

Statistical Physics of Random Binning ^{*}

Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E-mail: merhav@ee.technion.ac.il

Abstract

We consider the model of random binning and finite-temperature decoding for Slepian-Wolf codes, from a statistical-mechanical perspective. While ordinary random channel coding is intimately related to the random energy model (REM) – a statistical-mechanical model of disordered magnetic materials, it turns out that random binning (for Slepian-Wolf coding) is analogous to another, related statistical mechanical model of strong disorder, which we call the random dilution model (RDM). We use the latter analogy to characterize phase transitions pertaining to finite-temperature Slepian-Wolf decoding, which are somewhat similar, but not identical, to those of finite-temperature channel decoding. We then provide the exact random coding exponent of the bit error rate (BER) as a function of the coding rate and the decoding temperature, and discuss its properties. Finally, a few modifications and extensions of our results are outlined and discussed.

Index Terms Slepian-Wolf codes, error exponent, bit-error probability, finite-temperature decoding, random energy model, phase transitions, phase diagram.

^{*}This research was supported by the Israel Science Foundation (ISF), grant no. 412/12.

1 Introduction

The famous paper by Slepian and Wolf [16], on separate (almost) lossless compression and joint decompression of statistically dependent sources, has triggered an intensive research activity of information theorists during the last forty years. Among its various generalizations and modifications, several recent works have been dedicated to detailed performance analysis, first and foremost, to exponential error bounds for the Slepian–Wolf (SW) decoder. Specifically, Gallager [6] obtained a lower bound on the random coding error exponent associated with random binning, by employing a very similar technique to the one he used in his famous derivation of the random (channel) coding error exponent [5, Sections 5.5–5.6]. A few years later, this error exponent was shown by Csiszár, Körner and Marton [2], [4] to be achievable by a universal decoder, that is independent of the channel. In [3] Csiszár and Körner have studied universally achievable error exponents pertaining to linear codes, as well as ordinary (non-universal) expurgated exponents. Later, Csiszár [1] and Oohama and Han [13] have developed error exponents for situations of coded side information. For high rates at one of the two encoders, Kelly and Wagner [7] have improved relative to these results, but not in the general case.

This paper continues the above described line of work on exponential error bounds associated with random binning of SW codes, but unlike the previous works mentioned above, this one is more oriented to the statistical–mechanical point of view. Specifically, in analogy to the notion of finite–temperature decoding, originally proposed by Ruján [15] in the context of channel coding (see also [14, Section 6.3.3]), here we examine a similar finite–temperature decoder for SW codes, and analyze it from various aspects. In a nutshell, finite–temperature decoding amounts to an optimal symbol–error–probability decoder that is associated with the likelihood function, raised to some power $\beta \geq 0$, a parameter referred to as the *inverse temperature*, which is a term borrowed from equilibrium statistical mechanics and the Boltzmann–Gibbs distribution. In channel coding, the motivation for this parametrization by β could either stem from uncertainty concerning the channel SNR (which is analogous to the “temperature” of the real channel), or from the fact that it enables to analyze both the optimal symbol–error–probability decoder and the optimal block–error–probability decoder on the same footing ($\beta = 1$ and $\beta \rightarrow \infty$, respectively [14, p. 118]). In SW coding, the motivations are similar.

Focusing mostly on the “one-sided” version of the Slepian–Wolf setting, where one source is compressed while the other one is available (e.g., after perfect reconstruction) as side information at the decoder, we derive several results in this paper. First, we present a statistical–mechanical model, henceforth referred to as the *random dilution model* (RDM), which is a natural analogue of the random binning mechanism, exactly in the same way that the random energy model (REM) of statistical mechanics is the physical analogue of the random coding mechanism (see also [8], [14, Chapters 5 and 6] and references therein). The RDM was already mentioned briefly in an earlier work [10], but was not developed in detail therein.

Secondly, in analogy to the phase diagrams of finite-temperature channel coding, provided in [9] and [14, Sect. 6.3.3], here we provide a statistical–mechanical characterization in the form of a *phase diagram* of SW codes with random binning, in the plane of R vs. T , where R is the coding rate and $T = 1/\beta$ is the decoding temperature. As in channel coding, there are three different phases, in which the kinds of behavior of the posterior distribution of the source given the side information and the bin index, are completely different. We will elaborate on these phases in the sequel. Generally speaking, the phase diagram of a finite-temperature SW decoder appears similar to the “mirror image” of the one of channel coding, where the axis of the rate R is flipped over. On the one hand, this seems to make sense, in view of the fact that a SW code at rate R is nearly equivalent to a channel code at rate $H - R$, where H is the entropy of the compressed source. However, this equivalence is not perfect, as there are also some non-trivial differences between channel coding and SW coding. Accordingly, there are differences also in the phase diagram beyond the aforementioned “mirror reflection”.

Next, we derive the exact exponent of the symbol error probability of the finite-temperature decoder, as a function of R and β , and we make a few observations concerning the properties of the resulting error exponent, denoted $E(R, \beta)$. It turns out that $E(R, \beta)$ also exhibits phase transitions, which are related to the above mentioned phases of the posterior, but are not identical.

Finally, we outline a few extensions and modifications of the above described results, in several directions, including: mismatched decoding, universal decoding, variable-rate coding, and the “two-sided” SW problem, where both sources are compressed separately and decompressed jointly.

The outline of the paper is as follows. In Section 2, we establish notation conventions, define the

problem setting, and provide some physics background. In Section 3, we derive the phase diagram, and in Section 4, we derive the error exponent of the finite-temperature decoder.

2 Notation Conventions, Problem Setting and Background

2.1 Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\mathbf{X} = (X_1, \dots, X_N)$, (N – positive integer) may take a specific vector value $\mathbf{x} = (x_1, \dots, x_N)$ in \mathcal{X}^N , the N -th order Cartesian power of \mathcal{X} , which is the alphabet of each component of this vector.

The expectation operator will be denoted by $\mathbf{E}\{\cdot\}$. Logarithms and exponents will be understood to be taken to the natural base unless specified otherwise. The indicator function will be denoted by $\mathcal{I}(\cdot)$. The notation function $[t]_+$ will be defined as $\max\{t, 0\}$. For two positive sequences, $\{a_N\}$ and $\{b_N\}$, the notation $a_N \doteq b_N$ will mean asymptotic equivalence in the exponential scale, that is, $\lim_{N \rightarrow \infty} \frac{1}{N} \log\left(\frac{a_N}{b_N}\right) = 0$. Similarly, $a_N \dot{\leq} b_N$ will mean $\limsup_{N \rightarrow \infty} \frac{1}{N} \log\left(\frac{a_N}{b_N}\right) \leq 0$, and so on.

2.2 Problem Setting and Objectives

Let $\{(X_i, Y_i)\}_{i=1}^N$ be N independent copies of a random vector (X, Y) , distributed according to a given probability mass function $P(x, y)$, where x and y take on values in finite alphabets, \mathcal{X} and \mathcal{Y} , respectively. The source vector $\mathbf{x} = (x_1, \dots, x_N)$, which is a generic realization of $\mathbf{X} = (X_1, \dots, X_N)$, is compressed at the encoder by random binning, that is, each N -tuple $\mathbf{x} \in \mathcal{X}^N$ is randomly and independently assigned to one out of $M = e^{NR}$ bins, where R is the coding rate in nats per symbol. Given a realization of the random partitioning into bins (revealed to both the encoder and the decoder), let $f : \mathcal{X}^N \rightarrow \{0, 1, \dots, M - 1\}$ denote the encoding function, i.e., $u = f(\mathbf{x})$ is the encoder output. Accordingly, the inverse image of u , defined as $f^{-1}(u) = \{\mathbf{x} : f(\mathbf{x}) = u\}$, is the bin of all source vectors mapped by the encoder into u . The decoder has access to u and to $\mathbf{y} = (y_1, \dots, y_N)$, which is a realization of $\mathbf{Y} = (Y_1, \dots, Y_N)$, namely, the side information at the

decoder.

As is well known, the optimal decoder in the sense of minimum word error probability is the word-level maximum a-posteriori (MAP) decoder,

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in f^{-1}(u)} P(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x} \in f^{-1}(u)} P(\mathbf{x}, \mathbf{y}). \quad (1)$$

Similarly, the optimal decoder in the sense of minimum symbol error probability is given by the symbol-level MAP decoder,

$$\hat{x}_i = \arg \max_{x \in \mathcal{X}} \sum_{\mathbf{x} \in f^{-1}(u): x_i=x} P(\mathbf{x}, \mathbf{y}), \quad i = 1, 2, \dots, N. \quad (2)$$

Following the notion of finite-temperature decoding in channel coding [15] (see also [14, Section 6.3.3]), we consider a parametric family of decoders, that generalizes both (1) and (2), and which is of the form

$$\hat{x}_i = \arg \max_{x \in \mathcal{X}} \sum_{\mathbf{x} \in f^{-1}(u): x_i=x} P^\beta(\mathbf{x}, \mathbf{y}), \quad i = 1, 2, \dots, N, \quad (3)$$

where the parameter $\beta \geq 0$ is referred to as the *inverse temperature*, a term borrowed from equilibrium statistical physics (see next subsection). The motivation for considering the finite-temperature decoder is two-fold (see also [9] for even more motivations): First, as said, it is a common generalization of both the symbol-level MAP decoder ($\beta = 1$) and the word-level MAP decoder ($\beta \rightarrow \infty$). Secondly, in some important cases, it refers to a situation of a certain mismatch that may stem from uncertainty concerning the joint distribution of (X, Y) , or even more specifically, the quality of the ‘channel’ $P(y|x)$, connecting X to Y . For example, if (X, Y) is a double binary symmetric source (BSS), that is, X is a BSS and Y given X is generated by binary symmetric channel (BSC), then the choice of β manifests the decoder’s ‘belief’ concerning the quality of this BSC: $\beta < 1$ corresponds to a pessimistic decoder, whereas $\beta > 1$ is associated with an optimistic one.

