

Steven Turner

12/5/2008

Statistical Mechanics

Dr. Joshua Grossman

## The Entropy of Information

Physics is the study of motion. Although true, it is also the study of classification. Physical theories arise from observations of a system, mathematical ideas, and reproducible experiments. Information is the result of these activities while the Universe itself is not actually composed of this data. Because of this relationship, there are fundamental limits imposed by the laws of thermodynamics on information systems such as the human brain, the abacus, and all modern computation. The second law, increasing universal entropy, is impossible to avoid in the design and construction of information systems whether they are books, speech, or satellite communication channels. However, with a sound understanding of the second law, it is possible to exploit entropy to advantage through compression and error correction. Modern technology operates reliably and securely on these two fields of information theory, both of which result from the careful study of physics applied in a highly practical manner.

Estimated to be capable of storing the equivalent of 10 terabytes or 10 million photographs, the human brain attains the highest known density of information,  $1 \times 10^7$  atoms per bit. [7] The brain operates on vastly different principles than modern information storage devices, but the laws of physics apply to both in the same manner. Consider the cases of a person and computer attempting to store a particular book. Assuming the book occupies  $10 \times 10^7$  bits in both storage mediums, how is it possible for

both to store the contents of several books while saving space for other information? The answer is simple—most people do not save the irrelevant chapters and details of the book. The computer operates on a similar idea: compression. By joining repetitive information and throwing out useless data, it is possible to increase the storage density of a system. [3] First, we must understand information entropy and its relation to two-state systems as defined by:

$$H(p_1, \dots, p_n) = \sum p_i \log_2 \left[ \frac{1}{p_i} \right] = - \sum p_i \log_2 [p_i] \quad (\text{Eq. 1})$$

Here, the entropy is equal to the sum of all the probabilities of a given symbol in the text multiplied by the logarithm of the probability where any compression scheme must adhere to the following relation given by Shannon:

$$L(K) \geq H(S), \quad L(K) = \sum_{i=1}^n d_i p_i \quad (\text{Eq. 2})$$

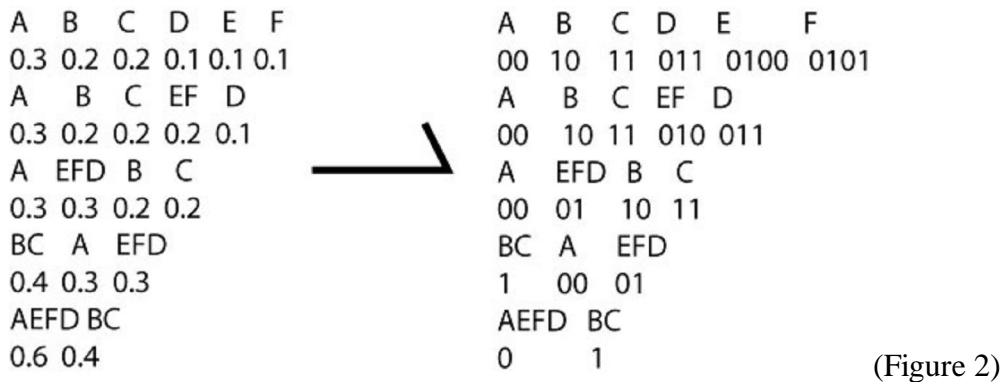
L(K) is the average length of a compression coding for a particular symbol of the source data lexicon. Thus, to preserve the original data integrity the coding length is only minimal if it is equal to the entropy of an information source and cannot exceed this limit.

[6] Applying these formulas to design several algorithms for efficient lossless compression including the popular Huffman, arithmetic, and Sequitur schemes have increased information density and transmission since the 1950's. [3] A useful case to understand information encoding is to examine the commonly used Huffman algorithm applied to the first six letters of the alphabet arranged from most to least probable.

Symbol	A	B	C	D	E	F
Probability	0.3	0.2	0.2	0.1	0.1	0.1

(Figure 1)

Next, the algorithm joins the last two symbols and adds their probabilities to construct a binary tree, recursively traversing the tree to generate the coding.



