UCLA

UCLA Previously Published Works

Title

Black holes and optimum coding

Permalink

<https://escholarship.org/uc/item/1mk5r6hx>

Journal

Physics Letters A, 304

Authors

Frieden, B. R.

Soffer, Bernard H

Publication Date

2002

Peer reviewed

ELSEVIER

Physics Letters A 304 (2002) 1–7

PHYSICS LETTERS A

www.elsevier.com/locate/pla

Black holes and optimum coding

B.R. Frieden^{*}, B.H. Soffer¹

Optical Sciences Center, University of Arizona, Tucson, AZ 85721, USA

Received 12 June 2002; accepted 28 August 2002

Communicated by C.R. Doering

Abstract

A black hole (BH) communicates information about its event area to an observer. This defines an information channel. We show that if the area is optimally encoded, the probability law on its *square root* follows approximate exponential decay, and the resulting entropy is precisely the Bekenstein–Hawking result including the important factor of 1/4. This result arises as well as the solution to: (a) maximum entropy subject to a fixed message length, or (b) minimum Fisher information subject to a normalization condition. A verifiable prediction is a randomly enlarged event area, implying a randomly enlarged irreducible mass value. These effects should be *observable* as an *enhanced number of BH's with large mass*, resulting in an increased occurrence of gravitational lensing, and an enhanced ability to entrap relatively distant stellar objects.

© 2002 Elsevier Science B.V. All rights reserved.

PACS: 4.70.Dy; 3.67.Hk

1. Introduction

Ideas from communication theory have been widely used in research on black holes [1–4]. Here we characterize the statistical behavior of the BH area as one of maximum cooperation with an observer. This is in the sense of presenting him with event areas whose occurrences as numbers are optimally encoded, and which can be transmitted, as well, to any other observer in optimally encoded form.

But, why nature should “want to cooperate” with the observer? Such a cooperative relationship is in part motivated by a recent view of the observer as “a participant”, in some sense, with the physics he observes [5]. This qualitative view has also been quantified as an information-based framework for constructing the Lagrangians, and probability laws, of physics [6–9]. Laws of biological growth follow as well from the approach [10,11].

Such a cooperative relationship between nature and the observer requires a communication channel. Such a channel consists of a message source, a statistical channel, an observation, and an estimate based upon the observation. The source message is generally an unknown parameter value that is under observation, the channel is an intervening statistical phenomenon, the observation is data about the parameter, and the

^{*} Corresponding author.

E-mail address: roy.frieden@optics.arizona.edu

(B.R. Frieden).

¹ Mailing address: 665 Bienvenida Ave., Pacific Palisades, CA 90272, USA.

estimate is of the parameter value and is formed out of the data.

A black hole and its observer can be placed within this context. The mean or *classical event area* of the BH is the unknown parameter that is *to be estimated* from the observations. The channel is the probability law governing the quantum mechanical fluctuations of the event area. The “observation” is indirect evidence such as Hawking radiation energy, gravitational lensing, and/or any other pertinent data that can be provided by a BH. Such evidence also distinguishes it from being a simple *black body* source. Finally, the estimate is of the BH area (although we do not directly address this aspect of the problem).

2. On optimal coding in general

Consider the digital transmission of some message, e.g., a sentence of English words, over some channel of communication to be chosen. The sentence is the *source* message (see above). The words must be encoded into digits before transmission. The *ideal* communication channel transmits source messages at the *highest possible* rate in time. It accomplishes this by achieving the minimum possible average length per coded symbol. Thus, the problem: What code optimally compresses the *source* messages into *coded* messages that have the *minimum possible average length*?

