

Information, Representation, and Structure

Subhash Kak

Oklahoma State University, Stillwater, OK 74078, USA

Email: subhash.kak@okstate.edu

Abstract. This paper investigates the consequences of the information-theoretic result that representations of numbers in base- e are most efficient. Since theories on complex system behavior in both natural and physical systems assume that Nature is optimal, as is done, for example, in the principle of least action, natural representations must be to the base e . Another way to interpret this fact is to take e as the information dimension of the data space. Some implications of this noninteger dimensionality are investigated. The approximate equivalent to such a space is the Menger sponge in which the recursion is taken to be random.

Keywords: dimensionality, fractals, noninteger spaces, informational logic

1. Introduction

Attempts to reconcile epistemic and ontic interpretations can help discover the implicit assumptions of theory [1]. Although information privileges the epistemic view, its inherent tension with the ontological view is ignored by a focus on probabilities [2] or on measurement [3]. Viewing information in the distribution of objects to different scales as in the large-scale structure of the universe [4-5] or in theories on the relationship of gravitation to quantum mechanics [6-7] from the perspective of the dimensionality of space may present new insights.

This paper is an attempt at bridging the epistemic and the ontic views by an explicit consideration of the nature of space by examining it with respect to information at a fundamental level. We do so by accepting that an idea like that of the principle of least action that endows Nature with “optimal” behavior also applies to information obtained through observation.

In the most abstract setting, the intuition of space may be viewed as emerging from a general mapping of observed data that is recognized through our cognitive structures as the familiar three dimensions. Since the most basic mapping is the representation of numbers for which the most efficient base is e [9], optimal mapping requires that space have the same dimension. We provide a constructive quantum mechanical proof of this assertion. The noninteger dimension of $e=2.718...$ means that planar structures

have substantial probability, and this may be viewed as a consequence of an intrinsic dynamics related to space.

2. Information dimension

One can think of information dimension of an object M as the amount of information necessary to specify the position of a point belonging to M , which is related to the representation of information to an appropriate base. A solid is three-dimensional because one needs three coordinates to specify any point inside.

Proposition 1. The amount of information required to specify a point in a space represents the *information dimension* of that space.

Not all physical shapes require integer dimensions. To see this, consider measuring a shape by a cube and then use smaller cubes with the scaling factor of ε , so that if N such smaller cubes are to be used, then we can write [10,11]:

$$N = \varepsilon^{-D} \quad (1)$$

The dimensionality associated with the shape will then be:

$$D = \lim_{\varepsilon \rightarrow \infty} \left(-\frac{\log N}{\log \varepsilon} \right) \quad (2)$$

Now we ask the question of the dimensionality of a general space. If space were d -dimensional, we could label the dimensions as 1, 2, 3, ... d . The probability of the use of each of the d dimensions may be taken to be the same and equal to $1/d$, and the information associated with each dimension is $\log d$.

Clearly, the location information will be greater if the dimensionality is higher. But the increase in information must be squared off against the extra burden entailed by the use of the larger set of dimensions. For two-dimensional space, the information value of each dimension is $\ln 2 = 0.693$ nats (=1 bit); for three-dimensional space, it is 1.099 nats (=1.585 bits); and for ten-dimensional space, it is 2.303 nats (=3.322 bits).

The efficiency of the representation of information per dimension is:

$$E(d) = \frac{\ln d}{d} \quad (3)$$

Its maximum value is obtained by taking the derivative of $E(d)$ and equating that to zero, which yields $d_{optimum} = e = 2.71828...$ In other words:

Theorem 1. The optimal number of information dimensions associated with space is e .

Table 1 gives the value of $E(d)$ in bits for d ranging from 2 to 10, together with the additional value for the optimum $d=e$ (Figure 1).

Table 1. Efficiency of space of dimensions ranging from 2 to 10

d	2	e	3	4	5	8	10
$E(d)$ bits	0.500	0.531	0.528	0.500	0.465	0.375	0.331

The efficiency for e dimensions is 0.531 bits whereas for $d=3$ it is 0.528 bits. The next best value coming at the bases 2 and 4 (where it is 0.500 bits). The three-dimensional space is off from the e -dimensional optimal space by about 0.003, or about 0.6 percent. One may propose that since our cognitions are based on counting, we associate the nearest integer space of 3 dimensions to space.

