A Statistical-Physics Refinement of Soft Covering

Neri Merhav

The Viterbi Faculty of Electrical and Computer Engineering,
Technion—Israel Institute of Technology, Technion City,
Haifa 3200003, Israel
merhav@ee.technion.ac.il

Abstract

We study the channel output distribution induced by a random code of rate R, from the perspective of statistical physics. The central object is the partition function $Z_n(\beta|\mathcal{C}) = \sum_{y^n}[P_{Y^n|\mathcal{C}}(y^n)]^\beta$, where y^n is the channel output vector, $\mathcal{C}$ is the code, and $\beta > 0$ plays the role of inverse temperature. More precisely, our focus is on the associated *annealed free energy*, $\psi(\beta, R) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[Z_n(\beta|\mathcal{C})]$, where the expectation is with respect to the randomness of $\mathcal{C}$. This quantity encodes the full Rényi spectrum of the output distribution. The single-letter formula derived for the annealed free energy decomposes into two branches which reflect a "competition" between two populations of codewords. One is the *bulk branch*, $\psi_{\rm b}(\beta, R)$, which is driven by typical codewords and the other one is the *sparse branch* $\psi_{\rm s}(\beta, R)$, which is driven by a-typical codewords, where the qualifiers 'typical' and 'atypical' are in a sense to become apparent later. We analyze the phase structure of each branch separately and characterize their competition. Both branches are derived for all $\beta > 0$. The phase boundary $R^\star(\beta)$, where the two branches are equal, is analyzed for $\beta \geq 1$, where it has an explicit closed-form expression. The phase diagram in the first quadrant of the (β, R) plane has four regions separated by three boundaries: $R = I^{\rm b}(\beta)$ (bulk branch transition), $R = R^\star(\beta)$ (bulk–sparse competition boundary), and $R = I^{\rm s}(\beta)$ (sparse branch transition), all meeting at the point $(\beta, R) = (1, I(X;Y))$, where $I(X;Y)$ is the mutual information induced by the input type and the channel. Applications to guesswork, channel resolvability, and hypothesis testing are discussed, and all results are illustrated with a numerical example of a Z-channel.

Keywords: random coding; soft covering; channel resolvability; free energy; phase transitions; annealed free energy; Rényi entropy; statistical mechanics; guesswork; hypothesis testing

Contents

1 Introduction

Given a codebook $\mathcal{C} = \{x^n(m)\}_{m=1}^M$ of size $M = e^{nR}$ and a discrete memoryless channel (DMC), $\{W(y|x)\}$, the induced channel output distribution is given by the mixture

$$P_{Y^n}(y^n) = \frac{1}{M}\sum_{m=1}^{M} W^n(y^n|x^n(m)). \tag{1}$$

This object is central to channel resolvability [1] and the soft-covering problem (see, e.g., [2], [3], [4] and references therein). Classical results, in this context, focus on a single threshold: if $R > I(X;Y)$, $I(X;Y)$ being the mutual information induced by the input distribution and the channel W, then the distribution P_{Y^n} converges to the i.i.d. product law $P_Y^{\otimes n}$ in total variation as well as in some other metrics between probability distributions.

This description, however, captures average behavior only. It says nothing about the internal geometry of P_{Y^n}: how the probability mass is distributed across output sequences, how many codewords support a typical and an atypical output, or how the self-information, $-\frac{1}{n}\log P_{Y^n}(y^n)$, fluctuates. These questions require going along a deeper journey, beyond the Shannon entropy and mutual information.

In this work, we study P_{Y^n} through the function,

$$Z_n(\beta) = \sum_{y^n}[P_{Y^n}(y^n)]^\beta = \sum_{y^n}\exp\{-\beta\log[1/P_{Y^n}(y^n)]\}, \quad \beta > 0, \tag{2}$$