The solution for the code is long known (see, e.g., [12]). It is called the “Huffman–Shannon–Fano” code. It achieves the desired minimum coded word length by coding the most frequently occurring word with a code of smallest length, less frequently occurring words with codes of larger length, etc. It can be shown that the average coded word length is a minimum if and only if the probability law for coded words of length l_n is an exponential of the form

$$P(l_n) = 2^{-l_n/\Delta l}, \quad l_n = \Delta l, 2\Delta l, 3\Delta l, \dots,$$

where

$$\sum_{n=1}^{\infty} P(l_n) = 1 \quad \text{and} \quad \langle l \rangle = 2\Delta l. \quad (1)$$

The normalization condition and computed mean value $\langle l \rangle$ easily follow by summing the given form

for $P(l_n)$. Here actually the “unit” of word length is $\Delta l = 1$, one digit (letter) long, but we leave Δl a variable to accommodate more general cases of word length “unit”.

The source is the generator of words, e.g., as they occur in a particular book whose contents are being communicated. The first equation (1) says that the ideal book contains words whose frequencies of occurrence fall off as powers of 2. Notice that the Shannon entropy is here

$$\begin{aligned} H &= - \sum_{n=1}^{\infty} P(l_n) \log_2 P(l_n) \\ &= \sum_{n=1}^{\infty} P(l_n) \frac{l_n}{\Delta l} = \frac{\langle l \rangle}{\Delta l} = 2 \text{ bits}, \end{aligned} \quad (2)$$

the latter by the far-right equation (1). The third equality shows that *the entropy is the mean length* per word (again using $\Delta l = 1$). This mean word length is also a minimum as well (see next paragraph), thereby achieving the goal of optimum coding. The fact that the mean word length is a minimum and equals the entropy of the source is the famous Huffman–Shannon–Fano result.

The proof of assertions (1) and (2) is as follows.

3. Alternative derivation using maximum entropy approach

The problem is one of constrained minimization, which is here

$$\begin{aligned} &\sum_{n=1}^{\infty} P(l_n) l_n + \lambda_1 \left[\sum_{n=1}^{\infty} P(l_n) - 1 \right] \\ &+ \lambda_2 \left[- \sum_{n=1}^{\infty} P(l_n) \log_2 P(l_n) + H_0 \right] = \text{extrem.} \end{aligned} \quad (3)$$

This represents a problem where the first term (the mean length) is to be extremized, subject to the second term (a normalization constraint) and the third term (the fixed source entropy value H_0 as a constraint). Since no terms $dP(l_n)/dl_n$ are present in the sums, the usual Euler–Lagrange solution is merely obtained by differentiating the sum with respect to each unknown $P(l_n)$ and equating the result to zero. The Euler–

Lagrange solution [13] is then

$$l_n + \lambda_1 - \lambda_2(\log_2 e)[1 + \ln P(l_n)] = 0, \quad \text{or} \\ P(l_n) = \exp[-1 + (\lambda_2 \log_2 e)^{-1}(l_n + \lambda_1)]. \quad (4)$$

The latter exhibits the basic exponential dependence (1) upon l_n , as required. The constants λ_1, λ_2 are fixed by satisfying the two constraints. The result is then (1) exactly. Finally, it can be shown that the extremum is a minimum, as required.

As an aside, the summation form of principle (3) makes it blind to which of the terms is to be extremized. Thus, the same problem can also be interpreted as one of *extremizing (maximizing) the entropy* subject to a constraint on the mean length. This is complementary to the preceding interpretation: the roles played by the entropy and the mean length are now exchanged.

4. Application to black hole physics

Keeping this coding effect in mind, we return to the *physical* problem of coding at hand. There is a communication channel consisting of a black hole and an observer. The black hole has a classical (or mean) total event area of value A_c . The presence of the BH is made evident to the observer in any number of ways. For example, there could be gravitational lensing, or the observation of the celebrated Hawking photons. Based upon such evidence the observer is to estimate the classical event area size.

