Computational Information Conservation Theory: An Introduction

RODOLFO A. FIORINI
Department of Electronics, Information and Bioengineering (DEIB)
Politecnico di Milano
32, Piazza Leonardo da Vinci, 20133 Milano
ITALY
rodolfo.fiorini@polimi.it http://www.linkedin.com/pub/rodolfo-a-fiorini-ph-d/45/277/498

Abstract: - Mankind's best conceivable worldview is at most a partial picture of the real world, a picture, a representation centered on man. We inevitably see the universe from a human point of view and communicate in terms shaped by the exigencies of human life, in a natural uncertain environment by incomplete knowledge. To grasp a more reliable representation of reality, researchers and scientists need two intelligently articulated hands: both stochastic and combinatorial approaches synergically articulated by natural coupling. In the past, many attempts to arrive to a system continuum-discrete unified mathematical approach have been proposed, all of them with big operational compromises. All these attempts use a top-down (TD) point-of-view (POV). From a computational perspective, all approaches that use a TD POV allow for starting from an exact global solution panorama of analytic solution families, which, unfortunately, offers a shallow local solution computational precision to real specific needs; in other words, overall system information from global to local POV is not conserved, as misplaced precision leads to information dissipation. On the contrary, to develop antifragile system, we need asymptotic exact global solution panoramas combined to deep local solution computational precision with information conservation. The first attempt to identify basic principles to achieve this goal for scientific application has been developing at Politecnico di Milano since the end of last century. In 2013, the basic principles on computational information conservation theory (CICT), for arbitrary-scale discrete system parameter from basic generator and relation, appeared in literature. A synthetic and effective comparative example from a CICT perspective is presented.

Key-Words: - CICT, information geometry, biological correlates, neurological correlates, natural uncertainty, epistemic uncertainty, entropy, wellbeing, health care quality, resilience, antifragility.

1 Introduction

Mankind's best conceivable worldview is at most a partial picture of the real world, a picture, a representation centered on man. We inevitably see the universe from a human point of view and communicate in terms shaped by the exigencies of human life in a natural uncertain environment by incomplete knowledge. Intrinsic randomness of a phenomenon (e.g. throwing a dice) or natural uncertainty cannot be reduced by the collection of additional data (big data approach) and it stems from the variability of the underlying biophysical process. Unlike natural uncertainty, epistemic uncertainty can be reduced by the collection of additional data. Epistemic uncertainty sources are still treated by risk analysis approach only, which provides an acceptable cost/benefit ratio to producer/manufacturer, but in some cases it may not represent an optimal solution to end user. In fact, deep epistemic limitations reside in some parts of the areas covered in decision making. More generally, decision theory, based on a "fixed universe" or a model of possible outcomes, ignores and minimizes the effect of events that are "outside model". Epistemic and aleatory uncertainties are fixed neither in space nor in time. What is aleatory uncertainty in one model can be epistemic uncertainty in another model, at least in part. And what appears to be aleatory uncertainty at the present time may be cast, at least in part, into epistemic uncertainty at a later date. Paradoxically if you don't know the underlying hidden generating process for the folded information you can't tell the difference between an information-rich message and a random jumble of letters. The observer, having incomplete information about any generating process, and no reliable theory about what the data correspond to, will always make inference mistakes, but these mistakes have a certain pattern. Statistical and applied probabilistic theory is the core of classic scientific knowledge; it is the logic of "Science 1.0"; it is the traditional instrument of risk-taking. Unfortunately, the "probabilistic veil" can be very opaque computationally, and misplaced precision leads to information dissipation and confusion. To grasp a more reliable representation of reality and to get stronger biological and physical system correlates, researchers and scientists need two intelligently articulated hands: both stochastic and combinatorial approaches synergically articulated by natural coupling. The former, applied to all branches of human knowledge under the "continuum hypotesis" assumption, has reached highly sophistication level, and a worldwide audience. Many "Science 1.0" researchers and scientists up to scientific journals assume it is the ultimate language of science. The latter, less developed under the "discrete hypotesis" assumption in specific scientific disciplines, has been considered in peculiar application areas only. It has been further slowly developed by a few specialists and less understood by a wider audience. Unfortunately, the above two mathematical research areas have followed separate development paths with no articulated synergic coupling. In the past, many attempts to arrive to a continuous-discrete unified mathematical approach have been proposed, all of them with big operational compromises. There are still several current attempts to unify continuous and discrete mathematics. One of the many approaches is the development of the time-scale Calculus (TSC), introduced by German mathematician S. Hilger in 1988 [19]. Since then hundreds of papers appeared in the theory and its applications to dynamic equations [4]. It has been created in order to unify continuous and discrete analysis, and it allows a simultaneous treatment of differential and difference equations, extending those theories to so-called dynamic equations. TSC is a unification of the theory of difference equations with that of differential equations, unifying integral and differential calculus with the calculus of finite differences, offering a formalism for studying hybrid discrete–continuous dynamical systems [3]. It has applications in any field that requires simultaneous modelling of discrete and continuous data. It gives a new definition of a derivative such that if one differentiates a function which acts on the real numbers then the definition is equivalent to standard differentiation, but if one uses a function acting on the integers then it is equivalent to the forward difference operator. Similar ideas have been used before and go back at least to the introduction of the Riemann–Stieltjes integral which unifies sums and integrals. The three most popular examples of TSC are differential calculus, difference calculus, and quantum calculus. Dynamic equations on a time scale have a potential for applications, such as in population dynamics. For example, they can model insect populations that evolve continuously while in season, die out in winter while their eggs are incubating or dormant, and then hatch in a new season, giving rise to a non–overlapping population. As a matter of fact, of a great deal of interest are methods being introduced for dynamic equations on time scales, which now are able to explain some discrepancies that have been encountered when results for differential equations and their discrete counterparts have been independently considered. The explanations of these seeming discrepancies are incidentally producing unifying results via time scales methods [3]. Nevertheless, current research in different but connected areas is approached by different methods that later have to be reducted (if it is possible) on a common ground [4]. All these unification attempts use a top-down (TD) point-of-view (POV). Unfortunately, from a computational pespective, all approaches that use a TD POV allow for starting from an exact global solution panorama of analytic solution families, which offers a shallow local solution computational precision to real specific needs (in other words, overall system information from global to local POV is not conserved, as misplaced precision leads to information dissipation [12]). In fact, usually further analysis and validation (by probabilistic and stochastic methods) is necessary to get localized computational solution of any practical value, in real application. A local discrete solution is worked out and computationally approximated as the last step in their line of reasoning, that started from an overall continuous system approach (from continuum to discrete ≡ TD POV). Unfortunately, on the other end, to develop resilient and antigfragile application,we need stronger biological and physical system correlates; in other words, we need asymptoctic exact global solution panoramas combined to deep local solution computational precision with information conservation and vice-versa. Can we achieve this goal? In section 2 we present the root of the problem and we give a problem formulation. We show a comparative example of Information Geometry Theory (IGT) to computational information conservation theory (CICT) approach, to better understanding their mutual operative advantages. In section 3 we draw a problem solution path. Last Section 4 sums up our findings from a top level perspective.