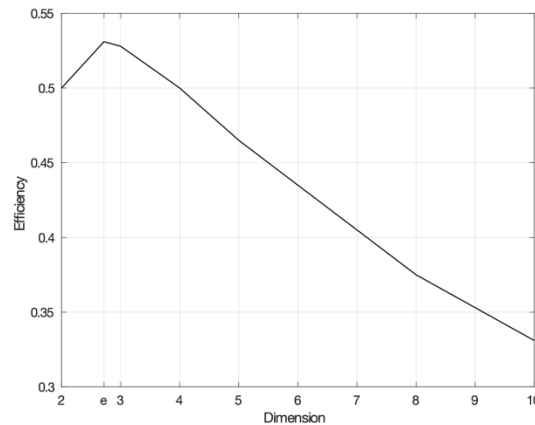


Figure 1. Efficiency of dimensions, 2 through 10

Proposition 2. The information dimension of a physical space is its physical dimensionality.

To visualize the e -dimensional space, one may consider it as shape or an abstract conception that is structured into different projections for small scale and large-scale phenomena. Its relation to the three-dimensional space is most clear at intermediate scales like the ones we encounter in everyday life. The terrestrial observer will see the large-scale as well as small-scale structures as a continuation of the nature of space at the terrestrial level.

Axiomatic foundations of non-integer spaces have been given by Wilson [12] and Stillinger [13]. In addition to the usual axioms that apply to Euclidean spaces, Stillinger needed to add two more axioms: one related to topology and another to integration measure. He further proposed that the realness of a fractional space less than 3 could be checked by experiments of sphere-packing but he acknowledged that to carry out such an experiment will not be an easy matter due to extreme constraints on accuracy.

Theorem 2. A unit cube in an e -dimensional space has $e^e \sim 15.154 \dots$ sub-cubes each of side $1/e$.

Proof. A unit 2-cube in a 2-dimensional space is a square and the total number of sub-cubes of side $1/2$ is $2^2 = 4$, and a unit 3-cube in a 3-dimensional space has sub-cubes of side $1/3$ that equal $3^3 = 27$; generalizing, we get the result.

The e -dimensional space is smaller than the 3-dimensional space. How it maps into the larger 3-dimensional space, consider how many sub-cubes of e -dimensions $1/e$ can be fitted in a 3-dimensional unit cube.

Theorem 3. The number of sub-cubes of side $1/e$ that go into a 3-dimensional unit cube is 20.085...

Proof. The volume of sub-cubes of side $1/e$ in a 3-dimensional space is e^{-3} . Therefore, the number of such sub-cubes that will go into a volume of 1 is $e^3 = 20.085 \dots$

Seen from the perspective of ordinary 3-dimensional space, the number of sub-cubes after n iterative operations is e^{3n} . Therefore, applying formula (2), we get the value of the dimension to be

$$D = \lim_{\varepsilon \rightarrow \infty} \left(-\frac{\log e^{3n}}{\log 3^{-n}} \right) = e. \quad (4)$$

The other implication of this recursive structure is that we are speaking of fractal or scale-invariant systems, examples of which are the Mandelbrot set and the Pythagoras tree (Figure 2), and many natural structures such as the tree fern tree or the snail shell.

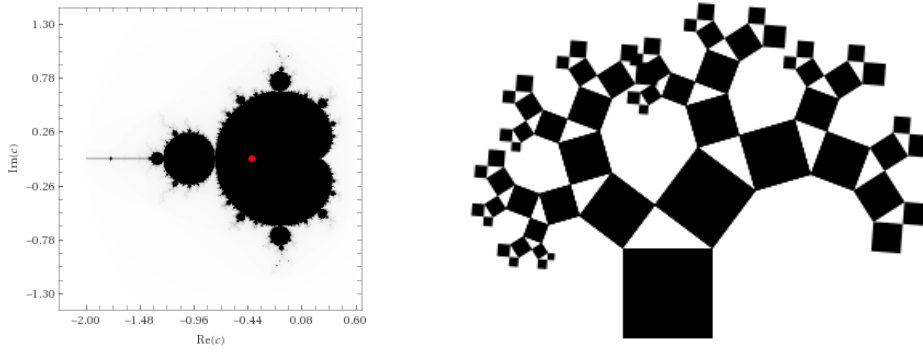


Figure 2. Mandelbrot set (left), and Pythagoras tree (right)

3. An approximation to the e-dimensional space

Now consider a deterministic model to help with the visualization of an e-dimensional space. We need a scaling transformation by which $e^3 \sim 20$ e-dimensional sub-cubes are seen as a subset of the 27 three-dimensional sub-cubes. This is done iteratively.