As show above, encoding “DEFCAB” results in the binary string, “01101000101110010.” Using equation one from above we can determine the entropy of the information source encoded.

$$H(p_1, \dots, p_n) = -(0.3 \log_2[0.3] + \dots + 0.1 \log_2[0.1])$$

$$H(p_1, \dots, p_n) = 1.59 \text{ (Figure 3)}$$

Here, “1.59” is the minimum code length for any compression scheme and our Huffman scheme has an average code length of 2.5 bits, validating the physical limit imposed by the relation in equation two. In general, digital information systems represent the symbols in an alphabet with 16 bits. Thus, the compression ratio of this Huffman information-coding scheme is approximately one to six—a dramatic saving of storage or transmission space.

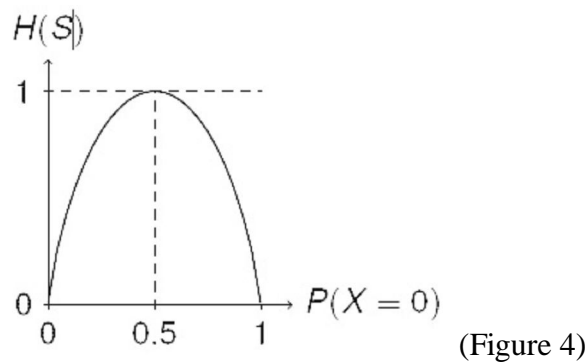
An additional benefit of compression is a decrease in energy required for storage. Every bit erased in memory decreases the data entropy by  $k \cdot \ln(2)$  joules per Kelvin but increases the total system entropy in similar fashion of Maxwell’s demon. [7] Extending this concept a step further, arguing that logic gates used in circuits effectively erase a

single bit per logic operation, logical thought processes must require energy. Statistical mechanics studies large systems and make predictions about their behavior. Given the current system of compressed data, a theory of reliable information transport is necessary in inherently unpredictable environments such as the Universe.

Information transport is hazardous. The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. [6] Consider the bit string of “101000” meaning, “Excellent work.” If even one bit is flipped in the message, for example, “101001” the meaning could become, “You are fired,” depending on the coding scheme. A transmission system consists of an information source, channel, and receiver. [6] Here, the channel represents a binary symmetric communications link such as a fiber optic cable or radio antenna. If the channel has a probability error “p” then the following relations hold true. When a zero is sent, a zero is received with probability “q” equal to 1-p. Similarly, when a one is sent, a one is received with probability “p.” In the extreme cases of probability equaling 0.0 or 1.0, it is trivial to read a message encoded over the channel as either no errors have occurred or all bits in the source message have been inverted. However, for a less extreme probability of random bits flipped in the code, it becomes impossible to determine the meaning. Entropy measures the average information present in a source symbol. [3] Modifying the formula discussed above it is possible to determine the information retained or lost over a noisy channel using conditional probabilities:

$$H(\mathbf{X} | \mathbf{y}_i) = - \sum_{j=1}^n P(\mathbf{x}_j | \mathbf{y}_i) \log_2 [P(\mathbf{x}_j | \mathbf{y}_i)] \quad (\text{Eq. 3})$$

Here,  $y_i$  is the symbol received and  $x_j$  is the symbol sent. [6] Then, the channel's ability to send information or capacity defined as one minus the entropy of the source allows us to maximize a coding scheme's effectiveness of transmitting data safely. It is possible to visualize a channel's capacity quite easily.