2 The Root of the Problem

The most fundamental concept of Mathematical Analysis is that of the function. Two sorts of functions are to be distinguished. First, functions in which the independent variable x may take every possible value in a given interval; that is, the variable is continuous. These functions belong to the domain of Infinitesimal Calculus (IC). Secondly, functions in which the independent variable x takes only given values; then the variable is discontinuous or discrete. In the same way, we talk of continuous probability distribution and discrete probability distribution. Unfortunately, to discrete variable the methods of IC are NOT applicable. To deal with discrete variables, we need the Finite Differences Calculus (FDC). The origin of this Calculus may be ascribed to Taylor [32], but the real founder of the theory was Jacob Stirling [31], who solved very advanced questions, and gave useful methods. Introducing the famous Stirling numbers, he paved the way even to an important part of modern combinatorics. The Stirling numbers form the backbone of the Calculus of Finite Differences. The first treatise on this Calculus is contained in [11], in which Leonardo Eulero was the first to introduce the symbol Δ for the differences, which is universally used. Since then many finite difference treatises were published, but the first most important authors to remember are Boole in 1860, Markoff in 1896, Selivanoff in 1904, Whittaker and Robinson in 1924, Nörlund in 1924, Steffensen in 1927, Scarborough in 1930, Kowalewski in 1932, Milne-Thomson in 1933 and Jordan in 1939 [22]. So, FDC deals especially with discrete functions, but it may be applied to continuous function too. As a matter of fact, it can deal with both discrete and continuous categories conveniently. In other words, if we want to achieve an overall system information conservation approach, we have to look for a convenient bottom-up (BU) POV (from discrete to continuum view ≡ BU POV) to start from first, and not the other way around! Then, a TD POV can be applied, if needed. We show a comparative example of Information Geometry Theory (IGT) to CICT approach, to better understanding their mutual operative advantages.

2.1 IGT

In 1945, by considering the space of probability distributions, Indian-born mathematician and statistician Calyampudi Radhakrishna Rao (1920-) suggested the differential geometric approach to statistical inference. He used Fisher information matrix in defining the metric, so it was called Fisher–Rao metric. In 1975, American statistician Bradley Efron (1938-) carried the argument a step forward when he introduced a new affine connection on the parameter space manifold, and thus shed light on the role of the embedding curvature of the statistical model in the relevant space of probability distributions. So, Information Geometry (IG) emerged from the study of the geometrical structure of a manifold of probability distributions under the criterion of invariance. IG reached maturity through the work of Shun'ichi Amari (1936-) and other Japanese mathematicians in the 1980s. Amari and Nagaoka's book [1] is cited by most works of the relatively young field due to its broad coverage of significant developments attained using the methods of information geometry up to the year 2000. Many of these developments were previously available in Japanese-language publications only. IG takes probability distributions for a statistical model as the points of a Riemannian manifold, forming a statistical manifold. The Fisher information metric provides the Riemannian metric. Moreover, a family of dually coupled affine connections are introduced. Mathematically, this is a study of a triple $\{\mathbf{M}, g, \mathrm{T}\}$, where $\mathbf{M}$ is a manifold, g is a Riemannian metric, and T is a third-order symmetric tensor. It largely focuses on typically multivariate, invariant and higher-order asymptotic results in full and curved exponential families, through the use of differential geometry and tensor analysis. Many important families of probability distributions are dually flat Riemannian manifolds. A dually flat manifold possesses a beautiful structure: it has two mutually coupled flat affine connections and two convex functions connected by the Legendre Transformation. It has a canonical divergence, from which all the geometrical structure is derived. During the last 15 years (1999-2014), it has been applied not only to statistical inferences but also to various fields of information sciences where probability plays an important role. The Kullback-Leibler divergence (introduced in 1951), for short KL-divergence, in probability distributions is automatically derived from the invariant flat nature. Moreover, the generalized Pythagorean and geodesic projection theorems hold. Conversely, we can define a dually flat Riemannian structure from a convex function. This is derived through the Legendre Transformation and Bregman Divergence (1967) connected with a convex function. The Fisher-information matrix is used to calculate the covariance matrices associated with Maximum-Likelihood Estimator (MLE). It can also be used in the formulation of test statistics, such as the Wald test. MLE gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However, in some complicated problems, difficulties do occur: in such problems MLEs are unsuitable or do not exist.

