a new location can be high. If acceleration is too low, then more resources will be lost while being in transit. The sum $C + Y$ represents the *total acceleration* *expenditure* of such movement (the cost C is negative and the gain Y is positive). The minimum for this function (if it exists) signifies the actual acceleration to a location with lower usage of resources. The minimum total cost during some small period of time can be deduced by solving:

$$\frac{d}{dt}(C+Y)=0$$

or:

$$\frac{\partial T}{\partial v_R}\frac{dv_R}{dt}+\frac{\partial T}{\partial R}\frac{dR}{dt}=0$$

and we can reduce to change in additional information:

$$\frac{\partial T}{\partial(\Delta i)}\times\frac{d(\Delta i)}{dv_R}\times\frac{dv_R}{dt}+\frac{\partial T}{\partial(\Delta i)}\times\frac{d(\Delta i)}{dR}\times\frac{dR}{dt}=0$$

Finally with Eq. (8) and (9) for computers n and m :

$$s\times i_n\times f_{mn}\times\frac{dv_R}{dt}+w\times(i_m/R^2)\times i_n\times f_{mn}\times\frac{dt}{dR}\times\frac{dR}{dt}=0$$

and with the same substitutions of constants as in Eq. (10):

$$\frac{dv_R}{dt}=-\frac{\Gamma\times i_m}{R^2}$$

This is the derivation of *Newtonian gravity* and so it must be:

$$\Gamma=G$$

where G is the gravitational constant. The Eq. (10) can now be written as:

$$dt_1=dt_2\Big/\sqrt{1-2\times G\times i_m/R\times c^2}$$

Physical interpretation: Gravity in FIT is a result of conservation of computing resources. Note that while we used "smooth" derivatives to deduce the acceleration, the change in density of information field is not smooth because information content is essentially an integer resource. Also, the acceleration differs with motion as the information throughput changes.

10. Quantum Effects

10.1 Uncertainty

Change of velocity is given as a set of facts with two components to it (one that describes speed and one that describes unit vector). For our purposes here, we can consider there to be two sets of facts (for speed and unit vector change). The information is a set of facts and the result of any of its use is always an integer. This means that change in speed of a standalone computer cannot take any arbitrary values, rather it must be discrete. The unit vector in space cannot be an arbitrary one, rather there are a finite number of directions in space that a standalone computer can move towards.

We used calculus in FIT under the assumption that the number of facts is statistically large but with understanding that in general the results obtained must be an approximation only, albeit a very good one in many cases.

We will call the speed and unit vector that would be calculated if there was no information loss an *ideal speed* and an *ideal unit vector*, and computation that would do that an *ideal computation*. We will denote p_v as the probability for the unit vector to equal an ideal unit vector, and p_s is the probability for the speed to be equal an ideal speed. Each is a function of the number of lost facts that affect the computation – the more facts lost, the lower such probabilities may be:

$$p_v=P(i_v,\delta i_v)\quad,\quad p_s=P(i_s,\delta i_s)$$

where i_v is the number of facts used to compute the unit vector, δi_v is the number of lost facts used to computer the unit vector, i_s is the number of facts used to compute the speed, δi_s is the number of lost facts used to compute the speed. Total loss of information limits the values for δi_v and δi_s , and total information about each is limited by the computer's storage:

$$\delta i_v+\delta i_s=(\Delta i)^2-\delta e\quad,\quad i_v+i_s=i^2 \tag{12}$$

The δe is the error in estimate of information loss. This estimate exists because a lost fact may turn out to be an ideal one (a correct one, i.e. it was not really lost). If the compressed information was not really lost, it may be $\delta e\approx\Delta i^2$; if the compressed information was lost entirely it may be $\delta e\approx 0$.

From above correlation between the information used to compute the speed and unit vector we get:

$$p_v=P(i_v,\delta i_v)=P(i^2-i_s,\Delta i^2-\delta i_s-\delta e)=P_1(i_s,\delta i_s,\delta e)$$

or generally:

$$p_v=f(p_s,\delta e)$$

This means the *the probability to compute ideal unit vector is correlated with the probability to compute ideal speed (and vice versa), even if either one is entirely unpredictable to us.* The correlation is likely to be a complementary one, according to Eq. (12). If the number of lost facts used to compute unit vector is very small, then the number of lost facts used to compute speed is likely large. In other words, the two values computed share the error space of computation and so the errors are likely to be complementary.