The event surface of the BH is assumed to be composed of a constant subdivision of elemental patch areas [1,14] of average size $\langle A \rangle$. Determining the latter is the objective of the observation. Suppose that each patch area randomly fluctuates in size about its mean value $\langle A \rangle$. The fluctuations are of quantum mechanical origin. A patch area is formed as a product $l_1 l_2$ of a randomly selected elemental patch length l_1 and a width l_2 . (This would restrict attention to generally rectangular areas. General ellipses could instead be generated as $\pi l_1 l_2$ if l_1 and l_2 now represent semi-major and semi-minor axes. Other shapes could likewise be formed.) As there is nothing to distinguish lengths from widths, these must obey the same probability law. Likewise they must be independent. It follows from the latter two conditions that

$$\langle A \rangle \equiv \langle l_1 l_2 \rangle = \langle l_1 \rangle \langle l_2 \rangle = \langle l \rangle^2. \quad (5)$$

The ansatz: Let the BH be maximally cooperative with the observer, obeying the probability law (1) on lengths, so as to *permit optimum coding* of its length values. Then by the far-right equation (1), and Eq. (5),

$$\langle A \rangle = 4\Delta l^2. \quad (6)$$

The factor 4 will turn out to be quite significant for a BH—in particular—communication problem.

Likewise the probability law on areas A_m , $m = 1, 2, \dots$, may be formed. As mentioned above, regard

$$A_m = m\Delta l^2, \quad m = 1, 2, \dots \quad (7)$$

Thus, a value of m represents an area. As an example, an area value $m = 4$ will occur as independently occurring *length* pairs of values 1, 4 or 2, 2 or 4, 1. By independence and Eq. (1), these event pairs have probabilities $2^{-1}2^{-4}$, $2^{-2}2^{-2}$, and $2^{-4}2^{-1}$, respectively. Also, as these joint events are disjoint, these probabilities add, so that $P(4) = 2^{-1}2^{-4} + 2^{-2}2^{-2} + 2^{-4}2^{-1}$. In the same way, the probability for a general area multiple m obeys

$$P(m) = \sum_{\{n\}=1}^m 2^{-(n+\frac{m}{n})}, \quad (8)$$

where the summation index $\{n\}$ means only over those values of n in the range $(1, m)$ for which quotient m/n is an integer. (For example, for the case $m = 4$ cited above, the allowed values are $n = 1, 2$ and 4.)

The probability law $P(m)$ obeying Eq. (8) is plotted as the solid curve in Fig. 1 over a limited range of m . The law is of course only defined at the isolated points $m = 1, 2, \dots$, but the points are connected so as to show the overall trend. The other two curves in the figure are likewise shown artificially as continuous curves for the same purpose. The curve $P(m)$ has a generally downward trend, although with *local oscillations* as addressed below.

The sum (8) is dominated by the term corresponding to a *square* area (or a circle in the case of general elliptical areas),

$$P(m) \sim 2^{-2\sqrt{m}}. \quad (9)$$

This is the dashed curve in Fig. 1. This curve is actually only defined at the points $m = 1, 4, 9, 16, 25$ that are perfect squares.

The curve (8) oscillates, since each local value depends upon the “richness” of the number m in

Fig. 1. Probability law $P(m)$ on multiplier m of the Planck area (solid curve); its dominant contributor (dashed curve)—a decreasing exponential function of the square-root of m ; and its lower bound (dotted curve)—a decreasing exponential function of m .

divisors. If m is a prime number then it is maximally poor in divisors, with only the terms $n = 1$ and $n = m$ contributing. The result is a lower-envelope curve to (8), the exponential

$$P_{\min}(m) = 2^{-m} \quad (10)$$

that tracks its local minima. This is shown as the dotted curve in the figure. The curve is defined only at the prime number points $m = 1, 3, 5, 7, 11, 13, 17, 23$. The spaces between points of this curve increase logarithmically with m (a known property of prime numbers). This slow dependence upon m is characteristic of the entire curve $P(m)$ (see Fig. 1).

The curve $P(m)$ in Fig. 1 has a very long tail. This is indicated by the fact that its mean $\langle m \rangle = 4$ and variance $\sigma^2 = 20$ are quite large. By comparison, the latter significantly exceeds the variance value 16 that would hold for an exponential curve (which already has a long tail). This longer tail will have an important, verifiable, physical consequence as discussed in Section 7.