where the second representation is readily recognized as a *partition function*, with $\beta > 0$ playing the role of *inverse temperature*, and $\log[1/P_{Y^n}(y^n)] = -\log P_{Y^n}(y^n)$ being the *energy function* (a.k.a. the *Hamiltonian*) associated with every *micro-state* y^n. In other words, this is identified as the canonical partition function of a statistical-mechanical system whose micro-states are channel output sequences. Two special values bracket the range: $Z_n(1) = 1$ (normalization, $\beta = 1$) and $\lim_{\beta\to\infty}[Z_n(\beta)]^{1/\beta} = \max_{y^n} P_{Y^n}(y^n)$ (the mode, or the ground state in the physics jargon). In general, $\frac{1}{1-\beta}\log Z_n(\beta)$ is exactly the definition of $H_\beta(P_{Y^n})$, the Rényi entropy of order $\beta \neq 1$, pertaining to the output distribution P_{Y^n}, which is non-increasing in β. The associated *free energy* $\psi(\beta, R)$ encodes the exponential growth/decay rate of $Z_n(\beta)$ as a function of n. Since $Z_n(\beta)$ depends on the random codebook, and actually, should be denoted $Z_n(\beta|\mathcal{C})$, we study the free energy behavior for the average code, which is identified with the *annealed free energy*,

$$\psi(\beta, R) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\{Z_n(\beta|\mathcal{C})\}, \tag{3}$$

provided that the limit exists. Here, $\mathbb{E}\{\cdot\}$ denotes the expectation operator with respect to (w.r.t.) the randomness of the code $\mathcal{C}$. Its phase structure — the rich dependence on β and R — is the main subject of this work.

It should be emphasized that many earlier statistical-mechanical analyses of random coding, were based on Derrida's random energy model (REM) (see, e.g., [5], [6] and references therein). These are based on partition functions whose micro-states were the channel inputs (i.e., the codewords) for a fixed channel output sequence y^n, that is,

$$Z_n^{\text{input}}(\beta) = \sum_{m=1}^{M}[W^n(y^n|x^n(m))]^\beta, \tag{4}$$

for fixed y^n, whereas, here the micro-states are the channel outputs, as mentioned earlier. This difference is significant because here, each energy term, $-\log P_{Y^n}(y^n)$, is itself a log-partition function over codewords. This two-level structure makes the problem considerably harder and richer than that of an ordinary REM.

The main contributions of this work are as follows. We derive the annealed free energy and analyze the phase structure of its two branches (Theorems 3.1 and 3.9). The free energy decomposes into a *bulk branch* and a *sparse branch*, reflecting a competition between two populations of codewords. The bulk branch is driven by *typical* codewords and the sparse branch is driven by *atypical* codewords, where the meanings of the qualifiers 'typical' and 'atypical' will become apparent in the sequel. Both branches are derived for all $\beta > 0$ and analyzed separately. The bulk branch, $\psi_{\mathrm{b}}(\beta, R)$ (Section 3), has a single phase transition at a critical rate $R = I^{\mathrm{b}}(\beta)$—the mutual information of the bulk optimizer (the unconstrained optimizer of the bulk branch)—where a rate constraint changes from inactive to active. A similar comment applies to the sparse branch, $\psi_{\mathrm{s}}(\beta, R)$, whose phase boundary is $R = I^{\mathrm{s}}(\beta)$. The total annealed free energy $\psi(\beta, R)$ has a phase boundary $R^\star(\beta)$ given by an explicit closed-form formula (Theorem 3.9(b), Section 3). Together, the three boundaries divide the (β, R) quadrant $\{R \geq 0,\ \beta \geq 1\}$ into four regions, denoted, A, B, C, and D. By construction, $\psi(\beta, R) \geq \psi_{\mathrm{b}}(\beta, R)$ always, with equality in region A and strict inequality in regions B, C, and D, where the sparse branch dominates. For the sparse branch, a fully explicit closed form formula, in terms of channel transition probabilities and β alone, holds whenever the rate is below $I^{\mathrm{s}}(\beta)$ (see eq. (29)). All results are illustrated in a numerical example in Section 4.