Taking a cue from quantum mechanics, we may speak of a *creation operator* that maps space into structure. Specifically, the $e^3 \sim 20$ sub-cubes of the e-space may be mapped by an appropriate iterative transformation in the 3-space. In other words, we need a mapping that takes us from the smaller sub-set of 20 sub-cubes to the larger 3-space of 27 sub-cubes.

Since a cube has six sides, this may be done by any mapping where one dark sub-cube is removed randomly from each of the six sides together with the one at the center. Although it is done uniformly in Figure 3, there is no reason why it cannot be done randomly. The seven extra sub-cubes represent the effect of the creation operator.

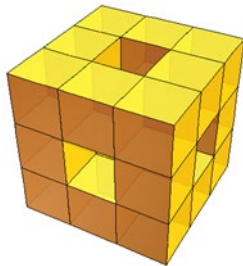


Figure 3. A mapping of 20 e-space sub-cubes into the unit cube

Let N_n be the number of dark boxes, L_n the length of a side of a light sub-cube, and V_n the fractional volume of the dark cubes after the n th iteration, then

$$N_n = 20^n \quad (5)$$

$$L_n = \left(\frac{1}{3}\right)^n = 3^{-n} \quad (6)$$

$$V_n = \left(\frac{20}{27}\right)^n \quad (7)$$

The dimension of such an iterative system will be:

$$D = \lim_{\varepsilon \rightarrow \infty} \left(-\frac{\log N_n}{\log L_n} \right) \quad (8)$$

$$= \frac{\ln 20}{\ln 3} = 2.7268 \dots \quad (9)$$

which is quite close to $e = 2.71828\dots$

The difference between the values of D and e is only 0.3% and, therefore, it is a good deterministic model to visualize the e -dimensional space. The difference of 0.3% was due to the fact that we used the integer value of 20 rather than the exact value of $e^3 = 20.085 \dots$

4. Iterative construction

Since the three-dimensional system is almost as efficient as the e -dimensional one, one would like to begin with an appropriate one-dimensional set and then generalize that to three dimensions.

It is surprising that the random mapping described above may be derived by the use of the one-dimensional Cantor set [14], of two kinds of elements that we label *dark* and *light* (Figure 4) One starts with a line segment of unit length that is *dark*, converts the middle third to *light*, then converts the middle thirds from the remaining two dark segments to *light*, and so on.

Formally, the Cantor set at the n th iteration, C_n , is:

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3} \right), \text{ for } n \geq 1, \text{ and } C_0 = [0,1] \quad (10)$$



Figure 4. The Cantor set through several iterations

Generalizing this to two dimensions. each cell may be transformed according to the rule of the Sierpiński carpet [15] of Figure 5 in which the light areas remain unchanged and each dark area is mapped into itself excepting that the middle sub-square (out of nine) is changed from dark to light:

$$0 \rightarrow \begin{vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}; 1 \rightarrow \begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{vmatrix} \quad (11)$$

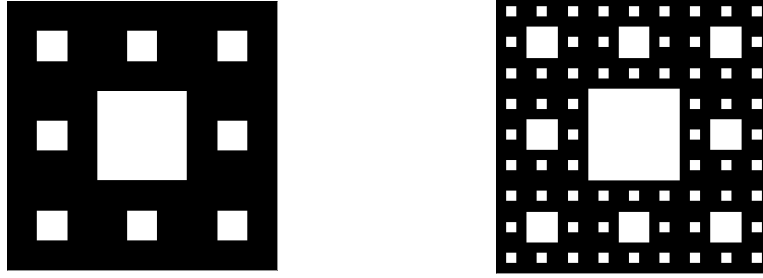


Figure 5. Iterative mapping of the Sierpiński carpet

Equivalently, one might use a random mapping where the 0 of the mapping for 1 is placed randomly in the matrix in the right-hand side. An example of this is given in Figure 6 where it was placed in the middle left corner in the first iteration and variously in the second iteration, and so on.