Physical interpretation: The uncertainty typically associated with quantum mechanics arises naturally in FIT due to limits of information processing, *and not because of either the implied uncertainty nor because of any measurement process.*

11. FIT and Physics

FIT is a purely informational theory applied to trivial physical concepts and it is a discrete theory at the core. Even being a discrete theory, it produces results in limiting cases identical to those of Einstein's SRT and GRT. However there are important differences. To start, FIT begins and ends its deliberations before the first principles of physics. SRT and GRT rely on first principles (i.e. relativity) and then establish some of their own, based on known experimental results (i.e. constancy of speed of light) as well as thought-experimental heuristics (i.e. equivalence principle). In contrast, FIT relies on largely axiomatic approach of the simplest model of information use and even more importantly on the necessity of information use on a foundational level.

Using FIT we have derived the notions of inertial and gravitational mass (as being information-based), the postulate of constancy of speed of light (as applicable to limiting cases), both velocity and gravitational time dilation (as limiting cases and without a notion of gravity), and covered (as examples) the De Sitter effect and Sagnac effect.

While relativity operates on a more heuristic approach and purports to say that time itself dilates, FIT takes a simpler approach. After all, *we don't know exactly what time is* but we do know how other phenomena relate to it. FIT examines the very core of the rate of physical processes. The central idea is that what powers those processes cannot be anything else but the use of information, because without it all that is left is truly random behavior. Thus, the rate of physical processes becomes directly related to the throughput of information. In FIT there is *real-time,* and it is the rate of processes of a faraway small mass (an unencumbered processing of information and thus the fastest), and there are *application-times* as the rate of processes measured by entities based on their own information throughput. In FIT, time itself remains constant, only the rate of processing information changes.

FIT does not use principle of relativity. The closest equivalent to principle of relativity in FIT is the assumption that the same information input produces the same information output. FIT examines more closely the very notion of movement from the perspective of information use. In FIT relative motion changes the throughput of information and it alters the precision of end-results of such information use. Not all relative motion is equal. Movement relative to some objects (generally those closer and more massive) counts more when it comes to information use (if we assume that all facts take the same time to process). Thus the very question of frames of reference with respect to laws of physics (as having a meaning beyond mere coordinate translation) does not exist in FIT. Instead, *laws of information use are the same for all information processing constituents.*

If we start with relativity as a first-order principle, we are essentially forfeiting the question of why relativity seems to work for us. For example if we knew the mechanism behind the seeming constancy of speed of light, are we sure such mechanism would absolutely validate relativity? FIT produces almost identical results to relativity, but not exactly the same. FIT shows that in Michelson Morley experiment the Earth practically becomes something akin to a local preferred frame of reference and the movement of Earth cannot be detected. However, FIT predicts *that a setup in motion relative to Earth would behave differently*.

FIT does not run into the question of whether time can dilate or not. Variable rate of information use is a concept that does not necessitate that question. It also entirely objectifies any physical system because it provides the answer to *why do physical processes slow down.* With that knowledge in hand, the answer to this question is revealed without any *actual* mechanical details, but rather as a generic model. The nature of time is spelled out clearer in FIT. A set of 100 facts will take the same time to process, *regardless of the state of motion.* It means *time itself is invariant with respect to information use.* However, the number of facts processed that *affect* physical change *depends on the state of motion* and the *measure of time is not invariant with respect to information use.* Because processing cycle takes a finite time to complete, there is a minimum time interval before physical change can occur. It means that *dt* in physical equations cannot really be considered infinitesimal.

FIT tells us that motion and reload effects cause changes in throughput of information use (*i.e.* performance-hit), which causes different rates of physical processes. This is what Einstein calls time dilation. FIT predicts the same expressions for time dilation as does SRT and GRT in limiting cases (however *without* using any physical postulates). In FIT those effects also *additionally* depend on the ratio of masses involved and the square of the distances in a way that is not trivial. FIT *derives* the constancy of speed of light in *a limiting case* as a consequence of the throughput of information being greater than zero. An important point is that *the constancy of speed of light and the "time dilation" have no causal relationship* (i.e. one does not cause the other), but rather are *both consequence of the limited information throughput*.

