Communication near the channel capacity with an absence of compression: Statistical Mechanical Approach

Ido Kanter and Hanan Rosemarin
Minerva Center and the Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

The generalization of Shannon's theory to include messages with given autocorrelations is presented. The analytical calculation of the channel capacity is based on the transfer matrix method of the effective 1D Hamiltonian. This bridge between statistical physics and information theory leads to efficient Low-Density Parity-Check Codes over Galois fields that nearly saturate the channel capacity. The novel idea of the decoder is the dynamical updating of the prior block probabilities which are derived from the transfer matrix solution and from the posterior probabilities of the neighboring blocks. Application and possible extensions are discussed, specifically the possibility of achieving the channel capacity without compression of the data.

Digital communication, which is the driving force of the modern information revolution, deals with the task of achieving reliable communication over a noisy channel. A typical communication channel is depicted in figure 1 (where $K \leq L$ and $K \leq N$).

Shannon, in his seminal work[1], proved that in order to overcome noise, redundancy must be added to the transmitted message. The channel capacity, which is the maximal rate ($R \equiv \frac{K}{N}$) is a function of the channel bit error rate (f), bit error rate (P_b) and the a priori bit probability (P)

$$R = \frac{1 - H_2(f)}{H_2(P) - H_2(P_b)} \tag{1}$$

where $H_2(x) \equiv -x\log_2(x) - (1-x)\log_2(1-x)$. The entropy, defined as the information content (in bits per symbol) of the message, is unity for unbiased ($P = 0.5$) messages and decreases with bias ($H_2(0) = H_2(1) = 0$).

To maximize channel throughput, the message is first compressed using algorithms (e.g., [2]) which approach entropy for large blocks, and then encoded assuming a unbiased message. Typical messages exhibit low autocorrelation coefficients (decreasing with the size of the message)[3]. For instance, the discrete periodic 2-point autocorrelation coefficient of a sequence $\{X_i\}$ is defined by

$$C_k = \frac{1}{L}\sum_{i=1}^{L} X_i X_{(i+k) \bmod L} \tag{2}$$

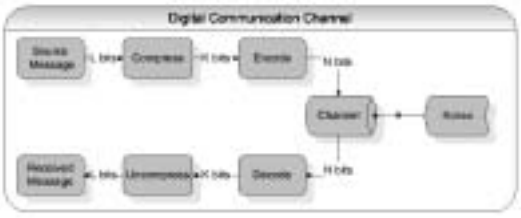

Figure 1: Digital Communication Channel

Compressible data exhibits enhanced autocorrelation coefficients, whose distribution has been extensively studied in numerous fields (e.g., linguistics, DNA research, heartbeat intervals, etc.)[4, 5].

In this paper we raise two questions:

- what is the channel capacity of correlated sequences?
- are these bounds achievable by an encoder alone, without the preceding compression phase?

We address these questions by using models from statistical mechanics, which lead to a new class of decoding algorithms. The decoder nearly approaches the capacity expected for perfect compression, without directly applying any compression.

The motivation for direct transmission of the uncompressed data is to overcome many difficulties, including:

- a single bit error in a compressed block would render the entire block useless. This is particularly important for applications where minimal distortion can be tolerated (e.g., audio/video transmission)
- compression of different packets of the same length would generate different length output messages, calling for greater complexity of the encoder
- the compression/decompression add delay to each transmission, and increase the complexity of the transmitter/receiver.

We start by calculating the entropy of correlated bit sequences. This is done by binning the source into K_o bits, where K_o is the highest autocorrelation coefficient taken, and using the transfer-matrix method. We proceed by integrating the results of the transfer matrix into a new decoding algorithm, based on a Low-Density Parity-Check Codes (LDPCC) over finite fields ($GF(q)$)[6].

For the sake of simplicity, we begin by demonstrating the entropy calculation for two autocorrelation coefficients,

namely, C_1 and C_2. The entropy of a set of binary sequences of length L is defined by the log of the number of possible sequences (Ω) divided by L. Calculating Ω with the correlation constraints (C_1, C_2)[7], is done by taking the trace over all possible states of the sequence obeying the constraints imposed as delta functions

$$\Omega = \mathrm{Tr}\,\delta\left(\sum_j x_j x_{j+1} - C_1 L\right)\delta\left(\sum_j x_j x_{j+2} - C_2 L\right) \quad (3)$$

where periodic boundary conditions are assumed for the indices. Using the Fourier representation of the delta functions and rearranging terms, one gets