In order to understand the behavior of the finite-temperature decoder (3), one has to first gain some insight concerning the behavior the posterior distribution induced by that decoder, namely,

$$P_\beta(\mathbf{x}|\mathbf{y}, u) = \begin{cases} \frac{P^\beta(\mathbf{x}, \mathbf{y})}{\sum_{\mathbf{x}' \in f^{-1}(u)} P^\beta(\mathbf{x}', \mathbf{y})} & \mathbf{x} \in f^{-1}(u) \\ 0 & \text{elsewhere} \end{cases} \quad (4)$$

Accordingly, we first focus on studying the properties of this posterior in the random binning regime. In particular, similarly as in [9] and [14], we view it as an instance of the Boltzmann–Gibbs

(B-G) distribution of statistical mechanics by rewriting it in the form

$$P_\beta(\mathbf{x}|\mathbf{y}, u) = \begin{cases} \frac{\exp\{-\beta\mathcal{E}(\mathbf{x}, \mathbf{y})\}}{\sum_{\mathbf{x}' \in f^{-1}(u)} \exp\{-\beta\mathcal{E}(\mathbf{x}', \mathbf{y})\}} & \mathbf{x} \in f^{-1}(u) \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

where $\mathcal{E}(\mathbf{x}, \mathbf{y})$ is the energy function (Hamiltonian), defined as $\mathcal{E}(\mathbf{x}, \mathbf{y}) \triangleq -\ln P(\mathbf{x}, \mathbf{y})$. We then study the phase diagram of the corresponding statistical–mechanical model in the plane of R vs. β , or more precisely, R vs. T , where $T = 1/\beta$ is the decoding temperature. In analogy to the role of the random energy model (REM) as the statistical–mechanical counterpart of ordinary random coding (see [8, Chap. 6], [14, Chapters 5 and 6]), it turns out that random binning corresponds to a somewhat different (though related) physical model, which we call the random dilution model (RDM), and which was first mentioned in [10]. The RDM and its relevance to random binning will be presented in the next subsection.

Our second objective would be to derive the *exact* error exponent of the symbol error probability, denoted $E(R, \beta)$, that is associated with the finite–temperature decoder (3), as a function of R and β , in the random binning regime. The properties of $E(R, \beta)$, as well as its phase diagram, will be studied in some detail.

Finally, we briefly discuss several variations of these results, covering situations of mismatch, universal decoding, variable–rate compression, and the “two–sided” version of SW coding, namely, separate encodings and joint decoding of both sources (as opposed to the “one–sided” version described above, of encoding and decoding of one source while the other one serves as side information at the decoder).

2.3 Background: The REM, the RDM and Random Binning

In ordinary random coding, the analysis of bounds on the probability of error (especially in Gallager’s method) are often associated with expressions of the form $\sum_{\mathbf{x} \in \mathcal{C}} P^\beta(\mathbf{y}|\mathbf{x})$, where $P(\mathbf{y}|\mathbf{x})$ is the conditional distribution pertaining to the channel, \mathcal{C} is randomly drawn codebook and $\beta > 0$ is a parameter. As described in [8, Chap. 6], from the statistical–mechanical perspective, this can be thought of as a partition function (depending on \mathbf{y}):

$$Z(\beta|\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{C}} e^{-\beta\mathcal{E}(\mathbf{x}, \mathbf{y})}, \quad (6)$$

where β is the inverse temperature and where here the energy function is $\mathcal{E}(\mathbf{x}, \mathbf{y}) = -\ln P(\mathbf{y}|\mathbf{x})$. The same partition function is relevant for the Boltzmann–Gibbs form of the finite–temperature posterior associated with the channel decoder, in the spirit of eqs. (4) and (5):

$$P_\beta(\mathbf{x}|\mathbf{y}) = \begin{cases} \frac{P^\beta(\mathbf{y}|\mathbf{x})}{Z(\beta|\mathbf{y})} & \mathbf{x} \in \mathcal{C} \\ 0 & \text{elsewhere} \end{cases} \quad (7)$$

$$= \begin{cases} \frac{\exp\{-\beta\mathcal{E}(\mathbf{x}, \mathbf{y})\}}{Z(\beta|\mathbf{y})} & \mathbf{x} \in \mathcal{C} \\ 0 & \text{elsewhere} \end{cases} \quad (8)$$

As the codewords are selected independently at random, then for fixed \mathbf{y} , the energy values $\{\mathcal{E}(\mathbf{x}, \mathbf{y}), \mathbf{x} \in \mathcal{C}\}$ are i.i.d. random variables. This is essentially the same as in the *random energy model* (REM), a well known model of disorder in statistical physics of spin glasses, which undergoes a phase transition: below a certain temperature ($\beta > \beta_c$), the system is frozen in the sense that the partition function is dominated by a non–exponential number of microstates $\{\mathbf{x}\}$ at the ground–state energy (zero thermodynamical entropy). This is called the *frozen phase* or the *glassy phase*. The other phase, $\beta < \beta_c$, is called the *paramagnetic phase* (see more details in [14, Chap. 5]). Owing to this analogy between the REM and random coding, the corresponding exponential error bounds associated with random coding then undergo a similar phase transition (see [8] and references therein).

In random binning, as opposed ordinary random coding, the mechanism is somewhat different, and the analogous statistical–mechanical model, which we call the RDM, is defined as follows. Consider a partition function of a certain physical system, with a microstate \mathbf{x} and Hamiltonian $\mathcal{E}(\mathbf{x})$, i.e.,

$$Z(\beta) = \sum_{\mathbf{x} \in \mathcal{X}^N} e^{-\beta\mathcal{E}(\mathbf{x})}, \quad \beta > 0, \quad (9)$$

where β is, as said, the inverse temperature. The *diluted* version of $Z(\beta)$, according to the RDM (hence the name), is defined as

$$Z_D(\beta) = \sum_{\mathbf{x} \in \mathcal{X}^N} I(\mathbf{x}) \cdot e^{-\beta\mathcal{E}(\mathbf{x})}, \quad (10)$$

where $\{I(\mathbf{x}), \mathbf{x} \in \mathcal{X}^N\}$ are i.i.d. Bernoulli random variables with $p \triangleq \Pr\{I(\mathbf{x}) = 1\} = 1 - \Pr\{I(\mathbf{x}) = 0\} = e^{-NR}$ for all $\mathbf{x} \in \mathcal{X}^N$, and $R \geq 0$ is a given constant. Thus, $Z_D(\beta)$ is a partial version of the full partition function $Z(\beta)$, with randomly chosen (surviving) microstates $\{\mathbf{x}\}$. Equivalently,

$Z_D(\beta)$ can be thought of as being defined just like $Z(\beta)$, but with an Hamiltonian redefined as $\mathcal{E}_D(\mathbf{x}) = \mathcal{E}(\mathbf{x}) + \psi(\mathbf{x})$, where $\psi(\mathbf{x})$ are i.i.d. random variables, taking the value $\psi(\mathbf{x}) = 0$ with probability e^{-NR} and the value $\psi(\mathbf{x}) = \infty$ with probability $1 - e^{-NR}$. From the physical point of view, $\psi(\mathbf{x})$ can be thought of as some disordered potential energy function, that due to long-range interactions, disables access to certain points in the configuration space (those that have not ‘survived’ the dilution).

Let $s(\epsilon)$ denote the normalized (per-particle) entropy as a function of the normalized energy, associated with the full system, $Z(\beta)$. More precisely, denoting by $\Omega_N(E)$ the number of vectors $\{\mathbf{x}\}$ for which $\mathcal{E}(\mathbf{x}) = E$, then

$$s(\epsilon) \triangleq \lim_{N \rightarrow \infty} \frac{\ln \Omega_N(N\epsilon)}{N}, \quad (11)$$

provided that the limit exists. Let $\Delta\epsilon$ be an arbitrarily small quantity (increment) of the normalized energy. Then,

$$Z_D(\beta) \approx \sum_i \left[\sum_{\mathbf{x}: Ni\Delta\epsilon \leq \mathcal{E}(\mathbf{x}) < N(i+1)\Delta\epsilon} I(\mathbf{x}) \right] \cdot e^{-\beta Ni\Delta\epsilon}. \quad (12)$$