Einstein's famous (and popular) v^2/c^2 and $2\times G\times m/R\times c^2$ ratios (used in SRT and GRT) are generalized in FIT with a single ratio $\Delta i^2/i^2$, which in limiting cases reduces to above Einstein's ratios. FIT's ratio $\Delta i^2/i^2$ essentially *represents information overload in an entity where information storage is not infinite.*

What of reciprocity of velocity time dilation in SRT? In SRT, when a spaceship approaches Earth, not only does the time slow down for a spaceship relative to Earth, but also equally does for Earth relative to a spaceship. SRT is a self-consistent theory and this seemingly contradictory statement works itself out. In FIT, the clock of a spaceship ticks slower because its throughput of information processing is lower. The clock on Earth also ticks slower but by such a smaller rate it is practically negligible (so the effect is not symmetrical). When we say that rate of a clock is slower, we mean *relative to the rate of physical processes of the smallest possible mass that is infinitely far away from other masses.* In both GRT and FIT there is the same gravitational time dilation (except that in FIT it is not called gravitational because *gravity is not necessary* to derive it). In FIT both of these "time dilation" effects decrease additionally due to ratio of masses and the square of distance in a non-trivial correlation. The end result is that at cosmological distances for macroscopic bodies, the time dilation between masses separated by large distances *may* vanish. However the time dilation due to the presence of nearby masses is still there and the exact results are no longer as simple to predict as they are in Einstein's view. In other words, the "time dilation" of a faraway mass *may have nothing to do with movement relative to us* and *everything to do with movement relative to masses near it*.

Recent observations of the lack of time dilation in quasars [2] are in agreement with FIT. Very large masses are much likely to not experience time dilation and can move with speeds that exceed speed of light relative to us for the same reason. This has been observed in "superluminal motion of galaxies"[3].

FIT also predicts that an experiment involving the speed of light in a setup moving relative to Earth should show a positive result. Indeed a recent modified Sagnac experiment [1] confirms this and its results are in mathematical agreement with FIT.

While FIT, like Einstein's relativity, predicts that nothing can move faster than 300,000km/s near massive bodies such as

Earth, in FIT this is different away from such bodies. Large masses away from bodies such as Earth can accelerate past 300,000km/s relative to them. While contradictory to contemporary body of physics, consider that in FIT the maximum local speed depends on the throughput of information processing, which in case of enough separation does not increase with speed any more.

Regarding superluminal speeds, FIT offers explanation that does not contradict current experimental evidence. Light emitted from objects moving away faster than 300,000km/s (superluminally) can still reach us because *the speed limit exists relative to a current constraint group*, and constraint group (a set of objects that have prevalent information influence) varies depending on location. For the same reason, a photon moving toward us superluminally will decrease its speed to 300,000km/s when it approaches. This is a simple and straightforward consequence of changing information influence, where speed limit is determined by a local constraint group.

Gravity in FIT is a consequence of a changing information field of an object and the conservation of resources. In FIT, gravity is not there to begin with and there is no "time dilation" to begin with. All of these effects are *derived from scratch* based on the information theory and are connected through information use on a fundamental level.

FIT is by nature a discrete theory which points to association with quantum effects. We have *derived the necessity for a variant of what's known as Heisenberg's Principle of Uncertainty to exist.* Current quantum physics leans toward a bias of not knowing what cannot be measured. FIT starts from objective reality where not knowing comes from limitations of information use (limited speed of processing and declining density of information field) and not from an innate quality of unknowing.

Hidden variables in physics have been debated for decades. The question is whether there can be a more "realistic" view of quantum effects that involves determinism and also if such a view would involve "locality" (or if it required propagation beyond speed of light). From the perspective of FIT, the important questions posed are inadequate and unfortunately they frame the possible answers. Determinism of computation *does not* imply ability to predict end-results *if resources are limited.* Limited resources generally mean loss of information and that means unpredictability. This is vastly different from a postulated and innate unpredictability. In FIT, *information is partially-local* because it is neither local nor non-local: information is fairly localized in its point of origin but it exists everywhere instantaneously, however not anywhere in its entirety at the same time. In FIT *movement of localized information is still shown to be subject to local speed limit however the influence of that information is instantaneous everywhere*.