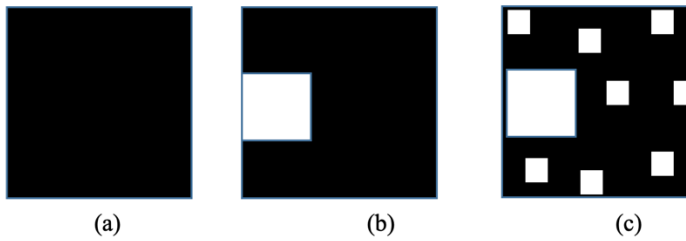


Figure 6. The two-dimensional iterative random mapping

The three-dimensional generalization of the Cantor set is the Menger sponge [16] whose first iteration was shown in Figure 2 with the second and third iterations shown in Figure 7:

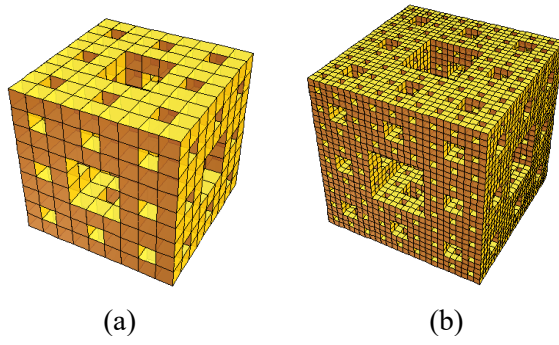


Figure 7. The evolution of the Menger sponge: (a) second iteration; (b) third iteration

Consider a model universe in which the *dark* cubes at each iteration creates 7 sub-cubes of *light* cubes. One may replace the mapping used in Figure 2 by modified forms. For example: (a) the *light* sub-cubes are not equal in size to the ones that remain *dark* (in other words, have a slower rate of conversion of dark to light), and (b) the transformation is done sequentially rather than at the same time for all regions. Doing so sequentially associates the process with a time variable. A steady-state model in which light and dark regions transform into each other at different rates may also be considered.

At the first iteration, the light density is $\frac{7}{27}$, or approximately 0.26. In the second iteration, it will be $\frac{7}{27} + \frac{20}{27} \times \frac{7}{27}$ which is approximately 0.45 and so on as shown in Figure 8.

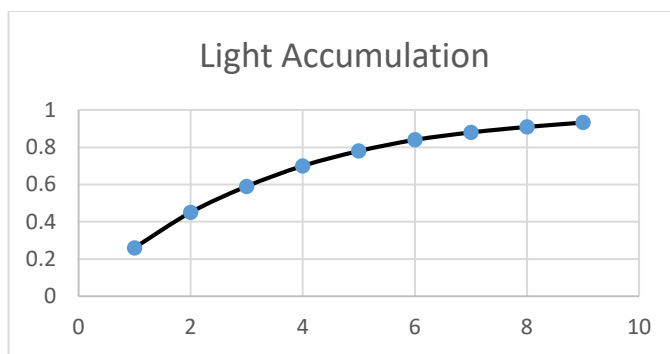


Figure 8. The increase in the density of light in the model universe

There need to be further investigations of stochastic versions of the model universe so that one can consider additional empirical aspects.

5. Measurements and dimensions

Consider a system being interrogated by the observer by means of an interaction. This interrogation will be visualized by means of the transfer of a state or node of the experimenter into the system.

Let *variety*, V , represent the novelty associated with each value of the data in terms of its diversity. We can associate variety not only each data point but associate it also with the original system as well as the aggregate system that includes the measurement node. One can compute V by looking at the novelty of combinations associated with the new information residing in the data vectors.

An intuitively satisfactory way to define dimensionality is to compute the infimum of the *variety* that can be associated with all the object (or node)-states within the system. The dimension, D , of the data is the minimum possible value of V .

Let the classical system, C , consist of n objects (or nodes in a network) and the observer, that is the measurement device, consist of m objects (or nodes). The information associated with the system will be maximized when the objects are associated with the equal probability.

The entropy for the system is its Shannon entropy:

$$H(C) = \ln n \quad (12)$$

The entropy of the $C+M$, that is the classical system together with the measurement apparatus is

$$H(C + M) = \ln(n + m) \quad (13)$$

The information obtained by the observer upon measurement is

$$\begin{aligned} H(C + M) - H(C) &= \ln(n + m) - \ln n \\ &= -\ln \frac{n+m}{n} \end{aligned} \quad (14)$$

Since information is the logarithm of the total possibilities, the variety associated with each data point is:

$$V_{each\ data\ point} = \left(\frac{n+m}{n}\right) \quad (15)$$

Since there are n data points,

$$V_{data} = \left(\frac{n+m}{n}\right)^n \quad (16)$$

An intuitively satisfactory way to define dimensionality is to compute the infimum of the *variety*, V , that can be associated with all the object (or node)-states within the system.