$$\begin{aligned}\Omega &= \frac{1}{(2\pi)^2}\int\int_{-\infty}^{\infty} \mathbf{d}y_1 \mathbf{d}y_2 e^{-iL(y_1 C_1 + y_2 C_2)} \\ &\times \mathrm{Tr}\prod_{j=1}^{\frac{L}{2}} e^{iB(y_1, y_2, x_{2j}\ldots x_{2j+3})}\end{aligned} \quad (4)$$

where B is given by

$$\begin{aligned}B(y_1, y_2, x_0, x_1, x_2, x_3) &\equiv \frac{y_1}{2}(x_0 x_1 + 2x_1 x_2 + x_2 x_3) \\ &+ y_2(x_0 x_2 + x_1 x_3)\end{aligned} \quad (5)$$

Using the transfer-matrix method[8], we group the sequence into blocks of 2 bits, obtaining the 4x4 matrix

$$V(y_1, y_2) = \begin{pmatrix} e^{2y_1+2y_2} & e^{y_1} & e^{-y_1} & e^{-2y_2} \\ e^{-y_1} & e^{-2y_1+2y_2} & e^{-2y_2} & e^{y_1} \\ e^{y_1} & e^{-2y_2} & e^{-2y_1+2y_2} & e^{-y_1} \\ e^{-2y_2} & e^{-y_1} & e^{y_1} & e^{2y_1+2y_2} \end{pmatrix} \quad (6)$$

representing all possible interactions between neighboring blocks. We proceed by replacing the trace with $\lambda_{max}^{\frac{L}{2}}$, where λ_{max} is the principal eigenvalue of $V(y_1, y_2)$. Using the method of Laplace integrals we obtain for the leading order[9] of Ω

$$\Omega = e^{-L\left(y_1^\star C_1 + y_2^\star C_2 - \frac{1}{2}\ln\lambda_{max}\left(y_1^\star, y_2^\star\right)\right)} \quad (7)$$

where $\left(y_1^\star, y_2^\star\right)$ are the solution of the saddle point equations. The entropy in bits is given by

$$H(C_1, C_2) = \frac{\frac{1}{2}\ln\lambda_{max}\left(y_1^\star, y_2^\star\right) - y_1^\star C_1 - y_2^\star C_2}{\ln 2} \quad (8)$$

The entropy of the correlated sequences are shown in figure 2 for all possible $\{C_1, C_2\}$ tuples. The area of non-zero entropy is depicted in figure 2 between the two straight lines, i.e., $\frac{-1-C_2}{2} < C_1 < \frac{1+C_2}{2}$. The parabolic line is $C_2 = C_1^2$, where the entropy reduces to the case of a single C_1 constraint, with the entropy $S = H_2\left(\frac{1+C_1}{2}\right)$[10]. This is the typical case for the C_1 constraint, since C_2 can be viewed as multiplying two consecutive C_1 pairs, hence it has the highest entropy for a given C_1.

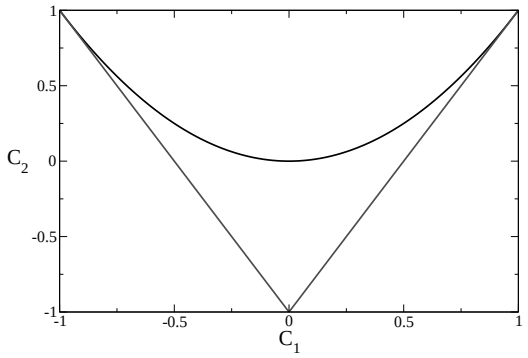

Figure 2: Phase space for sequences with C_1, C_2. Outside the triangle ($|C_1| \geq \frac{1+|C_2|}{2}$) the entropy is zero. The parabolic curve is $C_2 = C_1^2$.

Note that at the boundary of the phase space the entropy falls abruptly to zero (a first order phase transition). Another point, which might appear counter-intuitive, is the fact that the entropy of the constraint $\{C_1 = A, C_2 = B\}$ is not the same as $\{C_1 = B, C_2 = A\}$ [11].

The transfer-matrix solution shows that the ensemble of sequences obeying the correlation constraints are obtained from the thermal equilibrium solution of the 1-D Ising spin model (where the two states of the spin correspond to the binary values of the bit $\{0, 1\} \rightarrow \{1, -1\}$ [7]). The interaction length is the same as the correlation length, and the interaction strength is $J_i = -\frac{y_i^\star}{\beta}$ (where β is the inverse temperature). The corresponding Hamiltonian is $H = -\frac{y_1^\star}{\beta}\sum x_i x_{i+1} - \frac{y_2^\star}{\beta}\sum x_i x_{i+2}$. This physical observation is the key leading to our novel decoding algorithm.