Riemann surfaces are named after the 19th century mathematician Bernhard Riemann, who was the first to understand the importance of abstract surfaces, as opposed to surfaces arising concretely in some ambient space. Compact Riemann surfaces are geometric structures constructed by cleverly stretching, bending, and gluing together parts of the complex number plane. A surface becomes a Riemann surface when it is endowed with an additional geometric structure. One can think of this geometric structure as a so-called complex structure, which allows one to do complex analysis on the abstract surface. On a hyperbolic Riemann surface, we don't have straight lines, given the curvature of the surface. Instead, geometers consider "geodesics" when working in a non-Euclidian space. Geodesics are the natural analogue of lines: just as a line segment in a plane is the shortest path between two points, a geodesic on a Riemann surface is the curve with the shortest hyperbolic length between two points on the surface. Because of the compact closed nature of surfaces, geodesics can sometimes loop around the surface and come back to their starting point, and these are called closed geodesics. Closed geodesics that don't intersect themselves while tracing their path around the surface are called simple closed geodesics. The interesting hyperbolic Riemann surfaces all have a broad global structure that looks like some number of toruses or donuts glued together. It's possible to take two surfaces made out of the same number of donuts and deform by pulling or stretching on one surface to get the other surface, and so these surfaces are in some way related to each other. The number of donuts a surface is made out of is called the genus g of the surface. While the underlying topological surface remains the same, its geometric shape changes during a deformation. Riemann knew that these deformations depend on "6g-6" parameters or "moduli", meaning that the "moduli space" of Riemann surfaces of genus g has dimension "6g-6." However, this says nothing about the global structure of moduli space, which is extremely complicated and still very mysterious. For instance, when one considers only the closed geodesics that are simple, meaning that they do not intersect themselves, the growth of the number of geodesics of length at most L is no longer exponential in L but is of the order of "L6g-6." Recently, Mirzakhani showed that in fact the number is asymptotic to "c L6g-6" for large L (going to infinity), where the constant c depends on the hyperbolic structure [9,10]. The theory of Riemann surfaces and algebraic geometry are closely linked. Every complex curve is an algebraic curve, meaning that the complex curve, although defined abstractly, can be realized as a curve in a standard ambient space, in which it is the zero set of suitably chosen polynomials. Thus, although a Riemann surface is a priori an analytic object defined in terms of complex analysis on abstract surfaces, it turns out to have an algebraic description in terms of polynomial equations. An alternative but equivalent way of defining a Riemann surface is through the introduction of a geometry that allows one to measure angles, lengths, and areas. The most important such geometry is hyperbolic geometry (HG), the original example of a non-Euclidean geometry discovered by Bolyai, Gauss, and Lobatchevski. The equivalence between complex algebraic and hyperbolic structures on surfaces is at the root of the rich theory of Riemann surfaces. Sometimes properties of a fixed hyperbolic surface can be better understood by studying the moduli space (MS) that parametrizes all hyperbolic structures on a given topological surface. For example, thinking of Riemann surfaces as algebraic curves leads to the conclusion that MS itself is an algebraic object called an algebraic variety. In addition, MS has a metric whose geodesics are natural to study. These MSs have rich geometries themselves, and arise in natural and important ways in differential, hyperbolic, and algebraic geometry. There are also connections with theoretical physics, topology, and combinatorics. The behavior of geodesics in moduli space is even related to dynamical system on MS. Non-closed geodesics in MS are very erratic and even pathological, and it is hard to obtain any understanding of their structure and how they change when perturbed slightly. However, complex geodesics and their closures in MS are in fact surprisingly regular, rather than irregular or fractal [10]. It turns out that, while complex geodesics are transcendental objects defined in terms of analysis and differential geometry, their closures are algebraic objects defined in terms of polynomials and therefore have certain rigidity properties [9]. Rigidity is a fundamental phenomenon in HG and holomorphic dynamics. Its meaning is that the metric properties of certain manifolds or dynamical systems are determined by their combinatorics. Moreover, this phenomenon is intimately linked to the universality phenomenon, to basic measure-theoretical and topological properties of systems, to the problem of describing typical systems. An important development in Riemann surface theory has been the discovery of fertile connections between rational billiards, translation surfaces and flows on Teichmüller space and moduli space. In fact, the

Teichmüller space is the universal covering orbifold of the Riemann moduli space. A major focus of specific current mathematics research program is to explore this subject and its connections to hyperbolic geometry (HG), and the combinatorics of the complex of curves on a surface. So, Riemann surfaces and their MSs bridge several mathematical disciplines like HG, complex analysis, topology, and dynamics, and influences them all in return.