A few words are in order with regard to earlier related work and the differences relative to the present work. The model (1) is the canonical object of channel resolvability and the soft-covering problem. Han and Verdú [1] established the resolvability threshold $R = I(X;Y)$ under total variation, $\|P_{Y^n} - P_Y^{\otimes n}\|_{\mathrm{TV}} \to 0$ for $R > I(X;Y)$, and showed the threshold is tight. Hayashi [3] derived exponential convergence rates under total variation and normalized relative entropy, a.k.a. the Kullback-Leibler (KL) divergence. Hou and Kramer [4] extended the analysis to the Rényi divergence $D_\alpha(P_{Y^n} \| P_Y^{\otimes n})$, showing the critical rate remains $I(X;Y)$ for all $\alpha \in (0, \infty)$ but the exponents differ. Yu and Tan [7] derived exact error and strong-converse exponents for the soft-covering problem under KL divergence and total variation. In a slightly earlier but closely related and highly relevant work [8], the same authors characterized the *Rényi resolvability*—the minimum rate R required for the Rényi divergence $D_\alpha(P_{Y^n|\mathcal{C}} \| P_Y^{\otimes n})$ to vanish asymptotically. They showed that for $\alpha \leq 1$ the threshold remains $I(X;Y)$, while for $\alpha > 1$ it is strictly larger than $I(X;Y)$ and depends on α.

Our work is complementary to [8]: rather than characterizing the *threshold rate* at which a divergence to $P_Y^{\otimes n}$ vanishes, we study the exact value of the annealed free energy at every rate R and inverse temperature β, revealing a two-branch phase diagram with three distinct phase boundaries. We analyze the *below-threshold* regime with the same precision as the above-threshold regime, and expose qualitative phenomena—condensation, sparse-branch dominance, and the annealed/quenched gap—invisible to divergence-based analyses. In particular, our bulk branch boundary, $R = I^{\mathrm{b}}(\beta)$, for $\beta > 1$ is precisely the Rényi resolvability threshold of [8], now seen as one of three phase boundaries in a richer thermodynamic landscape.

In general, all earlier results, in this context, share a common feature: they measure how close P_{Y^n} is to a *target* distribution $P_Y^{\otimes n}$ via some metric, and they all identify $R = I(X;Y)$ as the single relevant threshold. The present paper asks a fundamentally

different and complementary question: not how close P_{Y^n} is to a target, but what is the *internal geometry* of P_{Y^n} itself?

Our partition function $Z_n(\beta)$ encodes the Rényi structure of P_{Y^n} without reference to any external target distribution, and reveals phenomena invisible to total variation or KL divergence. It turns out that the classical threshold, $R = I(X;Y)$, is only the *beginning* of the story: the (β, R) quadrant $\{\beta \geq 1,\ R \geq 0\}$ is divided into *four* distinct regions by three phase boundaries, all passing through the point $(\beta, R) = (1, I(X;Y))$. Concretely:

- Even for $R > I(X;Y)$ (where $\|P_{Y^n} - P_Y^{\otimes n}\|_{\mathrm{TV}} \to 0$), the annealed free energy reveals a *condensed phase*: when $R < I^b(\beta)$ for some $\beta > 1$, the bulk branch of ψ has an active rate constraint, signalling that probability mass concentrates on outputs supported by sub-exponentially few codewords — output condensation invisible to total variation distance.

- The annealed free energy has a phase boundary $R = R^\star(\beta)$ where the sparse branch takes over from the bulk branch, signalling that the atypical codewords dominate the ensemble average.

- Since $\log Z_n(\beta|\mathcal{C}) = (1-\beta)\, H_\beta(P_{Y^n|\mathcal{C}})$ exactly for every fixed codebook (by the definition of Rényi entropy), the bulk branch $\psi_{\mathrm{b}}(\beta, R)$ encodes the Rényi entropy rate $H_\beta(P_{Y^n|\mathcal{C}})/n$ of the output mixture for a typical codebook (for $\beta \geq 1$ and $R > R^\star(\beta)$, this is proved in Theorem 3.5; see also Appendix). The full annealed free energy thus encodes the complete Rényi-order profile of the output distribution, of which classical soft-covering ($\beta \to 1$) is a single cross-section.

In summary, this work does not contradict soft-covering results—it refines and extends them by revealing the fine structure of P_{Y^n} that lies beneath the single threshold $R = I(X;Y)$. These structural findings have direct operational consequences, which we now describe.