Observe that the expression in the square brackets is a binomial RV with exponentially about $e^{Ns(i\Delta\epsilon)}$ trials and probability of success $p = e^{-NR}$. Thus, in a typical realization the RDM, this number is about $\exp\{N[s(i\Delta\epsilon) - R]\}$ whenever $s(i\Delta\epsilon) - R \geq 0$ and by zero otherwise. Thus, in the limit of $\Delta\epsilon \rightarrow 0$,

$$\phi_D(\beta) = \lim_{N \rightarrow \infty} \frac{\ln Z_D(\beta)}{N} = \sup_{\{\epsilon: s(\epsilon) \geq R\}} [s(\epsilon) - R - \beta\epsilon] \quad (13)$$

$$= \begin{cases} \phi(\beta) - R & \beta < \beta_c \\ -\beta\epsilon_0 & \beta \geq \beta_c \end{cases} \quad (14)$$

where $\phi(\beta)$ is the asymptotic normalized log-partition function associated with the full system (before the dilution), ϵ_0 is ground-state (minimum energy state) of the RDM, i.e., the solution to the equation $s(\epsilon) = R$, and $\beta_c = s'(\epsilon_0)$, $s'(\cdot)$ being the derivative of $s(\cdot)$. Note that both β_c and ϵ_0 depend on R . Thus, the system undergoes a glassy phase transition (similarly to the REM) at inverse temperature β_c . At low temperatures, it freezes at zero entropy and $Z_D(\beta)$ is dominated by a subexponential number of configurations $\{\mathbf{x}\}$ at the ground-state energy $E_0 = N\epsilon_0$. This is then the glassy phase of the RDM. Above the critical temperature (the paramagnetic phase), the diluted system behaves essentially like the full one, i.e., it is dominated by exponentially many (actually,

about $e^{N[s(\epsilon_\beta)-R]}$ configurations with energy $E_\beta = N\epsilon_\beta$, where ϵ_β is the typical per-particle energy level pertaining to inverse temperature β in the full system, that is, the solution to the equation $s'(\epsilon) = \beta$. Note that

$$\beta_c(R) = s'[s^{-1}(R)], \quad (15)$$

where $s^{-1}(\cdot)$ is the inverse function of $s(\cdot)$ (in the range where it is monotonically increasing and hence one-to-one). In this range $s^{-1}(\cdot)$ is increasing as well, but since s' is decreasing, then β_c is a decreasing function of R . This means that, as the dilution becomes more aggressive, the critical temperature goes up. Since $s(\cdot)$ is concave (see, e.g., [8, pp. 13–14]) and increasing, $s^{-1}(\cdot)$ is convex, so the overall behavior depends on $s'(\cdot)$.

Example 1. Let $\mathcal{E}(\mathbf{x}) = \frac{\kappa}{2}\|\mathbf{x}\|^2$, with $\mathcal{X} = \{0, \pm a, \pm 2a, \dots\}$, i.e., an harmonic potential applied to particles in a grid with spacing a . Then, $\Omega_N(E)$ is approximately the volume of the shell of a hyper-sphere of radius $\sqrt{2E/\kappa}$, divided by an elementary volume of the grid cube a^N , which yields

$$s(\epsilon) = \frac{1}{2} \ln \frac{4\pi e \epsilon}{\kappa a^2}, \quad (16)$$

and so,

$$s'(\epsilon) = \frac{1}{2\epsilon}. \quad (17)$$

and

$$s^{-1}(R) = \frac{\kappa a^2}{4\pi e} \cdot e^{2R}. \quad (18)$$

Thus,

$$\beta_c(R) = \frac{2\pi e}{\kappa a^2} \cdot e^{-2R}, \quad (19)$$

meaning that the critical temperature grows exponentially with R . This concludes Example 1.

Remark 1. A slightly more general version of the RDM replaces the fixed parameter R by a function of ϵ , that is, $p = e^{-NR(\epsilon)}$. The analysis is essentially the same as before, except that now, the range of maximization $\{\epsilon : s(\epsilon) \geq R\}$ is replaced by $\{\epsilon : s(\epsilon) \geq R(\epsilon)\}$, which, depending on the form of the function $R(\cdot)$, might be rather different.

Finally, to see the relevance of the RDM to random binning for SW coding, let us return to the problem setting described in Subsection 2.2, and consider the partition function

$$Z(\beta|\mathbf{y}, u) = \sum_{\mathbf{x} \in f^{-1}(u)} P^\beta(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{X}^N} \exp\{-\beta \mathcal{E}(\mathbf{x}, \mathbf{y})\} \cdot \mathcal{I}[f(\mathbf{x}) = u], \quad (20)$$

pertaining to the Boltzmann-Gibbs distribution (5). We can think of this as an instance of the RDM, with $I(\mathbf{x}) = \mathcal{I}[f(\mathbf{x}) = u]$, i.e., the microstates $\{\mathbf{x}\}$ that ‘survive’ the dilution are only those for which the randomly selected bin index happens to coincide with the given u , which is the case with probability e^{-NR} , exactly like in the above defined RDM.

3 Phase Diagram of the Finite-Temperature Posterior

In this subsection, we characterize the phase diagram of the partition function (20), pertaining to the finite-temperature posterior (5), for a typical realization of the random binning scheme and a typical realization of (\mathbf{X}, \mathbf{Y}) . Generally speaking, this derivation is in the spirit of those in [8, Chap. 6] and [14, Section 6.3.3], but there are some important differences.

We begin by decomposing $Z(\beta|\mathbf{y}, u)$ as

$$Z(\beta|\mathbf{y}, u) = Z_c(\beta|\mathbf{y}, u) + Z_e(\beta|\mathbf{y}, u), \quad (21)$$

where $Z_c(\beta|\mathbf{y}, u) = e^{-\beta\mathcal{E}(\mathbf{x}, \mathbf{y})}$ is the contribution of the *correct* \mathbf{x} that was actually emitted by the source, whereas $Z_e(\beta|\mathbf{y}, u)$ is the sum of contributions of all other source vectors. For a typical realization (\mathbf{x}, \mathbf{y}) of (\mathbf{X}, \mathbf{Y}) , $\mathcal{E}(\mathbf{x}, \mathbf{y})$ is about $NH(X, Y)$ (by the weak law of large numbers), and so, $Z_c(\beta|\mathbf{y}, u)$ is about $e^{-\beta NH(X, Y)}$, where $H(X, Y)$ is the joint entropy of (X, Y) . What makes the real \mathbf{x} emitted having a special stature here is the fact that it surely survives the dilution, as u is, by definition, the bin index of \mathbf{x} .

We next address the behavior of the second term, $Z_e(\beta|\mathbf{y}, u)$. To this end, we need the entropy function $s(\epsilon)$ (see (11)) of the full (non-diluted) system. Using the method of types, it is easily seen that for a typical \mathbf{y} , this function is given by

$$s(\epsilon) = \max_{\{Q(x|y): -\sum_{x,y} P(y)Q(x|y) \ln P(x,y) = \epsilon\}} \sum_y P(y) \sum_x Q(x|y) \ln \frac{1}{Q(x|y)}. \quad (22)$$

It would be instructive and useful to characterize the form of the optimal ‘channel’ $\{Q(x|y)\}$, call it $\{Q^*(x|y)\}$, that achieves $s(\epsilon)$. Intuitively, $s(\epsilon)$ can be thought of as the overall per-particle entropy of a mixture of systems, indexed by y , each one with $NP(y)$ particles and Hamiltonian $\mathcal{E}(x, y) = -\ln P(x, y)$, where x plays the role of a micro-state and y is the index. In thermal equilibrium, all systems are at the same temperature, which we will denote by $\tau = 1/\alpha$, and

the Boltzmann factor is proportional to $e^{-\alpha\mathcal{E}(x,y)} = P^\alpha(x,y)$, where α is chosen so as to meet the constraint.

More precisely, let $\zeta(\alpha|y) = \sum_x P^\alpha(x,y)$, $\alpha \in \mathbb{R}$. We argue that the conditional distribution $Q^*(x|y)$ that achieves $s(\epsilon)$ is always of the form

$$Q^*(x|y) = Q_\alpha(x|y) \triangleq \frac{P^\alpha(x,y)}{\zeta(\alpha|y)}, \quad (23)$$

where α is chosen to satisfy the constraint

$$\sum_{x,y} P(y) Q_\alpha(x|y) \mathcal{E}(x,y) = - \sum_{x,y} P(y) Q_\alpha(x|y) \ln P(x,y) = \epsilon. \quad (24)$$

Note that for $\alpha \rightarrow \infty$, $Q_\alpha(x|y)$ tends to put all its mass on the letter x which maximizes $P(x,y)$ and the resulting energy is $\epsilon_{\min} = \sum_y P(y) \min_x \ln[1/P(x,y)]$, whereas for $\alpha \rightarrow -\infty$, $Q_\alpha(x|y)$ tends to put all its mass on the letter x which minimizes $P(x,y)$ and the resulting energy is $\epsilon_{\max} = \sum_y P(y) \max_x \ln[1/P(x,y)]$. Thus, as α exhausts the real line, the entire energy range $(\epsilon_{\min}, \epsilon_{\max})$ is covered. The optimality of $Q_\alpha(x|y)$ follows from the following consideration:

$$\begin{aligned} 0 &\leq \sum_{x,y} P(y) Q(x|y) \ln \frac{Q(x|y)}{Q_\alpha(x|y)} \\ &= \sum_{x,y} P(y) Q(x|y) \ln \frac{Q(x|y) \zeta(\alpha|y)}{P^\alpha(x,y)} \\ &= \sum_y P(y) \ln \zeta(\alpha|y) + \alpha \epsilon + \sum_y P(y) \sum_x Q(x|y) \ln Q(x|y) \end{aligned} \quad (25)$$

or

$$\sum_y P(y) \sum_x Q(x|y) \ln \frac{1}{Q(x|y)} \leq \sum_y P(y) \ln \zeta(\alpha|y) + \alpha \epsilon, \quad (26)$$

with equality for $Q(x|y) = Q_\alpha(x|y)$. It is also seen that

$$s(\epsilon) = \sum_y P(y) \ln \zeta(\alpha|y) + \alpha \epsilon, \quad (27)$$

where it should be kept in mind that α is itself a function of ϵ , defined by the constraint (24).