5. Planck length and area

As usual, in (6) take for the elemental length the Planck length,

$$\Delta l \equiv \Delta l_P = \sqrt{G\hbar/c^3}, \quad (11)$$

where G is the gravitational constant, $\hbar$ is Planck's constant divided by 2π , and c is the speed of light. Then (6) becomes

$$\langle A \rangle = 4\Delta l_P^2 = 4\frac{G\hbar}{c^3}. \quad (12)$$

The mean patch area size is exactly *four* Planck areas. This number will lead to the famous occurrence of the number four in the expression for the entropy of the black hole. Also, the elemental areas in Eq. (7) now become multiples of a Planck area Δl_P^2 ,

$$A_m = m\Delta l_P^2, \quad m = 1, 2, \dots \quad (13)$$

If the fluctuations in area A_m are to be of quantum origin, the curve $P(A_m)$ should depend upon the

Planck constant $\hbar$. However, the curve $P(m)$ given by Eq. (8) does not explicitly depend upon $\hbar$. Where do quantum effects enter in? By Eq. (7) a value of m corresponds to an area $m\Delta l^2 = m\Delta l_p^2$ by Eq. (13). Hence, the *area fluctuations* that obey Eq. (8) are *multiples* of the Planck area. This brings in both $\hbar$, as required, and the gravitational constant G . Consequently, the probability law $P(A_m)$ can be regarded as a quantum gravitational effect.

6. Calculation of the entropy

As the first step in calculation of the entropy, we need to know the expected number N of patches that lie within the event surface of area A_c . By the surface model given at the beginning of Section 4 this is

$$N = \frac{A_c}{\langle A \rangle}. \quad (14)$$

To further evaluate this we need to know the value of $\langle A \rangle$.

The calculation of the entropy follows an overall procedure given in [1] (second paper). The idea is to consider all the possible configurations of a BH that are *consistent with a given mean value of its energy*. This utilizes the concept of a “micro-canonical ensemble” in place of the usual “canonical ensemble” of statistical mechanics [15]. The mean value of the radiated energy from a BH is known to have the value [16]

$$\langle E \rangle = \left[(\exp(\hbar\omega_c/kT) - 1)^{-1} + \frac{1}{2} \right] \hbar\omega_c, \quad (15)$$

where ω_c is a classical oscillator frequency, k is the Boltzmann constant, and T is the BH temperature. The mean value (15) should occur at every patch of the event area since the event source patch is assumed typical of the event surface. Then by Eq. (14) the *total* mean energy is fixed as $(A_c/\langle A \rangle)\langle E \rangle$. Therefore by the second paragraph of this section, *we can now count states* subject to this total mean energy value.

A major premise of BH physics is the “no hair” hypothesis [15]: *It is impossible to know the internal states of a BH*. We will apply this hypothesis in various ways. First, it means that we can only *infer* the structure of the internal states. The inference shall be by the use the principle of maximum likelihood (ML),

which states that the unknowns of a problem are those that make observed data maximum probable.

Suppose that each patch can be in any of the states $a_1, \dots, a_Z$. We apply the ML principle to compute the average number Z of states/patch. Consider a *typical sequence* of states among the N patches, say

$$a_3 a_7 a_3 a_8 a_7 \dots a_{12}.$$

Notice that level a_3 has occurred $2 \equiv m_3$ times. A “typical sequence” is defined to be one for which each occurrence number m_i , $i = 1, \dots, Z$, is consistent with the *a priori* probability P_i of its state a_i . The obvious such value of m_i is $N P_i$ over the N patches. We can now form the likelihood law $P(m_i)$ for the occurrences.

As each of the m_i occurrences is independent, its likelihood $P(m_i)$ is a product of the individual probabilities a total of $N P_i$ times, or,

$$P(m_i) = P_i^{N P_i}. \quad (16)$$

This is to be maximized, by the ML principle. The maximization is through the choice of P_i . The solution is found by taking the logarithm and differentiating, with the result $P_i = 1/e$. Again invoking the “no hair” hypothesis, every state a_i must be equally probable, so that $P_i = 1/Z$. Consequently the average number of states/patch is

$$Z = e \text{ states/patch}. \quad (17)$$

We can now count up the total number of states over the N patches. This will likewise be an average number.