2.1.1 IG Example

As an application example, let us consider an application of IG to Image Processing. In the past, image models f were thought of having a scalar intensity $t \in \boldsymbol{R}$ at each pixel p (i.e. $f(p) = t$). By IG approach, we can have an univariate Gaussian probability distribution of intensities $n(\mu, \sigma^2) \in N$, i.e. image f is defined as the function:

$$f: \left\{ \begin{array}{c} \Omega \to N \\ p \mapsto n(\mu, \sigma^2) \end{array} \right\} \quad (1)$$

where Ω is the support space of pixels p (e.g. for 2D images $\Omega \subset \boldsymbol{Z}^2$) and N denotes the family of univariate Gaussian probability distribution functions (pdf). Most of current imaging sensors only produce single scalar values since the CCD (charge coupled Device) cameras typically integrates the light (arriving photons) during a given exposure time τ. To increase the signal-to-noise ration (SNR), exposure time is increased to $\tau' = \alpha\tau$, $\alpha > 1$. Let suppose that α is a positive integer number, this is equivalent to a multiple acquisition of α frames during τ for each frame (i.e., a kind of temporal oversampling). The standard approach only considers the sum (or average) of the multiple intensities [27], without taking into account the variance which is a basic estimator of the noise, useful for probabilistic image processing. By this approach, a gray scale image consists in considering that each pixel is described by the mean and the variance of the intensity distribution from its centered neighboring patch. In 2004, this model was used in [6] for computing local estimators which can be interpreted as pseudo-morphological operators. In fact, for a gray scale image, parameterized by the mean and the standard deviation of patches, it is possible to observe that the underlying geometry of this space of patches is not Euclidean, e.g. the geodesics are clearly curves [6]. Furthermore, this parametrization corresponds to one of the model of hyperbolic geometry (HG), the Poincaré upper-half plane (PUHP), specifically [26]. Henceforth, the corresponding image processing operators should be able to deal with Gaussian distributions-valued pixels. Specifically, morphological operators for images $f \in F(\Omega, N)$ involve that the space of Gaussian distributions (GD) N must be endowed of a partial ordering leading to a complete lattice structure. In practice, it means that given a set of Gaussian pdfs, we need to be able to define a Gaussian pdf which corresponds to the infimum (inf) of the set and another one to the supremum (sup). Mathematical morphology (MM) is a nonlinear image processing methodology based on the computation of sup/inf-convolution filters (i.e. dilation/erosion operators) in local neighborhoods [30]. MM is theoretically formulated in the framework of complete lattice and operators defined on them [17,29]. When only the supremum or the infimum are well defined, other morphological operators can be formulated in the framework of complete semilattices [18,24]. For further details, the interested reader is referred to [26, Chap 12]. This simple example was centered on the generalization of ordering structure for univariate GD. In the case of multivariate GD, we can consider to replace the PUHP model by the Siegel upper-half space (SUHS) [2].