Testing the physical mapping, we choose a pair of constraints (C_1, C_2), solve the transfer matrix model to obtain $y_1^\star, y_2^\star$, then select any temperature (e.g., $\beta = 1$) which gives the interactions J_1, J_2[12], and perform Monte-Carlo simulations, and indeed the system settles into an equilibrium state obeying the initial autocorrelation constraints.

The division into blocks of 2 bits, amplifies the fact that interactions affect only neighboring blocks, and the probability of finding a block in one of its four possible states *depends only on the state of its two neighbors*. Labeling the left, center and right blocks by l, c, r respectively, one gets for the joint probability of three consecutive blocks

$$P(c, l, r) = \frac{S_I(c)\, q_L^l S_L(l, c)\, q_R^r S_R(r, c)}{\mathrm{Tr}_{\{c,l,r\}} S_I(c)\, q_L^l S_L(l, c)\, q_R^r S_R(r, c)} \quad (9)$$

where

$$\begin{aligned}S_I(c) &= e^{y_1^\star c_1 c_2} \\ S_L(l, c) &= e^{y_1^\star l_2 c_1 + y_2^\star (c_1 l_1 + c_2 l_2)}\end{aligned} \quad (10)$$

$$S_R(r,c) = e^{y_1^\star c_2 r_1 + y_2^\star (c_1 r_1 + c_2 r_2)}$$

The first term in Equation 10 denotes the inner-block interactions, and the other terms denote the inter-block interactions. q_L^l/q_R^r is the posterior probability of finding the left/right block in a specific state given by the decoding algorithm below, and the subscripts 1,2 denote the bit number in the block.

The physical spin model emphasizes the dramatic difference between the case of biased bit probabilities, equivalent to an induced homogeneous field, where each spin is updated independently, to the case of correlations induced by inter-spin interactions.

The entropy of the class of correlated sequences, derived by the transfer-matrix method, can be plugged into equation 1. Using shannon's lower bound, gives the channel capacity of sequences with two autocorrelation coefficients

$$R^\star = \frac{1 - H_2(f)}{H_2(C_1, C_2) - H_2(P_b)} \tag{11}$$

The previous formulation is easily extended to higher autocorrelation coefficients. The procedure involves adding more delta functions to equation 3, resulting in a larger transfer matrix. For C_l , being the highest autocorrelation coefficient, one gets l variables ($y_1 \ldots y_l$), and a transfer matrix's dimension is $2^l \times 2^l$, which is solved numerically. The simulation results, shown in this paper, were conducted on sequences with 2 (C_1, C_2), and 3 (C_1, C_2, C_3) autocorrelation constraints.

Equation 9 is the crux of our decoding algorithm, it enables the decoder to set prior probabilities for each block based on the current state of its two neighbors.

Decoding of the autocorrelated sequences is based on LDPCC which have been shown to asymptotically nearly saturate Shannon's bound[13–15]. These codes easily lend themselves to decoding blocks of t bits by moving from boolean algebra to Galois-Field ($GF(q)$, with $q = 2^t$). This method was originally studied[6], as a means of increasing the degree of the nodes (connectivity of the graph) without introducing loops, which are known to degrade the decoder's performance.

The decoding consists of iterating two rounds of message passing (also known as belief-propagation), which update two probability matrices R_{mn}^a, Q_{mn}^a ($a \in 0..(q-1)$ stands for the state of the (m,n) element, where $m \in \left\{1 \ldots \frac{N}{t}\right\}$, and $n \in \left\{1 \ldots \frac{N+K}{t}\right\}$ is the block number [16]).