1. *Guesswork and Rényi entropy.* Arikan [9] showed that guesswork moments $\mathbb{E}[G(Y^n)^s]$ grow as $e^{nsH_{1/(1+s)}(P_{Y^n})}$. The annealed free energy encodes the Rényi entropy rate of the output mixture, extending Arıkan's analysis to the case where the source is itself a random-coding mixture. The phase structure of $\psi(\beta, R)$ produces regime changes in guessing complexity absent in the i.i.d. case.

2. *Hypothesis testing.* The Chernoff exponent [10] for testing $P_{Y^n|\mathcal{C}}$ against $P_Y^{\otimes n}$ equals $\xi(R) = \max_{0 \leq \beta \leq 1}[(1-\beta)\log|\mathcal{Y}| - \psi(\beta, R)]$, directly connecting our free energy to the optimal test. With P_Y being uniform, $\xi(R) = 0$ for $R \geq I(X;Y)$, meaning the two distributions are not exponentially distinguishable at and above the soft-covering threshold. The phase structure of $\psi(\beta, R)$ produces qualitative regime changes in $\xi(R)$ as R varies, and connects to the Rényi resolvability results of Yu and Tan [8]; see Section 5.

3. *Statistical mechanics of codes.* Sourlas [11] established the connection between linear codes and spin-glass models. Montanari [12] analyzed the phase transition in turbo codes. Mézard and Montanari [5] (Chapters 5 and 6) and Merhav [6] provided a comprehensive treatment of random coding via statistical physics, with the partition function summing over codewords (inputs) for a fixed received sequence. The present work complements [6] by placing the partition function over outputs: the resulting two-level hierarchical model

has a distinct phase structure governed by the code rate R rather than by signal-to-noise ratio.

The outline of the remaining part of this article is as follows. Section 2 defines the model and establishes the notation conventions. Section 3 derives the annealed free energy and analyzes the phase structure of its bulk and sparse branches; the phase boundary $R^*(\beta)$ has an explicit closed-form formula for $\beta \geq 1$. Section 4 presents the phase diagram (four regions, $\beta \geq 1$). Section 5 discusses applications and implications of our results. Finally, Section 6 summarizes the paper and provides an outlook.

2 Model, Definitions and Notation

Throughout the paper, n denotes the block length. An n-vector over alphabet $\mathcal{X}$ is displayed as $x^n = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, $\mathcal{X}$ being the single-letter finite alphabet, and similarly $y^n \in \mathcal{Y}^n$, where the single-letter alphabet $\mathcal{Y}$ is also finite. The empirical distribution (type) of x^n is denoted $\hat{P}_{x^n}(a) = \frac{1}{n}\#\{i : \ x_i = a\}$ for all $a \in \mathcal{X}$. A similar definition applies to joint types of pairs of sequences, (x^n, y^n). A DMC W is a stochastic matrix $W : \mathcal{X} \to \mathcal{Y}$ and for sequences x^n, y^n, we write $W^n(y^n|x^n) = \prod_{i=1}^n W(y_i|x_i)$. Throughout the sequel, all logarithms are natural (base e); Accordingly, information measures are given in nats. We use the standard exponential-equivalence notation $f(n) \doteq g(n)$ to mean $\lim_{n\to\infty} \frac{1}{n} \log \frac{f(n)}{g(n)} = 0$, i.e. f and g have the same exponential growth/decay rate. The notation of information measures is as follows. $D(P\|Q)$ the Kullback-Leibler (KL) divergence,

$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{5}$$

Throughout the paper, Q_{XY} denotes a joint distribution on $\mathcal{X} \times \mathcal{Y}$ with X-marginal given by $Q_X = P_X$, P_X being a fixed input distribution on the finite input alphabet $\mathcal{X}$. When Q_{XY} needs to appear as a *subscript* in an information measure, we abbreviate Q_{XY} by Q to avoid cumbersome notation. Thus, $H_Q(Y)$, $H_Q(Y|X)$, and $I_Q(X;Y)$ stand for $H_{Q_{XY}}(Y)$, $H_{Q_{XY}}(Y|X)$, and $I_{Q_{XY}}(X;Y)$ respectively, which are the entropy of Y, the conditional entropy of Y given X, and the mutual information between X and Y, respectively, all induced by Q_{XY}. On the other hand, when Q_{XY} appears as an argument of a certain functional, the full notation is used. The notation $I(X;Y)$ (without subscript) refers to the mutual information induced by the fixed single-letter channel input distribution P_X and the channel transition probability matrix W, i.e. $I(X;Y) = I_{P_X W}(X;Y)$. The KL divergence between a conditional distribution $Q_{Y|X}$ and W, with weighting by P_X, is defined by

$$D(Q_{Y|X}\|W|P_X) = \sum_{x\in\mathcal{X}} P_X(x) \sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{W(y|x)}. \tag{6}$$