Quantum physics is historically based on predicating reality based on what can be measured and the probabilities of such measurements. FIT predicates reality on a simplest possible information model. This information model *is independent of any measurements*. Because of that, simply put, information is processed even if no one is looking. Information we obtain from measurements (the discernable information) is just a subset of fundamental information. All our measuring tools process information too and they influence processing of information of what is being measured. However that influencing is *not* what brings fundamental uncertainty *per se*. It is the *limited resources* for processing information.

Lorentz invariant quantities and Lorentz covariant equations play an important role in physics. In FIT the available information would be Lorentz invariant if it were observable. In FIT, less is required of Nature that it is in physics. In physics, laws of Nature are likened to require the same "view of the world' to be "shown" to every observer. Arguably that is a beautiful and a succinct requirement in itself. In FIT, there is no requirement at all like this. One could say that no requirement is more beautiful than even the most beautiful requirement because a requirement (whether it is called a postulate or principle or whatever else) is just another name for a hunch (in absence of any deeper insights). What is not expected is that FIT comes out to support (in limiting cases) Lorentz invariance in physics without the requirement itself. FIT only requires that basic tenets of information use be met. It does not connect a "view of the world" in any way with laws of Nature. Yet, Einstein's equations that are consequence of Lorentz invariance also appear as a consequence of FIT. Why is this so? As we said, fundamental information *would be Lorentz invariant if it were observable.* Regardless of being observable or not (since FIT is a realism-based theory), information is a fundamental invariant quality in FIT which produces (as limiting cases) what is known as Lorentz-invariance.

An important question is that of the principle of relativity itself. Relativity is preserved in FIT on a more fundamental level. *Laws of information processing are the same for all physical fundamental entities.* This does not necessarily translate into Einstein's relativity because FIT transposes the essentials when compared to Einstein's relativity. In Einstein's relativity *physical laws govern fundamental physical entities* while in FIT *physical laws are created by information processing* of the fundamental physical entities. On a more informal level, Einstein's relativity is similar to the notion of laws in human society (i.e. people follow laws that rule the land) while FIT is similar to the individualism of persons that comprise the society (i.e. person's behavior is based on their own mind). Just like in human society the difference is nominally not apparent and the societal, economical and political laws can generally be used to predict how statistically large populations behave. But just like in human society, in limiting cases those laws break down exposing the fact that they are *only* approximations of many individual decisions and are not laws *per se*.

While FIT offers additional predictions that may prove it right or wrong (versus Einstein's relativity), it also offers something else: the underlying mechanism that does not depend on faith. We must admit that Einstein's postulates are faith-based. While the subject of this faith is often said to be elegant and proven, it can also be said to lack any true explanation and any underlying understanding whatsoever. That is the biggest weakness of Einstein's relativity even in the face of all of its success in the past century.

The arrow of time in physics poses an interesting question: can processes play backwards if the laws that govern them are T-symmetrical (meaning symmetrical in time)? In FIT the change in a system is due to available information. Take for example an accelerating object moving from one location to another. The object accelerated because available information in the first location affected it through information use. Once in the second location, the chances of available information there warranting the reversal are generally small. While a mathematical law that describes this movement seems to allow for reversibility of the process (by using negative time, i.e. a time in reverse), the informational foundation for such a law may not. In essence, if FIT is correct, the Nature is informational at core and mathematical in appearance (as a result of laws of information use). This is an extremely important differentiator. Mathematical representation allows for symmetry in time, whereas informational representation (of reality) generally does not. Mathematics of time in physics is a derivative and an approximation of information use on an antecedent level.

Unlike SRT, propagation of light in FIT can be in form of waves in a medium. Because FIT does not postulate constancy of light speed, a medium comprised of fundamental physical entities (i.e. computers in FIT) can propagate waves. The movement of such waves follows the same basic rules as the movement of individual entities and does not infringe upon the body of relativistic experiments, but it does answer a simple question of *what is waving* in a light wave. SRT prohibits the wave medium while FIT allows it. Yet FIT deduces the same effects as SRT, in limiting cases. It is somewhat ironic because inability to reconcile a propagation medium with the constancy of speed of light was the primary motivator for Einstein's relativity.