- Updating R_{mn} is based on the values of Q_{mn}, and the specific parity check. This stage is left unchanged.
- Updating Q_{mn} is based on R_{mn} and the a priori probabilities of the q states of the block

$$Q_{mn}^a = \alpha_{mn} P_n^a \prod_{j \in M(n) \backslash m} R_{jn}^a \tag{12}$$

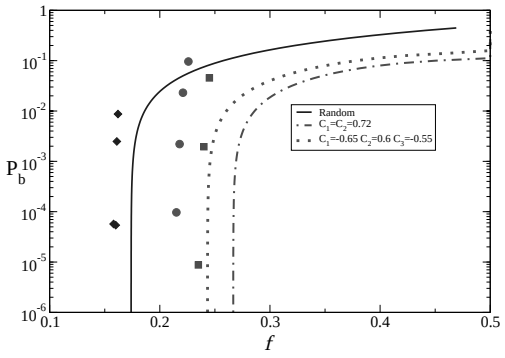

Figure 3: P_b (decoder bit error rate) vs. f (channel bit error rate) with rate $R = \frac{1}{3}$ for uncorrelated sequences; for sequences with $C_1 = C_2 = 0.72$; and for sequences with $C_1 = -0.65$, $C_2 = 0.6$, $C_3 = -0.55$. Results of simulations using KS codes[15], with $N = 10^4$ (9,999 for GF(8)), averaged over at least 2,000 samples, are given for the regular algorithm (♦) and for the new algorithm (■ and ● for the 2 and 3 autocorrelations respectively).

where $\alpha_{mn} = \frac{1}{\sum_{a=1}^{q} Q_{mn}^a}$ is a normalizing constant.

In this stage the a priori probabilities, P_n^a, of the block, are obtained by marginalizing the joint probability (equation 9)

$$\gamma_n^c = S_I(c)\left(\sum_{l=1}^{q} q_L^l S_L(l,c)\right)\left(\sum_{r=1}^{q} q_R^r S_R(r,c)\right)$$
$$P_n^c = \frac{\gamma_n^c}{\sum_{j=1}^{q} \gamma_n^j} \tag{13}$$

where l/r denotes the state of the $n-1/n+1$ block respectively, and q_L^l/q_R^r are their posterior probabilities. The meaning of equation 13 is summing the weighted contributions from all q^2 possible states of the two neighboring blocks to a possible state of the block in question. Thus, in each iteration the prior probability of a block is *dynamically updated* following its neighboring blocks.

The decoder performs the iterations until the message is decoded (all checks are satisfied), or one of two alternate halting criteria is met: (a) the maximal number of iterations is reached. In this paper the maximum was set to 500; (b) the decoder has not modified its estimate of the message in the last 50 iterations.

Simulations were run using the Binary Symmetric Channel (BSC), and the construction of the check matrix (H) used is based on KS Codes[15], for $R = \frac{1}{3}$, where the non-zero elements of H are randomly drawn from the range $1 \ldots (q-1)$.

Figure 3 shows the results of simulations for uncorrelated messages, and messages with 2 and 3 autocorrelation constraints[17]. The performance of all 3 cases is similar, both in the number of iterations, and in the distance from their respective limits[18, 19].

Although we demonstrated the ability to increase the possible noise rate the decoder can handle while keeping the rate fixed, the converse is also possible. Keeping the channel noise fixed, higher rates are achievable. Turning to figure 1, the rate is defined as $R \equiv \frac{K}{N}$, but our encoder compresses the messages as well, thus the achievable rate is actually greater. For $P_b = 0$, the total rate is $R^\star \equiv \frac{L}{N} = \frac{R}{H(X)}$ (where $H(X)$ is the entropy of the message); for $H(X) < R$ rates greater than 1 are feasible.

The complexity of the encoding process remains linear ($O(N)$), since the calculation of a small number of autocorrelation coefficients is linear. The decoder's complexity per iteration scales as $\frac{N}{t}q^2u$ for a LDPCC decoder over $GF(q)$, with $\frac{N}{t}$ blocks, and u checks per block[6]. Our decoder adds an order of $\frac{N}{t}\left(q^2+q\right)$ operations for the calculation of the block prior probabilities, calculating q probabilities, based on the pairs of q states of the 2 neighboring blocks, in equation 13. The algorithm lends itself to parallel implementation which can reduce the complexity to $O(1)$.

All simulations shown in this paper were done assuming BSC, however the formulation is channel independent, and can be applied to any channel (e.g., the popular Gaussian channel). The method can be also applied to different codes as well (e.g., Turbo-Code), by taking into account the dynamically updated block probabilities.

The method described in this paper can easily be extended to higher order correlation functions (e.g., 3-point correlation function, i.e., $C_{kk'} = \frac{1}{L}\sum_{i=1}^{L} X_i X_{i+k} X_{i+k'}$, with periodic boundary conditions). Adding more correlations has the effect of reducing the entropy by lifting the degeneracy associated with 2-point correlation functions. The size of the transfer-matrix, however, does not have to increase, thus the decoding complexity is unchanged. For example, a transfer matrix of two bits can accommodate 1 additional autocorrelation - C_{12}, and a transfer matrix of 3 bits can accommodate 3 additional autocorrelations - C_{12}, C_{13}, C_{123}, the last one being a 4-point correlation function[20].