2.2 CICT

Traditional Number Theory and modern Numeric Analysis use mono-directional interpretation (left-to-right, LTR) for $\boldsymbol{Q}$ Arithmetic single numeric group generator, so information entropy generation cannot be avoided in contemporary computational algorithm and application. On the contrary, according to CICT, it is quite simple to show information conservation and generator reversibility (right-to-left, RTL), by using basic considerations only. Traditional digital computational resources are unable to capture and to manage not only the full information content of a single Real Number $\boldsymbol{R}$, but even Rational Number $\boldsymbol{Q}$ is managed by information dissipation. In numeric representation of Rational Number $\boldsymbol{Q}$, rational proper quotient is represented by infinite repetition of a basic digit cycle, called "reptend" (the repeating decimal part). Let us consider fraction $1/D$, where D in $\mathbf{Z}$, called Egyptian fraction, with no loss of generality for common fraction. According to CICT, the first repetition of basic digit cycle of max length L corresponds to the first full scale interval where number information can be conserved completely, and is called "Representation Fundamental Domain" (RFD) [13]. Elementary number theory considerations give us worst case RFD word length $L = D - 1$ digits, if and only if 10 (by decimal base representation system, with no loss of generality) is a primitive root modulo D. Otherwise L is a factor of "D - 1." If the period of the corresponding repeating decimal to $1/D$ for prime D is equal to "$D - 1$", then the repeating decimal part is called "cyclic number" and D can be referred as "primitive number" or solid number (SN) [13] or "full reptend prime" elsewhere. Thus a SN is necessarily prime. It is a sufficient qualification. Conversely a prime number may not be a SN. In classical arithmetic long division algorithm (the one you learn to divide at elementary school), usual dominant result (quotient, Q) is important, and traditionally minority components (remainders, R) are always discarded. What a waste! In fact, Remainder R_L, at any division computation evolutive stage L, is the fixed multiplicative ratio of a formal power series associated to optimal decimal representations of $1/D$, at increasing arbitrary accuracy levels. In 2013, CICS showed that long arithmetic division minority components (Remainders, R), for long time concealed relational knowledge to theird ominant result (Quotient, Q), not only can always allow quotient regeneration from their remainder information to any arbitrary precision, but even to achieve information conservation and entropy minimization, in systems modeling and post-human cybernetic approaches [14,15]. According to CICT optimized infocentric worldview, symmetry properties play a fundamental role and affect word level structures and properties in analogous way to phoneme level and syllable level properties which create "double articulation" in human language, at least [20,21,25]. In case of figures or image sequences, Italian semiotician Umberto Eco (1932-) argued that "Iconic Language" has a "triple articulation": Iconic Figures, Semes (combinations of Iconic Figures), and Kinemorphs (combination of Semes), like in a classical movie. Therefore, traditional $\boldsymbol{Q}$ Arithmetic can be regarded as a highly sophisticated open logic, powerful and flexible optimized "OpeRational" (OR) LTR and RTL formal numeric language of languages, with self-defining consistent words and rules, starting from self-defined elementary generator and relation, based on recursively self-defining atom [13]. For instance, at any LTR computation stage, with remainder knowledge only, it is always possible to regenerate exact quotient and new remainder information at any arbitrary accuracy, with full information conservation. It is like to process tail information to regenerate the associated body information. Thanks to the above properties, the division algorithm can become free from trial and error like in Finite Segment P-adic representation systems, but with no usually associated coding burden. The rich operative scenario offered by combinatorial modular group theory is full of articulated solutions to information processing problems. One of the earliest presentations of a group by generator and relation was given by the Irish mathematician William Rowan Hamilton in 1856, in his Icosian Calculus, a presentation of the icosahedral group [7,16]. Every group has a presentation, and in fact many different presentations. A presentation is often the most compact way of describing the structure of the group. In abstract algebra, the "fundamental theorem of cyclic groups" states that every subgroup of a cyclic group G is cyclic. Moreover, the order k of any subgroup of a cyclic group G of order n is a divisor of n, and for each positive divisor k of n, the group G has exactly one subgroup of order k. This is just the first step to start an intriguing voyage from the concept of "presentation of a group" to the concept of "representation theory" for combinatorial modular group theory [8]. Furthermore, CICT sees rational geometric series as simple recursion sequences in a wider recursive operative framework where all algebraic recursion sequences of any countable higher order include all the lower order ones and they can be optimally mapped to rational number system $\boldsymbol{Q}$ OR representations and generating functions. For instance, arithmetic progression and Lucas sequences are recursion sequences of the second order. Lucas sequences are certain integer sequences that satisfy Lucas recurrence relation defined by polynomials $Un(P,Q)$ and $Vn(P,Q)$, where Un, Vn are specific polynomials and P, Q are fixed integer coefficients. Any other sequence satisfying this recurrence relation can be represented as a linear combination of the Lucas sequences $Un(P,Q)$ and $Vn(P,Q)$. Famous examples of Lucas sequences include the Fibonacci numbers, Mersenne numbers, Pell numbers, Lucas numbers, Jacobsthal numbers, and a superset of Fermat numbers. CICT is able to fold any recursion sequence of the first order into one digit number D_1, any recursion sequence of second order into a two digit number D_2, any recursion sequence of the third order into a three digit number D_3 and so on to higher orders. Then, you can interpret their asymptotic convergence ratios as increasing accuracy approximations to related asymptotic roots from corresponding first, second, third, ..., n-th order equations respectively. Thanks to this brand new knowledge and following this line of generative thinking, it is possible immediately to realize that traditional $\boldsymbol{Q}$ Arithmetic can be even interpreted, by new eyes, as a highly sophisticated open logic, powerful and flexible LTR and RTL evolutionary, generative, formal numeric language of languages, with self-defining consistent numeric words and rules, starting from elementary generator and relation (you get your specific formal numeric language by just simply choosing your most convenient numeric base to polynomially structure your information). American linguist, philosopher, cognitive scientist, logician, Avram Noam Chomsky's (1928-) Theory of Syntax came after his criticism of probabilistic associative models of word order in sentences by Markov process approaches, in 1957. As a matter of fact, since 1951, the inadequacy of probabilistic LTR models (Markov process) had already been noticed by American psychologist and behaviorist Karl Spencer Lashley (1890-1958), who anticipated Chomsky's arguments, by observing that probabilities between adjacent words in a sentence have little relation to grammaticality of the string. Ambiguity too provides a strong indication that sentences carry a structure. The treatment of numeric word, generator, relation and language by CICT largely draws its inspiration from many reliable research sources. Between them, the line of research started by Chomsky and French mathematician and Doctor of Western Medicine Marcel-Paul "Marco" Schützenberger (1920-1996) in the early 1960s occupies a singular place. Those fascinating viewpoints invite us to use a mind open logic approach to find new, more convenient solutions to old problems, always!