Now, as $Z_c(\beta|\mathbf{y}, u)$ is associated with the RDM, it has two phases: the paramagnetic phase and the glassy phase. in the plane of β vs. R , where the boundary is $\beta = \beta_c(R)$ with the entropy function $s(\cdot)$ as above. The contribution of $Z_c(\beta|\mathbf{y}, u)$ introduces a third phase – the so called, *ordered phase* or *ferromagnetic phase*. The ferromagnetic phase dominates the glassy phase when

$\beta H(X, Y) \leq \beta s^{-1}(R)$, namely, when $R \geq s[H(X, Y)]$. Now, we argue that $s[H(X, Y)] = H(X|Y)$. To prove this, first observe that obviously, $s[H(X, Y)] \geq H(X|Y)$ (by choosing $Q(x|y) = P(x|y)$). On the other hand, the reversed inequality is obtained by repeating eq. (25) with the choice $\alpha = 1$. Thus, the boundary between the ferromagnetic phase and the glassy phase is given by $R = H(X|Y)$ (the vertical line in the phase diagram of Fig. 1). Note also that $\beta_c[H(X|Y)] = 1$.

Now, the normalized log-partition function of the non-diluted system, $\phi(\beta)$, is obtained by

$$e^{N\phi(\beta)} = \sum_{\mathbf{x}} P^\beta(\mathbf{x}, \mathbf{y}) \quad (28)$$

$$= \sum_{\mathbf{x}} \prod_{i=1}^N P^\beta(x_i, y_i) \quad (29)$$

$$= \prod_{i=1}^N \left[\sum_x P^\beta(x, y_i) \right] \quad (30)$$

$$= \prod_{y \in \mathcal{Y}} \left[\sum_x P^\beta(x, y) \right]^{NP(y)} \quad (31)$$

$$= \exp \left\{ N \sum_{y \in \mathcal{Y}} P(y) \ln \left[\sum_x P^\beta(x, y) \right] \right\}, \quad (32)$$

i.e.,

$$\phi(\beta) = \sum_{y \in \mathcal{Y}} P(y) \ln \left[\sum_x P^\beta(x, y) \right]. \quad (33)$$

It follows that for the ferromagnetic component to dominate also the paramagnetic component, we must have

$$\beta H(X, Y) \leq - \sum_{y \in \mathcal{Y}} P(y) \ln \left[\sum_x P^\beta(x, y) \right] + R, \quad (34)$$

or, equivalently,

$$R \geq \beta H(X, Y) + \sum_{y \in \mathcal{Y}} P(y) \ln \left[\sum_x P^\beta(x, y) \right] \triangleq \Gamma(\beta), \quad (35)$$

and so the boundary is given by $R = \Gamma(\beta)$ or $T = 1/\Gamma^{-1}(R)$. The boundary between the glassy phase and the paramagnetic phase is, of course, $\beta = \beta_c(R)$ or $T = T_c(R) \triangleq 1/\beta_c(R)$, as mentioned already in the general discussion on the RDM. The phase diagram of finite-temperature random binning appears in Fig. 1, as a partition of the plane of $T = 1/\beta$ vs. R into the three regions mentioned. As can be seen, qualitatively speaking, it looks quite like the mirror image of the phase

diagram of random coding for channels [9], [14, p. 119, Fig. 6.5], since a rate R SW code essentially operates like a channel code at rate $H(X) - R$. However, the equations of the boundary curves $T = T_c(R)$ and $T = 1/\Gamma^{-1}(R)$ here and in channel coding are completely different due to several reasons:

1. In SW coding, the typical size of a bin (which is analogous to the size of the corresponding channel codebook) is a random variable, which fluctuates around $|\mathcal{X}|^N \cdot e^{-NR}$. Only about $\exp\{N[H(X) - R]\}$ members of this bin are typical to the source, but when error exponents and large deviations effects are considered, the a-typical bin members may also play a non-trivial role.
2. Unlike in traditional channel coding, the prior of the input is not necessarily uniform across the bin, as it depends on the source (just like in joint source–channel coding).
3. The compositions (types) of the codewords are random.

All these differences make the analogy between SW coding and channel coding rather non-trivial in our context. A few comments are in order concerning possible extensions and modifications of the above phase diagram.

1. *Variable–rate SW codes.* Variable–rate SW codes may be related to the generalized version of the RDM that allows R to be energy–dependent (see Remark 1). In the context of variable–rate SW coding, this requires a slight modification, as R may be allowed to depend only on \mathbf{x} , but not on \mathbf{y} . A plausible approach is then to let R depend on \mathbf{x} only via the type class of \mathbf{x} (see [18]).
2. *Mismatch.* The above analysis can easily be extended to a situation of a general mismatch. Suppose that the partition function is defined in terms of a mismatched model $\tilde{P}(\mathbf{x}, \mathbf{y})$, where we assume, without loss of generality, that $\tilde{P}(y) = P(y)$, because as far as the finite–temperature decoder is concerned, a general $\tilde{P}(\mathbf{x}, \mathbf{y})$ is equivalent to $P(\mathbf{y})\tilde{P}(\mathbf{x}|\mathbf{y})$, where $\tilde{P}(\mathbf{x}|\mathbf{y})$ is the conditional distribution induced by $\tilde{P}(\mathbf{x}, \mathbf{y})$. Accordingly, in Fig. 1, the ferromagnetic–glassy boundary would be replaced by the vertical straight line

$$R = -\mathbf{E} \ln \tilde{P}(X|Y) = -\sum_{x,y} P(x,y) \ln \tilde{P}(x|y),$$

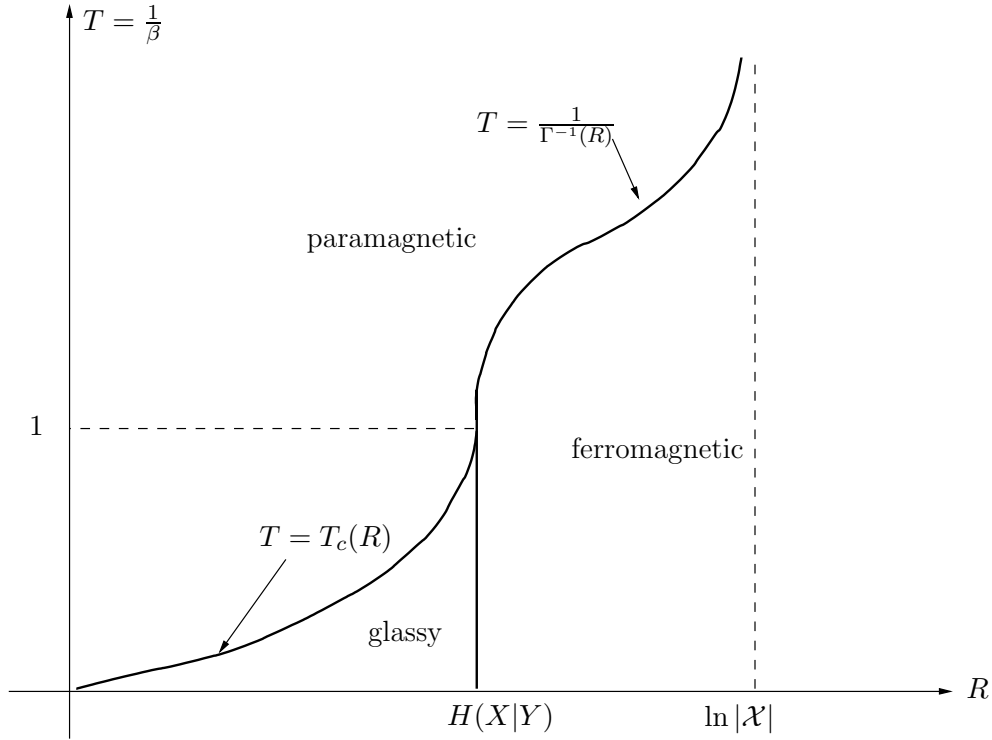


Figure 1: Phase diagram of $Z(\beta|\mathbf{y})$ (for a typical \mathbf{y}) in the plane of the decoding temperature T vs. the coding rate R .

the ferromagnetic–paramagnetic boundary would be modified to

$$R = \tilde{\Gamma}(\beta) = -\beta \mathbf{E} \ln \tilde{P}(X, Y) + \sum_y P(y) \ln \left[\sum_x \tilde{P}^\beta(x, y) \right],$$

and the paramagnetic–glassy boundary would become

$$\beta = \tilde{\beta}_c(R) = \tilde{s}'[\tilde{s}^{-1}(R)],$$

where $\tilde{s}(\epsilon)$ is defined similarly as $s(\epsilon)$, except that $P(x, y)$ is replaced by $\tilde{P}(x, y)$.