Speed of light is an important quantity in physics, not only in the area of its historical roots (optics and electromagnetic field) but elsewhere too. What is the speed of light in FIT? It is the ratio of physical length of a computer and the time it takes for it to complete a processing cycle. It has nothing to do with any particular field of physics but rather is a fundamental property of information processing and a physical measure a computer at that. If information processing is at the core of reality it would make sense that speed of light is related to it and not to any particular manifestation of it.

Is the central FIT assumption about finite throughput of information processing isomorphic to a postulate of constancy of speed of light (as both represent limitations)? It is not. Simply put, if the throughput of information were infinite, both the input and output of computation would be available at the same time. This means that current information before and after computation would be available at the same time and the actual computation could use either one (as there is no preference to use either one). It means that the entire current information set would be randomly chosen and this would result in loss of entire result of computation. On the other hand, if the throughput of computation were too slow, the time needed to compute the change in velocity would be too great. It is not difficult to imagine that this would lead to inability to form bound systems (i.e. "clusters of computers" in FIT). In this case computers would quickly move away and become isolated, eliminating the need for computing in the first place. This leaves a certain finite and high value for the throughput of computation (and consequently the speed of light) as the most likely one. Thus the need for finite throughput of computation is not isomorphic to the finite speed of information transfer.

Even though FIT purports to bring the more fundamental view of reality the question remains as to what exactly the physical nature of information is. We may accept that any physical action has a reason, and that reason is a euphemism for applied information use (even if such a line of thinking has been ignored over centuries). However the question remains of what may be the actual physical reality of information – on a fundamental level. What is the form of information? This question is not answered in FIT. What is answered is: *what information laws any such form must obey*? It is axiomatic that there is no way to produce non-random behavior other than by means of a discriminator of some sort. In the final analysis this discriminator is always information content. This is the reason to believe FIT is more than just a coincidental analysis or an isomorphism of sorts.

One of the defining features of FIT is that information has a definitive point of origin but it exists everywhere with information present on nearly every imaginable sphere around it at any point in time. Information moves as a computer moves. Because of this, *computers* ***can*** *interact directly and there is no need for a concept of "action at distance". For if any two computers occupy all of the space and thus the same space then the interaction is nearly instantaneous (even if incomplete due to declining density of information field around computers).* From a perspective of fundamental entities, the entire reality consists of the local information. In other words, *there is no significance in relative motion per se, just in the changes in local information field and varying throughput of local information use* (caused by movement). This is important because the thought experiments of SRT often speak of various observables relative to a moving frame of reference. If FIT is correct, the only true observables are the changes in local information.

From antiquity physics has relied on notion of force that is push or pull. The meaning of force thus has no deeper physical representation other than the observed one. FIT offers a notion of force as a pairing of facts (i.e. computation) and the acceleration as a result of such usage of facts. Similarly, the postulates of relativity and quantum physics have no deeper physical representation, being self-evident in nomenclature itself, with postulates leading the way. But more than that, these postulates are (just like the notion of force) *singularly borne out of experience.* FIT offers a more defined foundation that has usage of facts at its core. More importantly, in FIT all of these aspects of conventional physics (such as classic, relativistic or quantum physics) stem from the necessity of information use, or rather from impossibility of having (or constructing) a workable reality *without the use of information*. This impossibility is a simple premise that lies at the heart of FIT. Perhaps its ability to plainly corroborate classic, relativistic, gravitational and quantum effects without an artificial distinction is its defining characteristic.

References

[1] Ruyong Wang, Yi Zhengb, Aiping Yaob, Dean Langley, "Modified Sagnac experiment for measuring travel-time difference between counter-propagating light beams in a uniformly moving fiber", Physics Letters A, 312, 7-10, (2003).

[2] MRS Hawkins, "On time dilation in quasar light curves', Monthly Notices of the Royal Astronomical Society, Volume 405, Issue 3, pages 1940–1946, (July 2010).

[3] Richard Porcas, "Superluminal motions: Astronomers still puzzled". Nature 302, 5911 (1983).