A random codebook of rate $R > 0$ and block length n is a collection $\mathcal{C} = \{x^n(m)\}_{m=1}^M$, $M = \lfloor e^{nR} \rfloor$, where the codewords $x^n(m)$ are drawn independently and uniformly at random from the type class

$$\mathcal{T}(P_X) = \big\{x^n \in \mathcal{X}^n : \hat{P}_{x^n}(a) = P_X(a) \ \forall\, a \in \mathcal{X}\big\}, \tag{7}$$

the set of all n-sequences with empirical distribution exactly P_X. Given codebook $\mathcal{C}$, the induced output distribution is

$$P_{Y^n|\mathcal{C}}(y^n) = \frac{1}{M}\sum_{m=1}^{M} W^n(y^n|x^n(m)). \tag{8}$$

For $\beta > 0$, the partition function is defined as

$$Z_n(\beta|\mathcal{C}) = \sum_{y^n}[P_{Y^n|\mathcal{C}}(y^n)]^\beta. \tag{9}$$

The *annealed free energy* is

$$\psi(\beta, R) = \lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}_{\mathcal{C}}\big[Z_n(\beta|\mathcal{C})\big], \tag{10}$$

assuming that the limit exists. For a joint type Q_{XY} on $\mathcal{X}\times\mathcal{Y}$ with marginal $Q_X = P_X$, define:

$$\ell(Q_{XY}) = H_Q(Y|X) + D(Q_{Y|X}\|W|P_X). \tag{11}$$

Note that for (x^n, y^n) of joint type Q_{XY}, $W^n(y^n|x^n) = e^{-n\ell(Q_{XY})}$. We shall often use the identity

$$I_Q(X;Y) + \ell(Q_{XY}) = H_Q(Y) + D(Q_{Y|X}\|W|P_X). \tag{12}$$

Fix an output sequence y^n of type Q_Y. The type-class enumerator $N(Q_{XY}|y^n)$ counts how many codewords have a joint type (joint empirical distribution) Q_{XY} with y^n:

$$N(Q_{XY}|y^n) = \#\big\{m:\ (x^n(m), y^n) \in \mathcal{T}(Q_{XY})\big\}, \tag{13}$$

where $\mathcal{T}(Q_{XY})$ is the joint type class associated with the joint distribution Q_{XY}.

Using the method of types [13], [14], it is readily observed that given y^n, each randomly selected codeword, $x^n(m)$, satisfies $(x^n(m), y^n) \in \mathcal{T}(Q_{XY})$ with probability

$$\frac{|\mathcal{T}(Q_{X|Y}|y^n)|}{|\mathcal{T}(P_X)|} \doteq \frac{e^{nH_Q(X|Y)}}{e^{nH_Q(X)}} = e^{-nI_Q(X;Y)}, \tag{14}$$

where $\mathcal{T}(Q_{X|Y}|y^n) = \{x^n:\ (x^n, y^n) \in \mathcal{T}(Q_{XY})\}$ is the conditional type class, and where we have used the fact that $Q_X = P_X$. Since codewords are drawn independently at random from $\mathcal{T}(P_X)$, it is clear that $N(Q_{XY}|y^n)$ is a binomial random variable with M trials and a probability of single success of the exponential order of $e^{-nI_Q(X;Y)}$.

3 Annealed Free Energy and its Phase Structure

In this section, we derive a single-letter expression for the annealed free energy and investigate its phase structure in the (β, R) plane. The first main result is the following.