3. *Universal decoding.* It is interesting to analyze similarly the partition function pertaining to a finite–temperature version of the (universal) minimum conditional entropy decoder. The only difference is that here, the Hamiltonian is replaced by $\mathcal{E}(\mathbf{x}, \mathbf{y}) = -N \sum_y P(y) \sum_x Q(x|y) \ln Q(x|y)$, where $Q(x|y)$ is the conditional empirical distribution of X given Y , induced from (\mathbf{x}, \mathbf{y}) . Obviously, in this case, $s(\epsilon) = \epsilon$, and so, for a typical \mathbf{y} :

$$\lim_{N \rightarrow \infty} \frac{\ln Z_\epsilon(\beta|\mathbf{y})}{N} = \sup_{\{\epsilon: \epsilon \geq R\}} [\epsilon(1 - \beta) - R] \quad (36)$$

$$= \begin{cases} (1 - \beta) \ln |\mathcal{X}| - R & \beta < 1 \\ -\beta R & \beta \geq 1 \end{cases} \quad (37)$$

which means that the critical temperature is always $T_c = 1$, independently of R . Thus, the paramagnetic–glassy boundary becomes the horizontal straight line $T_c = 1$. The paramagnetic–ferromagnetic boundary is now

$$R = (1 - \beta) \ln |\mathcal{X}| + \beta H(X|Y),$$

or equivalently,

$$T = \frac{\ln |\mathcal{X}| - H(X|Y)}{\ln |\mathcal{X}| - R}.$$

The phase diagram is depicted in Fig. 2. As can be seen, the price of the universality is that the paramagnetic phase partly ‘invades’ into the previous area of the ferromagnetic phase, and that, similarly, the glassy phase has expanded at the expense of the paramagnetic phase.

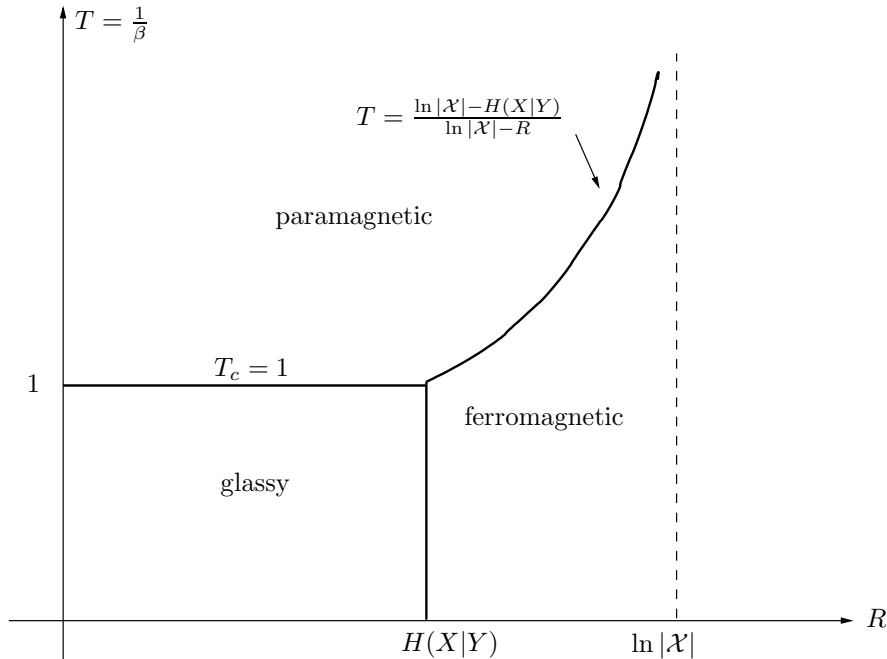


Figure 2: Phase diagram of the finite–temperature minimum entropy decoder in the plane of the decoding temperature T vs. the coding rate R .

4. *Two–Sided SW Coding.* When both \mathbf{x} and \mathbf{y} are encoded, at rates R_X and R_Y , respectively, the partition function becomes

$$Z(\beta|u, v) = \sum_{\mathbf{x}', \mathbf{y}'} P^\beta(\mathbf{x}', \mathbf{y}') \cdot I[f_X(\mathbf{x}') = u] \cdot I[f_Y(\mathbf{y}') = v]. \quad (38)$$

Here, we should distinguish between four terms:

$$\begin{aligned} Z_{cc}(\beta|u, v) &= P^\beta(\mathbf{x}, \mathbf{y}) = e^{-\beta \ln[1/P(\mathbf{x}, \mathbf{y})]} \\ Z_{cc}(\beta|\mathbf{y}, u, v) &= \sum_{\mathbf{x}' \neq \mathbf{x}} P^\beta(\mathbf{x}', \mathbf{y}) \cdot I[f_X(\mathbf{x}') = u] \end{aligned} \quad (39)$$

$$Z_{cc}(\beta|\mathbf{x}, u, v) = \sum_{\mathbf{y}' \neq \mathbf{y}} P^\beta(\mathbf{x}, \mathbf{y}') \cdot I[f_Y(\mathbf{y}') = v] \quad (40)$$

$$Z_{ee}(\beta|u, v) = \sum_{\mathbf{x}' \neq \mathbf{x}, \mathbf{y}' \neq \mathbf{y}} P^\beta(\mathbf{x}', \mathbf{y}') \cdot I[f_X(\mathbf{x}') = u] \cdot I[f_Y(\mathbf{y}') = v]. \quad (41)$$

As before, $Z_{cc}(\beta|u, v)$ is typically about $e^{-N\beta H(X, Y)}$. $Z_{cc}(\beta|\mathbf{y}, u, v)$ is exactly the same as the earlier $Z_e(\beta|\mathbf{y}, u)$, and so is, $Z_{cc}(\beta|\mathbf{x}, u, v)$, with the roles of \mathbf{x} and \mathbf{y} being interchanged. Thus, we define

$$s_{X|Y}(\epsilon) = \max_{\{Q(x|y): -\sum_{x,y} P(y)Q(x|y) \ln P(x,y) = \epsilon\}} \sum_{x,y} P(y)Q(x|y) \ln \frac{1}{Q(x|y)} \quad (42)$$

$$s_{Y|X}(\epsilon) = \max_{\{Q(y|x): -\sum_{x,y} P(x)Q(y|x) \ln P(x,y) = \epsilon\}} \sum_{x,y} P(x)Q(y|x) \ln \frac{1}{Q(y|x)} \quad (43)$$

and

$$s_{XY}(\epsilon) = \max_{\{Q(x,y): -\sum_{x,y} Q(x,y) \ln P(x,y) = \epsilon\}} \sum_{x,y} Q(x,y) \ln \frac{1}{Q(x,y)} \quad (44)$$

Therefore, we have

$$\lim_{N \rightarrow \infty} \frac{\ln Z_{cc}(\beta)}{N} = \sup_{\{\epsilon: s_{X|Y}(\epsilon) \geq R_X\}} [s_{X|Y}(\epsilon) - R_X - \beta\epsilon] \quad (45)$$

$$= \begin{cases} \phi_X(\beta) - R_X & \beta < \beta_X \\ -\beta\epsilon_X & \beta \geq \beta_X \end{cases} \quad (46)$$

and

$$\lim_{N \rightarrow \infty} \frac{\ln Z_{ce}(\beta)}{N} = \begin{cases} \phi_Y(\beta) - R_Y & \beta < \beta_Y \\ -\beta\epsilon_Y & \beta \geq \beta_Y \end{cases} \quad (47)$$

where

$$\phi_X(\beta) = \sum_y P(y) \ln \left[\sum_x P^\beta(x, y) \right], \quad (48)$$

$$\phi_Y(\beta) = \sum_x P(x) \ln \left[\sum_y P^\beta(x, y) \right], \quad (49)$$

ϵ_X is the solution to the equation $s_{X|Y}(\epsilon) = R_X$, $\beta_X = s'_{X|Y}(\epsilon_X)$, ϵ_Y is the solution to the equation $s_{Y|X}(\epsilon) = R_Y$, and $\beta_Y = s'_{Y|X}(\epsilon_Y)$. Similarly,

$$\lim_{N \rightarrow \infty} \frac{\ln Z_{ee}(\beta)}{N} = \sup_{\{\epsilon: s_{XY}(\epsilon) \geq R_X + R_Y\}} [s_{XY}(\epsilon) - R_X - R_Y - \beta\epsilon] \quad (50)$$

$$= \begin{cases} \phi_{XY}(\beta) - R_X - R_Y & \beta < \beta_{XY} \\ -\beta\epsilon_{XY} & \beta \geq \beta_{XY} \end{cases} \quad (51)$$

where

$$\phi_{XY}(\beta) = \ln \left[\sum_{x,y} P^\beta(x,y) \right], \quad (52)$$

ϵ_{XY} is the solution to the equation $s_{XY}(\epsilon) = R_X + R_Y$, and $\beta_{XY} = s'_{XY}(\epsilon_{XY})$. Here, the phase diagram, in the three-dimensional space (T, R_X, R_Y) is much more involved since each one of the terms $Z_{ee}(\beta)$, $Z_{ce}(\beta)$, $Z_{ec}(\beta)$ could be in two different phases, and on top of that, one should check when $Z_{cc}(\beta)$ dominates. We will not delve into it any further here, but only note that for $\beta \leq 1$ (which guarantees that all three erroneous partition functions are in the paramagnetic phases), $Z_{ee}(\beta)$, $Z_{ec}(\beta)$, $Z_{ce}(\beta)$ and $Z_{cc}(\beta)$, dominates, wherever $R_X + R_Y - \phi_{XY}(\beta)$, $R_X - \phi_X(\beta)$, $R_Y - \phi_Y(\beta)$, and $\beta H(X, Y)$, is the smallest among all four functions, respectively. In particular, we have the following conditions for reliable communication (where $Z_{cc}(\beta)$ dominates):

$$R_X > \beta H(X, Y) + \phi_X(\beta) \quad (53)$$

$$R_Y > \beta H(X, Y) + \phi_Y(\beta) \quad (54)$$

$$R_X + R_Y > \beta H(X, Y) + \phi_{XY}(\beta) \quad (55)$$

For $\beta = 1$, this boils down to the well-known achievability region of SW coding. Note that there are regions where either $Z_{ec}(\beta)$ or $Z_{ce}(\beta)$ dominate, which means that one of the sources is decoded reliably, while the other one is not. As expected, for $\beta = 1$, \mathbf{y} alone is decoded reliably within $\{(R_X, R_Y) : R_X < H(X|Y), R_Y > H(Y)\}$ and \mathbf{x} alone is decoded reliably within $\{(R_X, R_Y) : R_Y < H(Y|X), R_X > H(X)\}$.