Then since there are N patches and e states per patch, the *average number of states* L_A is simply

$$L_A = e^N. \quad (18)$$

If this is the only source of degeneracy of states, then [17] the Boltzmann entropy equals the Boltzmann constant k times the Shannon entropy in “nepits”, i.e., using natural logarithmic base e . (See [18] for a general discussion of the validity of this correspondence.) To compute the latter requires knowledge of the probability law on the *joint state* across the surface.

This implies that we cannot, even in principle, know that one joint state is more probable than another. In other words we have to admit of a flat probability density for the joint states. This is an admission of a situation of maximum disorder as well. Under

these equiprobable conditions the Shannon entropy is just the log of the total number of states (18). We can now proceed with the entropy calculation. The result is that the Boltzmann entropy is k times this, or

$$H_B \equiv k \ln(e^N) = kN = k \frac{A_c}{\langle A \rangle} = k \frac{A_c}{4\Delta l_p^2}. \quad (19)$$

Eqs. (12) and (14) were also used. This is the celebrated Bekenstein–Hawking area law, including the required factor of 4. In past work [19] the factor of 4 was accounted for by assuming a special value for the so-called “Immirzi parameter” [20], in the special case of a nonrotating black hole. Other attempts at obtaining the factor are discussed in [19] and [21].

7. Discussion

We have shown that the Bekenstein–Hawking area law for the entropy, including the necessary factor of 4, follows from an assumption that nature presents the observer with elemental event lengths that can be encoded with maximum efficiency. The latter means the accomplishment of a minimum average code length per number, where the minimum equals the entropy of the lengths. An added result is the prediction that patch areas should follow the occurrence law (8), with the simple approximation (9) representing the dominant term in (8) due to a square or circular area. Also, the lower-envelope curve (10) is of a pure exponential form.

The property of optimal coding was, at first, an *ansatz*. As we saw in Section 2 it follows whenever the occurrence law on fundamental events (here, length) is of a simple exponential form. Of course the exponential law, in the form of the Boltzmann energy distribution law, is also the basis for statistical mechanics. Then in view of oft-mentioned parallels between classical statistical mechanics and black hole phenomena it is perhaps not surprising to find that the exponential law likewise describes in some way black hole phenomena.

Furthermore, we showed in Section 3 that the *ansatz* in fact follows from requiring the probability law on length to obey *maximum entropy* in the presence of a fixed mean length. This follows the formulation of statistical mechanics developed by Jaynes quite awhile ago [22], and which is used widely today [23].

It has been shown [8] that the exponential law follows, as well, under conditions of maximum ignorance in the form of *minimum Fisher information*, subject to a normalization condition as prior knowledge.

Thus, the thesis of optimal coding is physically plausible, following from three different approaches. However, we could have alternatively postulated that the *areas*, rather than the lengths, are optimally encoded, and themselves follow the exponential law. Why should lengths, rather than areas, allow for optimal encoding? This question ties in with the long-standing enigma of why space–time coordinates seem to be the preferred coordinates for describing physical effects. This largely philosophical question invites some speculation. One possible reason is that lengths, or more generally space–time values, are simply the quantities whose measurements are most often taken in defining physical effects. As a corroboration, it is the measurement of space–time values, not areas, that elicit the laws of physics [6–11] via the EPI (extreme physical information) principle. As was discussed, the latter theory offers one justification for a cooperative nature as assumed in this paper; and the theory is based upon the measurement of space–time values, not areas.

Some immediate reasons for the lengths, and not areas, obeying the exponential law follow from the specifics of the given problem. If the areas obeyed an exponential law, the law equation (10) would be the exact answer and not merely a lower bound to it. This would *not* give rise to the important factor 4 in the Bekenstein–Hawking law. It also would not show the long tail region of the answer equation (8), important ramifications of which are discussed next.