2.2.1 CICT Example

As an example, we compare stochastic vs. combinatorially optimized noise in so called "true-color image" (24 bit or 16,777,216 color variations). Usually, true color is defined to mean at least 256 shades of red, green, and blue fundamental components, for a total of at least 16,777,216 color variations. Traditional psychophysical experiment attained that the human eye can discriminate up to ten million colors [23]. Let us think of an image as a square matrix of N^2 cells (pixels). In this case, to achieve maximum Shannon's entropy, denoted by H(x), each cell must have one shade of color, out of M = 16,777,216 possible ones (M values), different from all the other ones in the matrix. Our image can be thought as a system of linear equations that can be an under- or overdetermined system. In the case of an underdetermined system there are more M values (shades of color) than available equations N^2 (available cells). Hence an image with $N^2 < M$ pixels is unable to exploit the full dynamic range of a 16,777,216-shades of color ideal random noise source generator. On the other end, an image with $N^2 > M$ pixels is an overdetermined system and, inevitably, at least two pixels will have the same value (shade of color), lowering the value of H(x). Then, if we like to maximize H(x), it is sufficient to have N = 4096 and to use a 4096 by 4096 pixel image to be sure to visualize an instance image out of all the possible color variations generated by an "ideal random" color source (i.e. $16{,}777{,}216^{16{,}777{,}216}$, really a quite huge transcomputational number [5]). An ideal random color source always produces an apparently grey image, and human eye is quite sensitive to different grey shades! In this case we use, for simplicity, a computer generated pseudo-random source from open source software. Then, Shannon's entropy for random color image is H(x) = 0.999292377044885, in double precision. Random generation is quite efficient and gets high value of H(x), getting closer to theoretical maximum of $H_t(x)$ = 1,0. According to CICT perspective [12], it is possible to generate a corresponding visibly similar image to previous one, by using the so called Solid Number (SN) approach [13]. SN "Word Family Group," discussed in [13], shows combinatorially optimized exponential cyclic sequence (OECS) properties which can be conveniently used to get interesting algorithm in a quite simple way. In this case, combinatorial optimization is easily achieved by finding the best SN *p* which allows an optimal encoding of image information. In fact, in this case the closest SN to 16,777,216 is SN = 16,777,259 to be sure to visualize an instance image out of all the combinatorially optimized $(2*2^2 *16, 777, 258)$ = 134,218,064 theoretical encoded color combinations. A corresponding visibly similar image to the previous one is computed by SN algorithm, deterministically. Now, by design, Shannon's entropy for combinatorial true color image is H(x) = 0.99999999993863, in double precision arithmetic. As a matter of fact, our SN closest choice to 16,777,216 has given a less-than-optimal entropy encoding for true-color image. Nevertheless, this H(x) value is quite close to the ideal one (less than 10^{-10} difference). It is interesting to note that the corresponding values of H(x) for combinatorially optimized image still largely outperform the generated one by pseudo-random noise source (more than six orders of magnitude difference). So, even a less-than-optimal CICT combinatorially optimized entropy encoding solution can do better than a traditional digitally pseudorandom generated one. Furthermore, according to classic information theory point of view, the corresponding image Memory FootPrint

(MFP) should be at its maximum value of 56,174 Kb (a little more than 56 Mb) and its information content, considered as random noise, could not be lossless compressed further. That is not the case, because, thanks to its associated combinatorial structure, its MFP can be lossless compressed to a minimum of 179 Kb code, at the expense of a longer processing time, by usual programming tools, supported by MS Visual Studio development framework. In this case, it is possible to achieve an image Lossless Compression Ratio (LCR) close to 314:1, more precisely 313.8212290:1 in single precision arithmetic. This is the final evidence to verify the combinatorially optimized information encoding by SN.

3 CICT Problem Solution

The first attempt to identify basic principles, to synergically articulate CICT by natural coupling to IG, for scientific research and application, has been developing at "Politecnico di Milano" since the end of last century. In 2013, the basic principles on CICT, from discrete system parameter and generator, appeared in literature. To better understand the CICT fundamental relationship that tie together numeric body information of divergent and convergent monotonic power series in any base (in this case decimal, with no loss of generality) with D ending by digit 9 is given by the following correspondence equation (see [13]):

$$\frac{1}{D} = \sum_{k=0}^{\infty} \frac{1}{10^W} \left(\frac{\overline{D}}{10^W}\right)^k \Rightarrow\Leftarrow Div\left(\frac{1}{D}\right) = \sum_{k=0}^{\infty} (D+1)^k \quad (2)$$