4 Performance Evaluation

In this section, we provide an exact analysis of the error exponent, associated with the symbol error probability of the finite-temperature decoder (3). We then discuss some properties of the error exponent as a function of R and β and present a phase diagram. Finally, we discuss some modifications and extensions.

For the sake of simplicity of the exposition, and without any essential loss of generality, we will assume $\mathcal{X} = \{0, 1\}$ and evaluate the expected¹ bit-error rate (BER), $P_b(R, \beta, N) = \Pr\{\hat{x}_1 \neq x_1\}$

¹Expectation w.r.t. the random binning.

(which is the same as $\Pr\{\hat{x}_i \neq x_i\}$ for every $i = 1, 2, \dots, N$ due to symmetry). that is,

$$P_b(R, \beta, N) = \Pr \left\{ \sum_{\{\mathbf{x}': x'_1 \neq x_1\}} P^\beta(\mathbf{x}', \mathbf{y}) \cdot I[f(\mathbf{x}') = f(\mathbf{x})] \geq \sum_{\{\mathbf{x}': x'_1 = x_1\}} P^\beta(\mathbf{x}', \mathbf{y}) \cdot I[f(\mathbf{x}') = f(\mathbf{x})] \right\}, \quad (56)$$

or more precisely, the error exponent associated with $P_b(R, \beta, N)$:

$$\mathbb{E}(R, \beta) \triangleq \lim_{N \rightarrow \infty} \left[-\frac{\ln P_b(R, \beta, N)}{N} \right]. \quad (57)$$

For later use, we also define the following notation.

$$\epsilon(Q_{XY}) \triangleq \frac{1}{N} \ln P(\mathbf{x}, \mathbf{y}) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q_{XY}(x, y) \ln P_{XY}(x, y), \quad (58)$$

where Q_{XY} is understood here to be the joint empirical distribution of $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^N \times \mathcal{Y}^N$. Similarly, for a generic \mathbf{x}' , the corresponding auxiliary random variable will be denoted by X' , so that the joint empirical distribution of $(\mathbf{x}', \mathbf{y})$ will be denoted by $Q_{X'Y}$. In general, depending on the context, Q_{XY} and $Q_{X'Y}$ (or just Q) may also denote generic joint distributions on $\mathcal{X} \times \mathcal{Y}$, not necessarily empirical distributions pertaining to sequences of finite length. For a given Q_{XY} , let us define

$$A(Q_{XY}, R, \beta) \triangleq \min_{Q_{X'Y}} \left\{ [R - H_Q(X'|Y)]_+ : \epsilon(Q_{X'Y}) + \frac{1}{\beta} [H_Q(X'|Y) - R]_+ \geq \epsilon(Q_{XY}) \right\}, \quad (59)$$

where $H_Q(X'|Y)$ is the conditional entropy of X' given Y associated with $Q_{X'Y}$. Finally, define

$$E(R, \beta) = \min_{Q_{XY}} [D(Q_{XY} \| P) + A(Q_{XY}, R, \beta)], \quad (60)$$

where $D(Q_{XY} \| P)$ is the relative entropy (Kullback–Leibler divergence) between $\{Q_{XY}(x, y)\}$ and $\{P(x, y)\}$, i.e.,

$$D(Q_{XY} \| P) = \sum_{x,y} Q_{XY}(x, y) \ln \frac{Q_{XY}(x, y)}{P(x, y)}. \quad (61)$$

The following theorem presents our main result in this section.

Theorem 1 *For the ensemble of random binning pertaining to SW codes, as described in 2.2,*

$$\mathbb{E}(R, \beta) = E(R, \beta). \quad (62)$$

It is easy to see that $E(R, \beta)$ is non-decreasing both in β and R . In particular,

$$E(R, \infty) = \min_{Q_{Y|X}} \left\{ D(Q_{XY} \| P) + \min_{Q_{X'|Y}: \epsilon(Q_{X'Y}) \geq \epsilon(Q_{XY})} [R - H_Q(X'|Y)]_+ \right\}, \quad (63)$$

which agrees with the error exponent of the word error probability, as expected. On the other hand, for $\beta = 1$, the finite-temperature decoder minimizes the BER and hence maximizes the exponent. Consequently, $E(R, \beta)$ must be a constant for all $\beta \geq 1$, which is equal to $E(R, \infty)$. It is also easy to verify that $E(R, \beta)$ vanishes for every $\beta \leq \Gamma^{-1}(R)$, i.e., beyond the paramagnetic-ferromagnetic boundary curve. Thus, $E(R, \beta)$ has three phases in the plane of β vs. R : (i) $\beta \leq \Gamma^{-1}(R)$ or $R \leq H(X|Y)$ (the union of the paramagnetic and glassy phases of the posterior), where $E(R, \beta) = 0$, (ii) $R > H(X|Y)$ and $\beta \geq 1$, where $E(R, \beta) = E(R, \infty)$ (first ferromagnetic sub-phase), and (iii) $R > H(X|Y)$ and $\Gamma^{-1}(R) \leq \beta < 1$ (second ferromagnetic sub-phase), where $E(R, \beta) < E(R, \infty)$ is monotonically non-decreasing both in β and R . The phase diagram of $E(R, \beta)$ is depicted in Fig. 3. As can be seen, it is related to the phase diagram of the finite-temperature partition function, but somewhat different. Here the paramagnetic and the glassy phases are united (in both of them $E(R, \beta) = 0$), but the ferromagnetic phase is subdivided into two new phases, as described above.

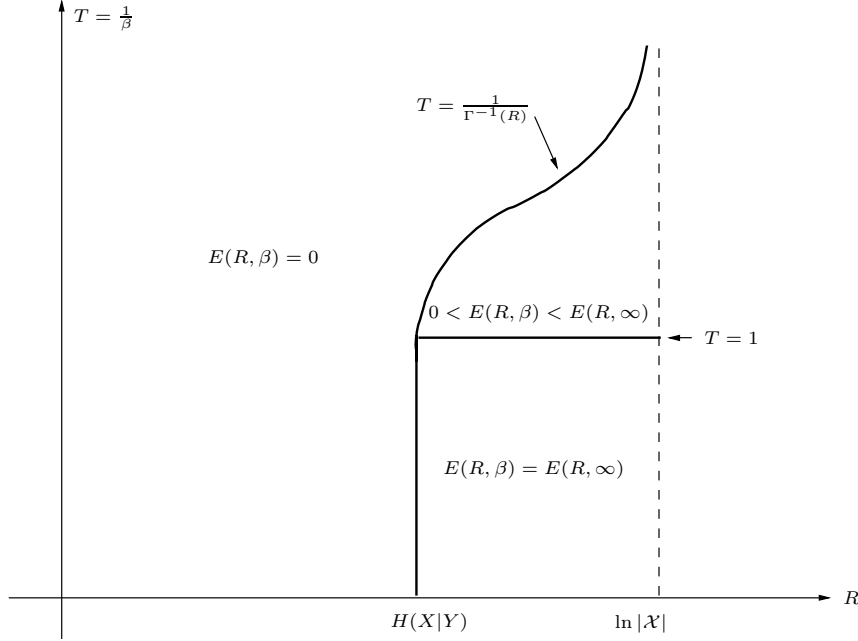


Figure 3: Phase diagram of $E(R, 1/T)$ in the plane of the decoding temperature T vs. the coding rate R .

Remark 2. In order to analyze the performance of other decoding metrics that depend on (\mathbf{x}, \mathbf{y})

only via their joint type, one should simply replace the definition of $\epsilon(Q_{XY})$ by the corresponding metric, for example, a mismatched metric $\epsilon(Q_{XY}) = \sum_{x,y} Q_{XY}(x,y) \ln \tilde{P}(x,y)$, or the minimum conditional entropy metric $\epsilon(Q_{XY}) = -H_Q(X|Y)$. Concerning the latter, the phase diagram of the error exponent will be based on Fig. 2 in the same way that the phase diagram of Fig. 3 is based on the phase diagram of Fig. 1. Here, however, there is no apparent subdivision of the ferromagnetic phase $E(R, \beta) > 0$ by the line $T = 1$. In other words, there are just two phases, $E(R, \beta) > 0$ and $E(R, \beta) = 0$.