Definition. The dimension, D , of the data is the minimum possible value of V .

The dimension of all linear data will be one and that of data associated with a plane will be two. This idea of dimension derived from information considerations can be viewed as being consistent with its intuitive meaning [9].

The true variety of the data is obtained when $n \rightarrow \infty$:

$$V = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e^m. \quad (17)$$

Considering $n=1$, $m=1$ the space has a dimension of 2, for it is associated with the pair of states associated with the system and the observer. With $n=2$, $D=2.25$, that indicate correlations between the two objects and the one observer. Beyond this the value builds up to 2.718 as shown in Figure 9.

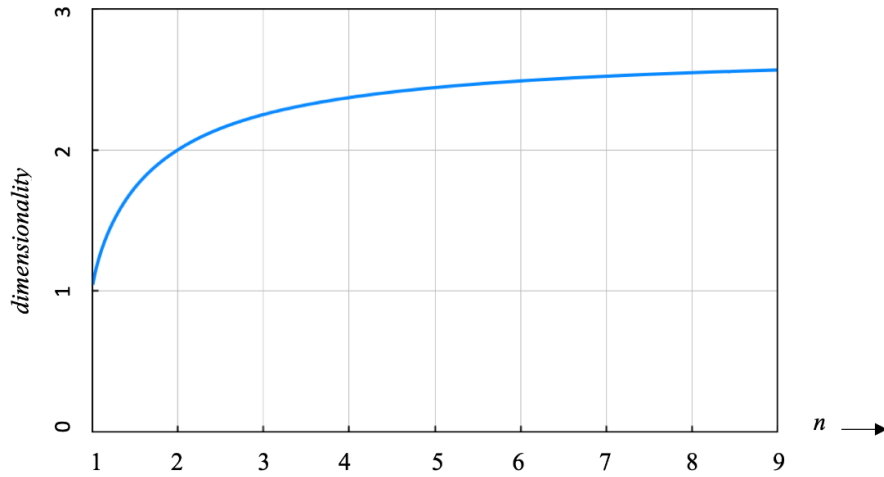


Figure 9. Dimensionality versus number of system states

$$D = \min_m \{e^m\} = e^m \quad (18)$$

This result is identical to that obtained on probabilistic grounds for number representation systems where it was shown that number representation to the base e is optimal [9]. That previous paper did not present the physical intuition behind this non-integer dimension, and now we have addressed that issue.

The information obtained by the measurement apparatus will increase exponentially with the capacity of the apparatus.

6. Minimum dimensionality of quantum data

In the quantum case we must first specify the way the measurement is performed. The measurement is an interaction between the measured system S and the measuring apparatus M .

Before the interaction, M is prepared in a ready-to-measure state $|p_0\rangle$, eigenvector of the pointer observable P of M , and the state of S is a superposition of the eigenstates $|a_i\rangle$ of an observable A of S . The interaction introduces a correlation between the eigenstates $|a_i\rangle$ of A and the eigenstates $|p_i\rangle$ of P :

$$|\Psi_0\rangle = \sum_i \alpha_i |a_i\rangle \otimes |p_0\rangle \rightarrow |\Psi_1\rangle = \sum_i \alpha_i |a_i\rangle \otimes |p_i\rangle \quad (19)$$

In the orthodox Copenhagen Interpretation, the pure state $|\Psi\rangle$ is assumed to “collapse” to one of the components of the superposition, say $|a_k\rangle \otimes |p_k\rangle$, with probability $|\alpha_k|^2$.

The state of the composite system after measurement is represented by the mixture ρ^c :

$$\rho^c = \sum_i |\alpha_i|^2 |a_i\rangle \otimes |p_i\rangle \langle a_i| \otimes \langle p_i| \quad (20)$$

where the probabilities $|\alpha_i|^2$ represent expectations associated with different eigenstates.

Let the system S consist of n states and the measuring apparatus M , which defines the observer, consist of m states. The composite system $S+M$ has $n+m$ states.