where $\overline{D}$ is the additive 10^W complement of D, i.e. $\overline{D} = (10^W - D)$, W is the word representation precision length of the denominator D and "Div" means "Divergence of". Further generalizations related to D ending by digit 1 or 3 or 7 are straightforward. Furthermore, When $\overline{D} > D$ the formal power series on the left of eq.(2) can be rescaled mod$D$, to give multiple convergence paths to $1/D$, but with different "convergence speeds." The total number of allowed convergent paths, as monotonic power series, is given by the corresponding Q_L value, at the considered accuracy level L. So, increasing the level of representation accuracy, the total number of allowed convergent paths to $1/D$, as monotonic power series (as allowed conservative paths), increases accordingly and can be counted exactly, and so on, till maximum machine word length and beyond: like discrete quantum paths denser and denser to one another, towards a never ending "blending quantum continuum," by a TD perspective. Rational representations are able to capture two different type of information at the same time, modulus (usual quotient information) and associated inner or intrinsic period information which an inner phase can be computed from. So, rational information can be better thought to be isomorphic to vector information rather than to usual scalar one, at least. Furthermore, our knowledge of RFD Q_L and corresponding RFD R_L can allow reversing numeric power convergent sequence to its corresponding numeric power divergent sequence uniquely. Reversing a convergence to a divergence and vice-versa is the basic property to reach information conservation, i.e. information reversibility, as from eq.(2). CICT results have been presented in term of classical power series to show the close relationships to classical and modern control theory approaches for causal continuous-time and discrete-time linear systems. Usually, the continuous Laplace transform is in Cartesian coordinates where the *x-axis* is the Real axis and the discrete Z-transform is in circular coordinates, where the *Rho-axis* is mapping the Real axis. By using this approach, it is possible to generate LTR and RTL remainder sequences that show same quotient body information (arbitrary-scale periodic) and specific quotient head and tail information to compute deterministic boundary values, to sustain body periodicity with no information dissipation (full information conservation and reversibility) [13]. Traditional rational number system $\boldsymbol{Q}$ properties allow to compute evolutive irreducible co-domain for every computational operative domain used. Then, all computational information usually lost by using traditional computational approach can be captured and recovered by a corresponding complementary co-domain, step-by-step. Then co-domain information can be used to correct any computed result, achieving computational information conservation. CICT is offering an operational optimized infocentric worldview [12]. CICT fundamental relation (eq.(2)) allows to focus our attention on combinatorially optimized number pattern generated by LTR or RTL phased generators and by convergent or divergent power series with no further arbitrary constraints on elementary generator and relation. Thanks to subgroup interplay and intrinsic phase specification through polycyclic relations in each SN remainder sequence, word inner generator combinatorial structure can be arranged for "pairing" and "fixed point" properties for digit group with the same word length. As a matter of fact, those properties ("pairing" and "fixed point") are just the operational manifestation of universal categorical irreducible dichotomy hard-wired into integer digit and digit group themselves (i.e. "evenness" and "oddness") and to higher level structures (i.e. "correspondence" and "incidence"). Actually, since space is limited, the discussion here will not be extended further to the subgroup interplay of the family group and polycyclic groups. We refer the interested reader to good general references on polycyclic groups [28,33].

4 Conclusion

The final result is CICT new awareness of a hyperbolic framework of coded heterogeneous hyperbolic structures, underlying the familiar Euclidean surface representation system. CICT emerged from the study of the geometrical structure of a discrete manifold of ordered hyperbolic substructures, coded by formal power series, under the criterion of evolutive structural invariance at arbitrary precision. It defines an arbitrary-scaling discrete Riemannian manifold uniquely, under HG metric, that, for arbitrary finite point accuracy level L going to infinity (exact solution theoretically), is isomorphic to traditional IG Riemannian manifold. In other words, HG can describe a projective relativistic geometry directly hardwired into elementary arithmetic long division remainder sequences, offering many competitive computational advantages over traditional Euclidean approach. It turns out that, while free generator exponentially growing sequences can be divergent or convergent, their closures can be defined in terms of polynomials. Furthermore, combinatorially OECS have strong connection even to classic DFT algorithmic structure for discrete data, Number-Theoretic Transform (NTT), Laplace and Mellin Transforms [12]. In this way, even simple scalar moduli can emerge out from sequences of phased generators. CICT can help to reach a unified vision to many current biophysics and physics problems and to find their optimized solutions quite easily. Expected impacts are multifarious and quite articulated at different system scale level. One of the first practical result was that usual elementary arithmetic long division remainder sequences can be even interpreted as combinatorially optimized coding sequences for hyperbolic geometric structures, as point on a discrete Riemannian manifold, under HG metric, indistinguishable from traditional random noise sources by classical Shannon entropy, and contemporary most advanced instrumentation systems, as discussed in Section 2.2.1. Specifically, CICT showed that classical Shannon entropy computation is completely unable to reliably discriminate so called computational "random noise" from any combinatorially optimized encoded message by OECS, called "deterministic noise" in [12]. As a matter of fact, for any free generator, CICT can provide us with an "ecoco-domain" multiscale evolutive structured family of sequences that can be used for checking for the presence of a specific generator in laboratory or system "background noise" [12]. Following CICT approach, it is possible even to extend the classic Shannon entropy concept to arrive to a stronger and specific "Coherent Shannon entropy" (CSE) approach. Second result was to realize that classical experimental observation process, even in highly ideal operative controlled condition, like the one achieved in contemporary most sophisticated and advanced experimental laboratories like CERN, can capture just a small fraction only, with misplaced precision, of overall ideally available information from unique experiment. The remaining part is lost and inevitably added to something we call "background noise" or "random noise" usually, in every scientific experimental endeavor. CICT can help us to develop strategies to gather much more reliable experimental information from single experiment and to conserve overall system information. In this way, coherent representation precision leads to information conservation and clarity. Specifically, high reliability organization (HRO), mission critical project (MCP) system, very low technological risk (VLTR) and crisis management (CM) system will be highly benefitted mostly by these new techniques. The latest CICT claim has been that the "external" world real system physical manifestation properties and related human perception are HG representation based, while Euclidean approximated locally. Furthermore, the fundamental play of human information observation interaction with an external world representation is related by the different manifestation and representation properties of a unique fundamental computational information structuring principle: the Kelvin Transform (KT). KT is key to efficient information representation, structuring "external space" information to an "internal representation" and vice-versa by inversive geometry.