Proof of Theorem 1. The proof is similar to the proof of Theorem 1 in [11]. For a given $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^N \times \mathcal{Y}^N$, and a given joint probability distribution $Q_{X'Y}$ on $\mathcal{X} \times \mathcal{Y}$, let $\Omega_1(Q_{X'Y})$ denote the number of $\{\mathbf{x}'\}$ within the bin of \mathbf{x} , such that $x'_1 \neq x_1$ and such that the empirical joint distribution with \mathbf{y} is given by $Q_{X'Y}$, that is

$$\Omega_1(Q_{X'Y}) = \sum_{\mathbf{x}': x'_1 = x_1} \mathcal{I}\{(\mathbf{x}', \mathbf{y}) \in \mathcal{I}(Q_{X'Y})\} \cdot \mathcal{I}[f(\mathbf{x}') = f(\mathbf{x})]. \quad (64)$$

For a given (\mathbf{x}, \mathbf{y}) , the BER is first calculated w.r.t. the randomness of the bins of codewords with $x'_1 \neq x_1$, but for a given binning of those with $x'_1 = x_1$. We henceforth denote $\mathcal{C}_0 = \{\mathbf{x}' : \mathbf{x}'_1 = x_1, f(\mathbf{x}') = f(\mathbf{x})\}$ and $\mathcal{C}_1 = \{\mathbf{x}' : \mathbf{x}'_1 \neq x_1, f(\mathbf{x}') = f(\mathbf{x})\}$.

For a given \mathcal{C}_0 and (\mathbf{x}, \mathbf{y}) , let

$$r \triangleq \frac{1}{N} \ln \left[\sum_{\mathbf{x}' \in \mathcal{C}_0} P^\beta(\mathbf{x}', \mathbf{y}) \right], \quad (65)$$

and so, the BER becomes

$$\Pr \left\{ \sum_{\mathbf{x}' \in \mathcal{C}_1} P^\beta(\mathbf{x}', \mathbf{y}) \geq e^{Nr} \right\},$$

where it is kept in mind that r is a function of \mathcal{C}_0 and (\mathbf{x}, \mathbf{y}) . Now,

$$\Pr \left\{ \sum_{\mathbf{x}' \in \mathcal{C}_1} P^\beta(\mathbf{x}', \mathbf{y}) \geq e^{Nr} \right\} = \Pr \left\{ \sum_{Q_{X'Y}} \Omega_1(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \geq e^{Nr} \right\} \quad (66)$$

$$\doteq \Pr \left\{ \max_{Q_{X'Y}} \Omega_1(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \geq e^{Nr} \right\} \quad (67)$$

$$= \Pr \bigcup_{Q_{X'Y}} \left\{ \Omega_1(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \geq e^{Nr} \right\} \quad (68)$$

$$\doteq \sum_{Q_{X'|Y}} \Pr \left\{ \Omega_1(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \geq e^{Nr} \right\} \quad (69)$$

$$\doteq \max_{Q_{X'|Y}} \Pr \left\{ \Omega_1(Q_{X'Y}) \geq e^{N[r-\beta\epsilon(Q_{X'Y})]} \right\}, \quad (70)$$

Now, for a given $Q_{X'|Y}$, $\Omega_1(Q_{X'Y})$ is a binomial random variable with exponentially $e^{NH_Q(X'|Y)}$ trials and probability of ‘success’ e^{-NR} . Thus, similarly as in [11] and [17], a standard large deviations analysis yields

$$\Pr \left\{ \Omega_1(Q_{X'Y}) \geq e^{N[r-\beta\epsilon(Q_{X'Y})]} \right\} \doteq e^{-NE_0(r,\beta,R,Q_{X'Y})}, \quad (71)$$

where

$$\begin{aligned} E_0(r,\beta,R,Q_{X'Y}) &= \begin{cases} [R - H_Q(X'|Y)]_+ & \beta\epsilon(Q_{X'Y}) \geq s \\ 0 & \beta\epsilon(Q_{X'Y}) < r, \beta\epsilon(Q_{X'Y}) \geq r - H_Q(X'|Y) + R \\ \infty & \beta\epsilon(Q_{X'Y}) < r, \beta\epsilon(Q_{X'Y}) < r - H_Q(X'|Y) + R \end{cases} \\ &= \begin{cases} R - H_Q(X'|Y) & \beta\epsilon(Q_{X'Y}) \geq r, H_Q(X'|Y) \leq R \\ 0 & \beta\epsilon(Q_{X'Y}) \geq r, H_Q(X'|Y) \geq R \\ 0 & \beta\epsilon(Q_{X'Y}) < r, \beta\epsilon(Q_{X'Y}) \geq s - H_Q(X'|Y) + R \\ \infty & \beta\epsilon(Q_{X'Y}) < r, \beta\epsilon(Q_{X'Y}) < r - H_Q(X'|Y) + R \end{cases} \\ &= \begin{cases} R - H_Q(X'|Y) & \beta\epsilon(Q_{X'Y}) \geq r, H_Q(X'|Y) \leq R \\ 0 & \beta\epsilon(Q_{X'Y}) \geq r - [H_Q(X'|Y) - R]_+, H_Q(X'|Y) \geq R \\ \infty & \beta\epsilon(Q_{X'Y}) < r - [H_Q(X'|Y) - R]_+ \end{cases} \\ &= \begin{cases} [R - H_Q(X'|Y)]_+ & \beta\epsilon(Q_{X'Y}) \geq r - [H_Q(X'|Y) - R]_+ \\ \infty & \beta\epsilon(Q_{X'Y}) < r - [H_Q(X'|Y) - R]_+ \end{cases} \end{aligned} \quad (72)$$

Therefore, $\max_{Q_{X'|Y}} \Pr\{\Omega_1(Q_{X'Y}) \geq e^{N[r-\beta\epsilon(Q_{X'Y})]}\}$ decays according to

$$E_1(r,\beta,R,Q_Y) = \min_{Q_{X'|Y}} E_0(r,\beta,R,Q_{X'Y}),$$

which is given by

$$E_1(r,\beta,R,Q_Y) = \min\{[R - H_Q(X'|Y)]_+ : \beta\epsilon(Q_{X'Y}) + [H_Q(X'|Y) - R]_+ \geq r\} \quad (73)$$

with the understanding that the minimum over an empty set is defined as infinity. Finally, $P_b(R,\beta,N)$ is the expectation of $e^{-NE_1(r,\beta,R,Q_Y)}$ where the expectation is w.r.t. the randomness of the binning in \mathcal{C}_0 and the randomness of (\mathbf{X}, \mathbf{Y}) . This expectation will be taken in two steps: first, over the randomness of the binning in \mathcal{C}_0 while \mathbf{x} (the real transmitted source vector) and \mathbf{y} are held fixed, and then over the randomness of \mathbf{X} and \mathbf{Y} . Let \mathbf{x} and \mathbf{y} be given and let $\delta > 0$ be arbitrarily small. Then,

$$P_b(\mathbf{x}, \mathbf{y}) \stackrel{\Delta}{=} \mathbf{E}[\exp\{-NE_1(r,\beta,R,Q_Y)\} | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}]$$

$$\begin{aligned}
&= \sum_r P(r|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) \cdot \exp\{-NE_1(r, \beta, R, Q_Y)\} \\
&\leq \sum_i \Pr\{i\delta \leq r < (i+1)\delta | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}\} \cdot \exp\{-NE_1(i\delta, \beta, R, Q_Y)\}, \quad (74)
\end{aligned}$$

where i ranges from $\frac{1}{N\delta} \ln P(\mathbf{x}, \mathbf{y})$ to some constant, which is immaterial for our purposes. Now,

$$\begin{aligned}
e^{nr} &= P^\beta(\mathbf{x}, \mathbf{y}) + \sum_{\mathbf{x}': x'_1=x_1} P^\beta(\mathbf{x}', \mathbf{y}) \cdot \mathcal{I}[f(\mathbf{x}') = f(\mathbf{x})] \\
&= e^{N\beta\epsilon(Q_{XY})} + \sum_{Q_{X'|Y}} \Omega_0(Q_{X'|Y}) e^{N\beta\epsilon(Q_{X'|Y})}, \quad (75)
\end{aligned}$$

where Q_{XY} is the empirical distribution of (\mathbf{x}, \mathbf{y}) and $\Omega_0(Q_{X'|Y})$ is the number of codewords in $\mathcal{C}_0 \setminus \{\mathbf{x}\}$ whose joint empirical distribution with \mathbf{y} is $Q_{X'|Y}$. The first term in the second line of (75) is fixed at this stage. As for the second term, we have (similarly as before):

$$\Pr \left\{ \sum_{Q_{X'|Y}} \Omega_0(Q_{X'|Y}) e^{N\beta\epsilon(Q_{X'|Y})} \geq e^{nt} \right\} \doteq e^{-NE_1(t, \beta, R, Q_Y)}. \quad (76)$$

On the other hand,

$$\Pr \left\{ \sum_{Q_{X'|Y}} \Omega_0(Q_{X'|Y}) e^{N\beta\epsilon(Q_{X'|Y})} \leq e^{Nt} \right\} \doteq \Pr \bigcap_{Q_{X'|Y}} \left\{ \Omega_0(Q_{X'|Y}) \leq e^{N[t - \beta\epsilon(Q_{X'|Y})]} \right\}. \quad (77)$$