The entropy associated with the system S is [17]:

$$S(\rho) = -\text{tr}(\rho \ln \rho) \quad (21)$$

The maximum value of this is $\ln n$ when all eigenstates are equally probable.

The maximum entropy of the system S before the measurement is $\ln n$, and of the composite system $S+M$ after the measurement is $\ln(n+m)$.

The measurement is associated with a change in entropy of S that equals:

$$\ln(n+m) - \ln n = \ln \frac{(n+m)}{n} \quad (22)$$

The proof thereafter is identical to one for the classical case shown above.

The least amount of information, D , required to specify a point in a space (or states of the system) is the information dimension of that space. It measures the span associated with the data. It is quite clear that the minimum value of R is obtained when $k=1$.

Statistical view of D

In contrast to the implicit assumption that object and observer are apart in the classical case, in the quantum case when $n=1$, the space has a dimension of 2, for it is associated with the pair of states associated with the system and the observer. With $n=2$, $D=2.25$, that indicate correlations between the two objects and the one observer. Beyond this the value builds up to 2.718 as shown in Figure 9.

Since information is a statistical measure, this may be interpreted to mean that on an average one requires D pieces of information. A dimension of $D=2.718$ is closer to a three-dimensional system than a plane as shown in Figure 10. Let the probability of structures that are 2-dimensional be p . Solving for p , we obtain that it is 0.282.

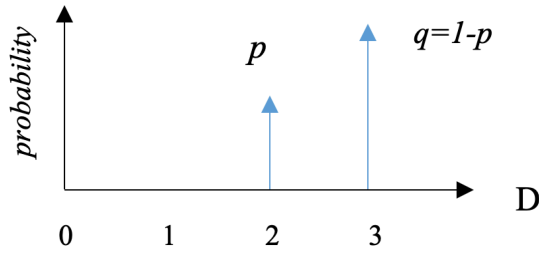


Figure 10. The dimensional probabilities for $D=e$

Roughly speaking this means that on an average 28% of the structures will be effectively planar. If one were to imagine such a system with a random initial distribution of objects, after sufficient of time, it would achieve a distribution where about 28% of the objects will be in planar arrangement.

But this is only possible if a dynamic can be associated with the contents of the space. This dynamic will make objects come closer together with time. In other words, a noninteger space is associated with an attraction field, and this means that the properties of integer and noninteger spaces are very different.

Since the surface area is related to r^2 , the attraction force will be inversely proportional to the square of the separation. In the general case, the space associated with the data cannot be two-dimensional. If the space were 2-dimensional with $1 < D < 2$, the proportionality with respect to $1/r$ would lead to infinite attraction at each point because the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots = \sum_{i=1}^{\infty} \frac{1}{i} \quad (23)$$

is divergent.

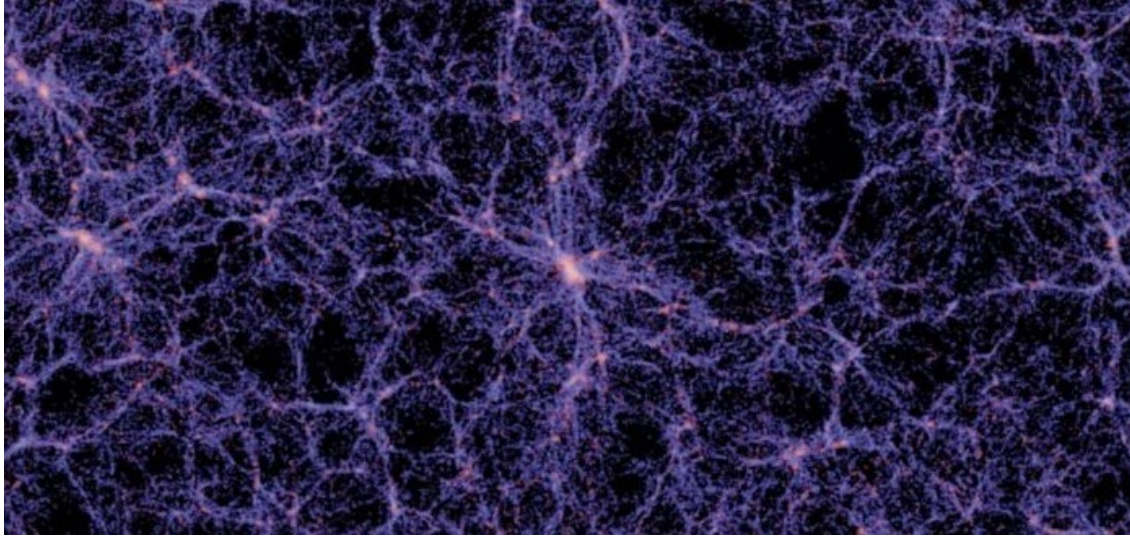


Figure 11. The distribution of matter in the universe [22]