Theorem 3.1 (Annealed Free Energy, $\beta > 0$)**.** *For all $\beta > 0$:*

$$\psi(\beta, R) = \max\big\{\psi_b(\beta, R),\ \psi_s(\beta, R)\big\}, \tag{15}$$

with

$$\psi_b(\beta,R)=\sup_{\substack{Q_{XY}:\,Q_X=P_X\\ I_Q(X;Y)\le R}}\big[H_Q(Y)-\beta(I_Q(X;Y)+\ell(Q_{XY})\big], \tag{16}$$

$$\psi_s(\beta,R)=R(1-\beta)+\sup_{\substack{Q_{XY}:\,Q_X=P_X\\ I_Q(X;Y)>R}}\{H_Q(Y|X)-\beta\ell(Q_{XY})\}. \tag{17}$$

For convenience, we define the functional

$$F(Q_{XY}):=H_Q(Y|X)-\beta\ell(Q_{XY}), \tag{18}$$

which is the common building block of both branches. Using the identity, $H_Q(Y)=I_Q(X;Y)+H_Q(Y|X)$, we may present the two branches of the annealed free energy as

$$\psi_{\rm b}(\beta,R)=\sup_{\substack{Q_{XY}:\,Q_X=P_X\\ I_Q(X;Y)\le R}}\big[(1-\beta)I_Q(X;Y)+F(Q_{XY})\big] \tag{19}$$

$$\psi_{\rm s}(\beta,R)=R(1-\beta)+\sup_{\substack{Q_{XY}:\,Q_X=P_X\\ I_Q(X;Y)>R}}F(Q_{XY}). \tag{20}$$

Proof of Theorem 3.1. We begin with the trivial identity,

$$\mathbb{E}[Z_n(\beta)]=\mathbb{E}\left[\sum_{y^n}[P_{Y^n|\mathcal{C}}(y^n)]^\beta\right]=\sum_{y^n}\mathbb{E}\left\{\big[P_{Y^n|\mathcal{C}}(y^n)\big]^\beta\right\}. \tag{21}$$

We next focus on $\mathbb{E}[P_{Y^n|\mathcal{C}}(y^n)^\beta]$ for a fixed y^n. Define $S(y^n)=\sum_{m=1}^M W^n(y^n|x^n(m))$, so that $P_{Y^n|\mathcal{C}}(y^n)=S(y^n)/M$. Since $M=e^{nR}$ is deterministic,

$$\mathbb{E}\big[P_{Y^n|\mathcal{C}}(y^n)^\beta\big]=M^{-\beta}\mathbb{E}\left\{[S(y^n)]^\beta\right\}=e^{-n\beta R}\mathbb{E}\left\{[S(y^n)]^\beta\right\}. \tag{22}$$

The problem reduces to computing $\mathbb{E}[S(y^n)^\beta]$. For each y^n,

$$\begin{aligned}
\mathbb{E}\{[S(y^n)]^\beta\} &= \mathbb{E}\left[\sum_{\substack{Q_{XY}:\,Q_X=P_X\\ Q_Y=\hat{P}_{y^n}}}N(Q_{XY}|y^n)\cdot e^{-n\ell(Q_{XY})}\right]^\beta\\
&\doteq \mathbb{E}\left[\sum_{\substack{Q_{XY}:\,Q_X=P_X\\ Q_Y=\hat{P}_{y^n}}}[N(Q_{XY}|y^n)]^\beta\cdot e^{-n\beta\ell(Q_{XY})}\right]\\
&= \sum_{\substack{Q_{XY}:\,Q_X=P_X\\ Q_Y=\hat{P}_{y^n}}}\mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\}\cdot e^{-n\beta\ell(Q_{XY})}\\
&\doteq \max_{\substack{Q_{XY}:\,Q_X=P_X\\ Q_Y=\hat{P}_{y^n}}}\mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\}\cdot e^{-n\beta\ell(Q_{XY})},
\end{aligned} \tag{23}$$

where the dotted equalities are since the sum contains at most poly(n) non-negative terms. Now, the evaluation of $\mathbb{E}\{N(Q_{XY}|y^n)]^\beta\}$ requires the following lemma, which is Theorem 4.2 of [15] and the proof is therein.