We have demonstrated the ability to nearly saturate Shannon's limit without compression mechanisms. The algorithm can be used for utilizing specific noise statistics as well. This work paves the way for additional extensive theoretical research and various practical implementations.

Fruitful discussions with E. Kanter are acknowledged. We would like to thank Shlomo Shamai for critical comments on the manuscript.

[1] C. E. Shannon. A mathematical theory of communication. *Bell System Technical J.*, 27:398-403, 1948

[2] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Information Theory*, IT-24:530-536, 1978

[3] The reduced autocorrelation coefficients which occur for typical messages (high entropy), are not to be confused with low-autocorrelated sequences which have extremely low entropy. For a recent discussion see: L. Ein-Dor et al., *Phys. Rev. E* **65**, 020102, 2002

[4] A. Bunde and S.Havlin. Fractals in Science. *Springer-Verlag Berlin*, 1994

[5] I. Kanter and D. Kessler. *Phys. Rev. Lett.***74**, 4559, 1995

[6] M. C. Davey and D. J. C. Mackay. Low Density Parity Check Codes Over GF(q). *IEEE Communications Letters*, Vol 2, No. 6, June 1998

[7] Note that $X_i \in \{-1, 1\}$ instead of $\{1, 0\}$ (the original binary values). This mapping was introduced by N. Sourlas, *Nature* **339**, 6227, 1989

[8] R. J. Baxter. Exactly Solved Models In Statistical Mechanics. *Academic Press*, 1982

[9] Corrections to the leading order are of the type $\lambda_{max}^{\frac{L}{2}}\left(1+\sum_i\left(\frac{\lambda_i}{\lambda_{max}}\right)^{\frac{L}{2}}\right)$, see for instance [8]

[10] The entropy of a single correlation constraint is obtained in a straightforward fashion by performing a gauge. For C_1 the gauge $\sigma_i = x_i x_{i+1}$ gives L new variables with a bias $P = \frac{1+C_1}{2}$, and the entropy is given by Shannon's result, eq. 1

[11] Note that imposing short range autocorrelation constraints, C_k, induces long repetitions, symbols, in the sequence. For example, for positive C_1 and C_2, one finds subsequences of the same binary digit, whose length is much greater than 2 (the autocorrelation range).

[12] The sign of each interaction can be positive or negative, leading to frustration. As the entropy approaches the transition, the absolute values of the interactions ($|J_i|$) diverge.

[13] S. Y. Chung, G. D. Forney, T. J. Richardson and R. L. Urbanke. On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit. *IEEE Commun. Lett.* **5** (2). 2001

[14] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke. Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes. *IEEE Trans. Information Theory*, vol. **47**, pp. 619-656, 1978

[15] I. Kanter and D. Saad. *Phys. Rev. Lett.* **83**, 2660, 1999

[16] The range of m and n is specified for the MN class of LDPCC, for Gallager-type codes $m \in \left\{1 \ldots \frac{N-k}{t}\right\}$, $n \in \left\{1 \ldots \frac{N}{t}\right\}$

[17] The interaction strengths for the simulations shown in figure 3 are: $(y_1^\star \simeq 0.30, y_2^\star \simeq 0.61)/(y_1^\star \simeq -0.31, y_2^\star \simeq 0.23, y_3^\star \simeq -0.23)$ for the 2/3 autocorrelation constraints respectively.

[18] The number of decoder iterations ranges from < 10 iterations when the noise level is low, to a few tens of iterations when the threshold of the code is approached. The maximal noise level tolerated by the decoder is within 10% of f_∞ (the bound for the noise level for infinite length messages) in all 3 cases.

[19] Given the channel characterization, f, the usual method would consist of applying standard compression algorithms (e.g. compress, gzip, bzip2), and then sending the compressed message at the appropriate rate. For the 2 classes shown in figure 3, the measured compression in simulations is about 80% of the entropy (for sequences up to 10^6). Our results are superior in that they utilize the message entropy to a greater extent, as can be seen from figure 3.

[20] The number of autocorrelation functions grows exponentially with the block size, as do the number of symbols for a given block, which are used in most compression techniques.