We add that fractal dimensions have been observed in the large-scale structure of the universe [18-21]. Figure 11 presents an image of the distribution of matter in the universe generated by a simulation run modeling led by researchers at the U.S. Department of Energy's Argonne National Laboratory showing clearly its self-similar or fractal characteristics [22].

7. Conclusions

We examined implications of the information-theoretic result that optimal representation requires that the corresponding data space have an information dimension of e . This corresponds to just over 20 sub-cubes of side $1/e$ in the three-dimensional unit cube. We provided a constructive classical and quantum mechanical proof of this assertion. We proposed an approximation to this space in terms of a random recursive Menger sponge, whose fractal nature may be seen across different scales.

REFERENCES

1. Kak, S., Epistemic view of quantum communication. In Quantum Foundations, Probability and Information, A. Khrennikov and B. Toni (editors). Springer-Verlag International, 2018, pp. 119-128.
2. Zurek W.H. Decoherence, einselection, and the quantum origins of the classical. *Rev Mod Phys* 75, 715-775 (2003)
3. Kak, S. From the no-signaling theorem to veiled non-locality. *NeuroQuantology* 12, 12-20 (2014)

4. Misner, C.W., Thorne, K.S. and Wheeler, J.A.: Gravitation. Freeman, San Francisco, (1973).
5. Peebles, P. J. E.: The large-scale structure of the universe. Princeton University Press. (1980)
6. Milgrom, M. Alternatives to dark matter. Comments Astrophysics 13 215 (1989).
7. Ashtekar, A., Rovelli, C., Smolin. L.: Gravitons and loops, Physical Review D44, 1740 (1991).
8. Rovelli, C. Quantum Gravity. Cambridge University Press. (2004).
9. Kak, S. The base-e representation of numbers and the power law. Circuits Systems and Signal Processing, 2020 (in press); also arXiv:2003.02345 (2020).
10. Mandelbrot, B.B. The Fractal Geometry of Nature. Macmillan (1983)
11. Falconer, K. Fractal Geometry: Mathematical Foundations and Applications. John Wiley (2003)
12. Wilson, K.G. and Kogut, J., The renormalization group and the q expansion. Physics Reports 12, 75-199 (1974).
13. Stillinger, F.H., Axiomatic basis for spaces with noninteger dimensions. Journal of Mathematical Physics, 18: 1224- 1234 (1977).
14. Mattila, P. Geometry of Sets and Measures in Euclidean Space: Fractals and rectifiability. Cambridge University Press (1999)
15. Whyburn, G. Topological chcracterization of the Sierpinski curve. Fund. Math. 45, 320–324 (1958)
16. Iwaniec, T. and Martin, G. Geometric function theory and non-linear analysis. The Clarendon Press Oxford University Press (2001)
17. Kak, S., State ensembles and quantum entropy. Int. J. Theoretical Physics 55, 3017-3026 (2016).
18. Einasto, J., Htsi, G., Kuutma, T., Einasto, M.: On fractal properties of the cosmic web. arXiv:2002.02813 (2020)
19. Tarasov, V.E., Anisotropic fractal media by vector calculus in non-integer dimensional space. Journal of Mathematical Physics 55, 083510 (2014).
20. Larson, R. B., Turbulence and star formation in molecular clouds. Mon. Not. R. Astron. Soc. 194, 809–826 (1981).
21. de Vega, H., Snchez, N. & Combes, F., Self-gravity as an explanation of the fractal structure of the interstellar medium. Nature 383, 56–58 (1996).
22. Heitmann, K. et al., The Q Continuum Simulation: Harnessing the Power of GPU Accelerated Supercomputers. The Astrophysical Journal Supplement Series, Volume 219, Issue 2, article id. 34, 13 pp. (2015).

International Conference on Recent Trends in Mathematics and Its Applications to Graphs, Networks and Petri Nets, New Delhi, July 20-25, 2020.