When the probability of an error occurring in a message is exactly one half, the entropy and channel capacity are maximized as mentioned previously. However, for the same probability without an error-handling scheme, it is impossible to determine the intended meaning of a message. It is possible to devise an algorithm for error-correction that operates over a noisy channel at capacity. Unfortunately, in modern digital computing, these codes are computationally expensive and thus most implementations today use approximate methods for maintaining data integrity. [3]

Error-correcting coding schemes of information traditionally operate using basic linear algebra. Using the previous example of “Excellent work” and “You’re fired,” two binary vectors or linear codes represent this information:

$$\mathbf{u} = (1, 0, 1, 0, 0, 0)$$

$$\mathbf{v} = (1, 0, 1, 0, 0, 1) \text{ (Figure 5)}$$

It is then possible to define the Hamming distance, from Richard Hamming, between these codes as  $d(u,v)$  equivalent to the number of “places” these codes differ. [3] In this

example  $D(u,v)$  equals one since the sixth bit is the only location that changes which is defined as the syndrome of a code received.

The simplest method of error correction is to replicate bits over a channel. For example, “1” becomes “111” creating a minimum distance between any received code word of one. [3] There are other schemes, including even parity, Hamming, Golay, Cyclic, and Reed-Solomon, which offer varying degrees of error detection, correction, and transmission rates and depend on linear independence. [3] One of the most famous of these schemes is the Golay “24” coding used on the Voyager spacecrafts which detects four errors and corrects three. Using a binary Golay, a sender starts with a  $12 \times 24$  generator matrix composed of an identity matrix in the first 12 columns and the remaining columns represent integer squares modulus 11, where a “1” is placed at the number computed. The remaining rows are then permutation of this algorithm. If Voyager wishes to send a message,

$$\mathbf{m} = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0) \text{ (Figure 6)}$$

it computes the codeword by multiplying this message by the generator matrix,  $G$ , and transmitting:

$$\mathbf{mG} = (1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0) \text{ (Figure 7)}$$

Generalizing this scheme, the ground station receives the message, computes the syndrome, and adds an appropriate column vector from  $G$  to the received message to arrive at the original. Using Golay codes it is possible for the receiver to determine if the message contained errors, correct those errors, or request resubmission if the codeword is not a member of the Golay vector field if the number of errors is greater than four.

These methods of controlling entropy introduced over a channel pertain to digital information. The human brain too has methods for guarding against errors—the skull protects from impacts and neurons grow redundant connections to ensure redundancy should damage occur. By manipulating information with statistical mechanics, it is possible to achieve maximum throughput over a channel, increasing entropy, while ensuring reliable transit.

Information theory does not exist without physics. In this brief overview, compression and error-correction are means of maximizing the density and reliability of information. However, these theories require the existence of basic physical principles, as both matter and energy are compulsory to the storage and transmittance of information. Returning from these higher abstractions to the application of statistical mechanics to information systems, there is a fundamental limit on the number of states available in the universe, which restricts our ability to process data inferred from the environment. Thus, with respect to the second law of thermodynamics, it is necessary for the universe and physical theories to exist before information. As a result, the extent of human memory and our understanding of the universe through physics is fundamentally limited by available matter and energy.

References:

1. T. M. Cover and Joy A. Thomas, *Elements of information theory*, Wiley series in telecommunications (New York: Wiley, 1991).
2. E.T. Jaynes, "Information Theory and Statistical Mechanics," *The Physical Review* 106, no. 4 (May 15, 1957): 620-630.
3. Wade Trappe, *Introduction to Cryptography: With Coding Theory*, 2nd ed. (Upper Saddle River, N.J: Pearson Prentice Hall, 2005).
4. Aleksandr Akovlevich Khinchin, *Mathematical Foundations of Information Theory*, New Dover ed. (New York: Dover Publications, 1957).
5. Susan Loepp, *Protecting Information: From Classical Error Correction to Quantum Cryptography* (Cambridge: Cambridge University Press, 2006).
6. C.E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* 27 (October 1948): 379-423, 623-656.
7. Schiller, Christoph. *The Motion Mountain Physics Textbook*. Vers. 19th. 2004. 5 Dec. 2006. 25 Nov. 2008
8. Charles Bennett and Rolf Landauer, "The Fundamental Physical Limits of Computation," *Scientific American* 253, no. 1 (1985): 48-56.