References:

[1] Amari, Shun'ichi, *Methods of Information Geometry*, Oxford University Press 2000. (Amari S., Nagaoka H., *Methods of Information Geometry*, originally published in Japanese in 1993 and published in English in 2000 with the American Mathematical Society, AMS).

[2] Barbaresco, F., Information geometry of covariance matrix: cartan-siegel homogeneous bounded domains, Mostow/Berger fibration and fréchet median. In: Nielsen F., Bhatia R. (eds.), *Matrix Information Geometry*, 2013, pp.199-255, Heidelberg: Springer.

[3] Bohner, M., Peterson, A., *Dynamic Equations on Time Scales*, Berlin: Birkhäuser, 2001.

[4] Bohner, A., Peterson A., Eds., *Advances in Dynamic Equations on Time Scales*, Berlin: Birkhäuser, 2003.

[5] Bremermann, H.: Quantum noise and information, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability,* Univ. of California Press, Berkeley, California, 1965.

[6] Căliman, A., Ivanovici, M., Richard, N. Probabilistic pseudo-morphology for grayscale and color images. *Pattern Recognition*, No.47, 2004, pp.721-35.

[7] Chandler, B. M., *The history of combinatorial group theory: A case study in the history of ideas, studies in the history of mathematics and physical sciences* (1st ed.), Springer, 1982, pp.387–407.

[8] Curtis, C. W., Pioneers of Representation Theory: Frobenius, Burnside, Schur, and Brauer, *History of Mathematics*, American Mathematical Society, Providence, 2003.

[9] Eskin A. and Mirzakhani M., Invariant and stationary measures for the SL2(R) action on moduli space, preprint 2013; see arXiv:1302.3320.

[10] Eskin A., Mirzakhani M., and Mohammadi A., Isolation, equidistribution, and orbit closures for the SL2(R) action on moduli space, preprint, 2013; see arXiv:1305.3015.

[11] Eulero L., *Institutiones Calculi Differentialis*, Academia Imperialis Scientiarum Petropolitanae, 1775.

[12] Fiorini R., How Random is Your Tomographic Noise? A Number Theoretic Transform (NTT) Approach, *Fundamenta Infomaticae*, Vol.135, No.3-4. 2014, pp.1-36.

[13] Fiorini, R. A., Laguteta, G.: Discrete Tomography Data Footprint Reduction by Information Conservation, *Fundamenta Informaticae*, Vol.125, No.3-4, 2013, pp. 261–272.

[14] Fiorini R. A., Santacroce G., Economic Competitivity in Healthcare Safety Management by Biomedical Cybernetics ALS, *Proc. International Symposium, The Economic Crisis: Time For A Paradigm Shift - Towards a Systems Approach*, Universitat de València, January 24-25, 2013.

[15] Fiorini R. A., Santacroce G., Systems Science and Biomedical Cybernetics for Healthcare Safety Management, *Proc. International Symposium, The Economic Crisis: Time For A Paradigm Shift - Towards a Systems Approach*, Universitat de València, January 24-25, 2013.

[16] Hamilton W. R., Memorandum respecting a new System of Roots of Unity, *Philosophical Magazine*, No.12, 1856, pp. 446.

[17] Heijmans HJAM, *Morphological Image Operators*, Boston: Academic Press, 1994.

[18] Heijmans HJAM, Keshet R., Inf-semilattice approach to self-dual morphology, *J. Math. Imaging Vis.*, Vpl.17, No.1, 2002, pp. 55-80.

[19] Hilger S., Ein Masskettenkalkül mit Anwendung auf Zentrumsmannigfaltigkeiten, PhD Thesis, Universität Würzburg, 1988.

[20] Hockett C., *A Course in Modern Linguistics*, Macmillan: New York, 1958.

[21] Hockett C., The Origin of Speech, *Scientific American*, No.203, 1960, pp.89–96.

[22] Jordan C., *Calculus of Finite Differences*, Budapest, 1939.

[23] Judd D., Wyszecki G., *Color in Business, Science and Industry*, Wiley Series in Pure and Applied Optics (third ed.), Berkeley, California, 1975.

[24] Keshet R., Mathematical morphology on complete semilattices and its applications to image processing. *Fundamenta Informaticae*, No.41, 2000, pp. 33-56.

[25] Manovich L., The Language of New Media, MIT Press, Cambridge, 2001.

[26] Nielsen F., Ed., *Geometric Theory of Information*, Springer International Publishing Switzerland, 2014.

[27] Sbaiz L., Yang F., Charbon E. Süsstrunk S., Vetterli M., The gigavision camera. In: *Proceedings of IEEE ICASSP'09*, 2009, pp.1093-1096.

[28] Segal D., *Polycyclic groups*, Cambridge University Press, Cambridge, 1983.

[29] Serra J., *Image Analysis and Mathematical Morphology, Vol. II: theoretical advances*, London:Academic Press 1988.

[30] Soille P., *Morphological Image Analysis*, Berlin: Springer-Verlag, 1999.

[31] Stirling J., *Methodus Differentialis*, London 1730.

[32] Taylor B., *Methodus Incrementorum*, London, 1717.

[33] Wehrfritz BAF., *Group and Ring Theoretic Ptoperties of Polycyclic Groups*, London: Springer-Verlag, 2009.