The long tail of the curve $P(m)$ plotted in Fig. 1 implies that, randomly, *very large* values $m\Delta l_p^2$ of patch area (13) can occur. These are for area multiples m that are rich in divisors, such as $m = 2^n$, n a large integer. Such large areas A imply correspondingly large irreducible mass values M_{ir} , since [24]

$$M_{ir} = \frac{c^2}{4G} \sqrt{\frac{A}{\pi}}. \quad (20)$$

The increased mass should enhance the ability of a BH to entrap even *distant* stellar bodies. This entrapment effect is a prediction of the approach, and is a candidate for future observational verification. Other mass-dependent effects should likewise be affected by the augmented mass value: e.g., enhanced gravitational

lensing. Also, Eqs. (8) and (20) together give a prior probability law on the irreducible mass associated with any event patch. The extra-long tail of (8) implies that a *survey* of stellar BH masses should show a corresponding bias toward the occurrence of large masses.

References

- [1] J.D. Bekenstein, Phys. Rev. D 30 (1984) 1669;
J.D. Bekenstein, Stud. Hist. Philos. Mod. Phys. 32 (2001) 511.
- [2] W. Israel, in: V. de Sabbata, Z. Zhang (Eds.), Black Hole Physics, Kluwer, Dordrecht, 1992.
- [3] W. Kundt, Nature 259 (1976) 30.
- [4] D.C. Wilkins, J. Gen. Rel. Grav. 11 (1979) 45.
- [5] J.A. Wheeler, in: J.J. Halliwell, J. Perez-Mercader, W.H. Zurek (Eds.), Physical Origins of Time Asymmetry, Cambridge Univ. Press, 1994, pp. 1–29.
- [6] B.R. Frieden, B.H. Soffer, Phys. Rev. E 52 (1995) 2274.
- [7] B.R. Frieden, W.J. Cocke, Phys. Rev. E 54 (1996) 257.
- [8] B.R. Frieden, Physics from Fisher Information, Cambridge Univ. Press, 1998.
- [9] B.R. Frieden, A. Plastino, Phys. Lett. A 278 (2001) 299.
- [10] B.R. Frieden, A. Plastino, B.H. Soffer, J. Theor. Biol. 208 (2001) 49.
- [11] R.A. Gatenby, B.R. Frieden, Cancer Res. 62 (2002) 3675.
- [12] F.M. Reza, An Introduction to Information Theory, McGraw-Hill, New York, 1965.
- [13] G.A. Korn, T.M. Korn, Mathematical Handbook for Scientists and Engineers, 2nd Edition, McGraw-Hill, New York, 1968.
- [14] A. Alekseev, A.P. Polychronakos, M. Smedback, hep-th/0004036.
- [15] S.W. Hawking, Phys. Rev. D 13 (1976) 191.
- [16] J.D. Bekenstein, Phys. Rev. D 7 (1973) 2333.
- [17] L. Brillouin, Science and Information Theory, Academic Press, New York, 1956.
- [18] R. Kikuchi, B.H. Soffer, J. Opt. Soc. Am. 67 (1977) 1656.
- [19] A. Ashtekar, J. Baez, A. Corichi, K. Krasnov, Phys. Rev. Lett. 80 (1998) 904.
- [20] G. Immirzi, Class. Quantum Grav. 14 (1997) L177.
- [21] C. Rovelli, gr-qc/9803024.
- [22] E.T. Jaynes, Phys. Rev. 106 (1957) 620;
E.T. Jaynes, Phys. Rev. 108 (1957) 171.
- [23] E.T. Jaynes, W.T. Grandy, P.W. Miloni (Eds.), Physics and Probability: Essays in Honor of E.T. Jaynes, Cambridge Univ. Press, Cambridge, 1993.
- [24] C.W. Misner, K.P. Thorne, J.A. Wheeler, Gravitation, Freeman, New York, 1973.