Now, if there exists at least one $Q_{X'|Y}$ for which $R < H_Q(X'|Y)$ and $H_Q(X'|Y) - R > t - \beta\epsilon(Q_{X'|Y})$, then this $Q_{X'|Y}$ alone is responsible for a double exponential decay of the probability of the event $\{\Omega_0(Q_{X'|Y}) \leq e^{N[t - \beta\epsilon(Q_{X'|Y})]}\}$, let alone the intersection over all $Q_{X'|Y}$. On the other hand, if for every $Q_{X'|Y}$, either $R \geq H_Q(X'|Y)$ or $H_Q(X'|Y) - R \leq t - \beta\epsilon(Q_{X'|Y})$, then we have an intersection of polynomially many events whose probabilities all tend to unity. Thus, the probability in question behaves exponentially like an indicator function of the condition that for every $Q_{X'|Y}$, either $R \geq H_Q(X'|Y)$ or $H_Q(X'|Y) - R \leq t - \beta\epsilon(Q_{X'|Y})$, or equivalently,

$$\Pr \left\{ \sum_{Q_{X'|Y}} \Omega_0(Q_{X'|Y}) e^{N\beta\epsilon(Q_{X'|Y})} \leq e^{Nt} \right\} \doteq \mathcal{I} \left\{ R \geq \max_{Q_{X'|Y}} \{H_Q(X'|Y) - [t - \beta\epsilon(Q_{X'|Y})]_+\} \right\}. \quad (78)$$

Let us now find what is the minimum value of t for which the value of this indicator function is unity. The condition is equivalent to

$$\max_{Q_{X'|Y}} \min_{0 \leq a \leq 1} \{H_Q(X'|Y) - a[t - \beta\epsilon(Q_{X'|Y})]\} \leq R \quad (79)$$

or:

$$\forall Q_{X'|Y} \exists 0 \leq a \leq 1 : H_Q(X'|Y) - a[t - \beta\epsilon(Q_{X'Y})] \leq R, \quad (80)$$

which can also be written as

$$\forall Q_{X'|Y} \exists 0 \leq a \leq 1 : t \geq \beta\epsilon(Q_{X'Y}) + \frac{H_Q(X'|Y) - R}{a} \quad (81)$$

or equivalently,

$$t \geq \max_{Q_{X'|Y}} \min_{0 \leq a \leq 1} \left[\beta\epsilon(Q_{X'Y}) + \frac{H_Q(X'|Y) - R}{a} \right] \quad (82)$$

$$= \max_{Q_{X'|Y}} \left[\beta\epsilon(Q_{X'Y}) + \begin{cases} H_Q(X'|Y) - R & H_Q(X'|Y) - R \geq 0 \\ -\infty & H_Q(X'|Y) < R \end{cases} \right] \quad (83)$$

$$= \max_{\{Q_{X'|Y} : R \leq H_Q(X'|Y)\}} [\beta\epsilon(Q_{X'Y}) + H_Q(X'|Y)] - R \quad (84)$$

$$\triangleq r_0(Q_Y). \quad (85)$$

It is easy to check that $E_1(t, \beta, R, Q_Y)$ vanishes for $t \leq r_0(Q_Y)$. Thus, in summary, we have

$$\Pr \left\{ e^{nt} \leq \sum_{Q_{X'|Y}} \Omega_0(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \leq e^{n(t+\epsilon)} \right\} \doteq \begin{cases} 0 & t < r_0(Q_Y) - \epsilon \\ e^{-nE_1(t, \beta, R, Q_Y)} & t \geq r_0(Q_Y) \end{cases} \quad (86)$$

Therefore, for a given (\mathbf{x}, \mathbf{y}) , the expected error probability w.r.t. the randomness of the binning at \mathcal{C}_0 yields

$$P_e(\mathbf{x}, \mathbf{y}) = \mathbf{E}\{e^{-N[E_1(r, \beta, R, Q_Y)]} | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}\} \quad (87)$$

$$\leq \sum_i \Pr \left\{ e^{Ni\delta} \leq \sum_{Q_{X'|Y}} \Omega_0(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \leq e^{N(i+1)\delta} \right\} \times \exp\{-NE_1(\max\{i\delta, \beta\epsilon(Q_{X'Y})\}, \beta, R, Q_Y)\} \quad (88)$$

$$\leq \sum_{i \geq r_0(Q_Y)/\delta} \exp\{-NE_1(i\delta, \beta, R, Q_Y)\} \times \exp\{-NE_1(\max\{i\delta, \beta\epsilon(Q_{X'Y})\}, \beta, R, Q_Y)\}, \quad (89)$$

where the expression $\max\{i\delta, \beta\epsilon(Q_{X'Y})\}$ in the argument of $E_1(\cdot, Q_Y)$ is due to the fact that

$$r = \frac{1}{N} \ln \left[e^{N\beta\epsilon(Q_{X'Y})} + \sum_{Q_{X'|Y}} \Omega_0(Q_{X'Y}) e^{N\beta\epsilon(Q_{X'Y})} \right] \quad (90)$$

$$\geq \frac{1}{N} \ln \left[e^{N\beta\epsilon(Q_{X'Y})} + e^{Ni\delta} \right] \quad (91)$$

$$\doteq \max\{i\delta, \beta\epsilon(Q_{XY})\}. \quad (92)$$

By using the fact that δ is arbitrarily small, we obtain

$$\begin{aligned} P_e(\mathbf{x}, \mathbf{y}) &\doteq \exp\{-NE_1(\max\{r_0(Q_Y), \beta\epsilon(Q_{XY})\}, \beta, R, Q_Y)\} \\ &= \exp\{-N \max\{E_1(r_0(Q_Y), \beta, R, Q_Y), E_1(\beta\epsilon(Q_{XY}), \beta, R, Q_Y)\} \\ &= \exp\{-NE_1(\beta\epsilon(Q_{XY}), \beta, R, Q_Y)\} \end{aligned} \quad (93)$$

since the dominant contribution to the sum over i is due to the term $i = r_0(Q_Y)/\delta$ (by the non-decreasing monotonicity of the function $E_1(\cdot, Q_Y)$). After averaging w.r.t. (\mathbf{X}, \mathbf{Y}) , we obtain

$$E(R, \beta) = \min_{Q_{XY}}\{D(Q_{XY}\|P) + E_1(\beta\epsilon(Q_{XY}), \beta, R, Q_Y)\} \quad (94)$$

$$= \min_{Q_{XY}}\{D(Q_{XY}\|P) + A(Q_{XY}, R, \beta)\}, \quad (95)$$

completing the proof of Theorem 1.

References

- [1] I. Csiszár, “Linear codes for sources and source networks: error exponents, universal coding,” *IEEE Trans. Inform. Theory*, vol. IT-28, no. 4, pp. 585–592, July 1982.
- [2] I. Csiszár and J. Körner, “Towards a general theory of source networks,” *IEEE Trans. Inform. Theory*, vol. IT-26, no. 2, pp. 155–165, March 1980.
- [3] I. Csiszár and J. Körner, “Graph decomposition: a new key to coding theorems,” *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 5–12, January 1981.
- [4] I. Csiszár, J. Körner, and K. Marton, “A new look at the error exponent of a discrete memoryless channel,” *Proc. ISIT ‘77*, p. 107 (abstract), Cornell University, Ithaca, New York, U.S.A., 1977.
- [5] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley 1968.
- [6] R. G. Gallager, “Source coding with side information and universal coding,” LIDS-P-937, M.I.T., 1976. Available on-line at: [<http://web.mit.edu/gallager/www/papers/paper5.pdf>].

- [7] B. G. Kelly and A. B. Wagner, “Improved source coding exponents via Witsenhausen’s rate,” *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 5615–5633, September 2011.
- [8] N. Merhav, “Statistical physics and information theory,” *Foundations and Trends in Communications and Information Theory*, vol. 6, nos. 1–2, pp. 1–212, 2009.
- [9] N. Merhav, “Relations between random coding exponents and the statistical physics of random codes,” *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 83–92, January 2009.
- [10] N. Merhav, “Erasure/list exponents for Slepian–Wolf decoding,” *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4463–4471, August 2014.
- [11] N. Merhav, “Exact random coding exponents of optimal bin index decoding,” *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 6024–6031, October 2014.
- [12] N. Merhav, “Erasure/list exponents for Slepian–Wolf decoding,” *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4463–4471, August 2014.
- [13] Y. Oohama and T. S. Han, “Universal coding for the Slepian–Wolf data compression system and the strong converse theorem,” *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.
- [14] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford University Press, 2009.
- [15] P. Ruján, “Finite temperature error–correcting codes,” *Phys. Rev. Let.*, vol. 70, no. 19, pp. 2968–2971, May 1993.
- [16] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT–19, no. 4, pp. 471–480, January 1973.
- [17] N. Weinberger and N. Merhav, “Codeword or noise? Exact random coding exponents for joint detection and decoding,” *IEEE Trans. Inform. Theory*, vol. 60, no. 9, pp. 5077–5094, September 2014.

- [18] N. Weinberger and N. Merhav, “Optimum trade-off between the error exponent and the excess-rate exponent of variable-rate Slepian–Wolf coding,” submitted to *IEEE Trans. Inform. Theory*, January 2014. Available on-line at: <http://arxiv.org/pdf/1401.0892.pdf>