Lemma 3.2 (Moments of Binomial Enumerator)**.** *Let* $N \sim \text{Binomial}(e^{nA}, e^{-nB})$ *with* $A, B > 0$ *and* $\beta > 0$*. Then,*

$$\mathbb{E}\{N^\beta\} \doteq \begin{cases} e^{n\beta(A-B)} & A > B \\ e^{-n(B-A)} & A < B \end{cases} \tag{24}$$

The case $A = B$ corresponds to a non-exponential behavior of $\mathbb{E}\{N^\beta\}$ as can be seen by observing the limiting behavior of both cases. (As a side remark, when N is Binomial$(e^{nA}, \lambda e^{-nA})$, for some constant $\lambda > 0$, then in the limit of $n \to \infty$, N becomes a Poissonian random variable with parameter λ, that is, $\Pr\{N = k\} \to \frac{\lambda^k e^{-\lambda}}{k!}$, for every non-negative integer k, and therefore, the asymptotic βth moments are constants).

Lemma 3.2 is now used with the assignments $A = R$ and $B = I_Q(X;Y)$. Accordingly, one must distinguish between types $\{Q_{XY}\}$ for which $R > I_Q(X;Y)$ as opposed to those with $R < I_Q(X;Y)$. Now,

$$\begin{aligned}
\mathbb{E}\{[S(y^n)]^\beta\} &\doteq \max_{\substack{Q_{XY}:\, Q_X = P_X \\ Q_Y = \hat{P}_{y^n}}} \mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\} \cdot e^{-n\beta\ell(Q_{XY})} \\
&= \max\Bigg\{ \max_{\substack{Q_{XY}:\, Q_X = P_X,\ Q_Y = \hat{P}_{y^n} \\ I_Q(X;Y) \le R}} \mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\} \cdot e^{-n\beta\ell(Q_{XY})}, \\
&\qquad \max_{\substack{Q_{XY}:\, Q_X = P_X,\ Q_Y = \hat{P}_{y^n} \\ I_Q(X;Y) \ge R}} \mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\} \cdot e^{-n\beta\ell(Q_{XY})} \Bigg\} \\
&\doteq \max\Bigg\{ \max_{\substack{Q_{XY}:\, Q_X = P_X,\ Q_Y = \hat{P}_{y^n} \\ I_Q(X;Y) \le R}} e^{n\beta[R - I_Q(X;Y)]} \cdot e^{-n\beta\ell(Q_{XY})}, \\
&\qquad \max_{\substack{Q_{XY}:\, Q_X = P_X,\ Q_Y = \hat{P}_{y^n} \\ I_Q(X;Y) \ge R}} e^{-n[I_Q(X;Y) - R]} \cdot e^{-n\beta\ell(Q_{XY})} \Bigg\}.
\end{aligned} \tag{25}$$

Since the latter expression involves maximization over $\{Q_{XY}\}$ when $Q_Y = \hat{P}_{y^n}$ is held fixed, it is clear that it depends on y^n only via $\hat{P}_{y^n}$. The number of $\{y^n\}$ with $\hat{P}_{y^n} = Q_Y$ is of the exponential order of $e^{nH_Q(Y)}$. Multiplying the per-y^n contributions by $e^{nH_Q(Y)}$ and by $e^{-n\beta R}$, yields the total contribution of type Q_Y, and finally, summing over all $\{Q_Y\}$, which is exponentially equivalent to maximizing over all $\{Q_Y\}$, yields the asserted expression of $\psi(\beta, R)$. The part pertaining to types with $I_Q(X;Y) \le R$ is identified with $\psi_{\text{b}}(\beta, R)$, and the one with $I_Q(X;Y) > R$ is associated with $\psi_{\text{s}}(\beta, R)$. □

For future reference, we need the following definition.

Definition 3.3 (Sparse-feasible and bulk-feasible types)**.** *For a given rate* $R > 0$*, a joint distribution* Q_{XY} *with* $Q_X = P_X$ *is called:*

- sparse-feasible *if* $I_Q(X;Y) > R$*, i.e., it satisfies the constraint of the sparse branch optimization;*
- bulk-feasible *if* $I_Q(X;Y) \le R$*, i.e., it satisfies the constraint of the bulk branch optimization.*

Every Q_{XY} *is either sparse-feasible or bulk-feasible (or both, if* $I_Q(X;Y) = R$*).*

3.1 The Sparse Branch: Formula and Phase Structure

The sparse branch of the annealed free energy,

$$\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + \sup_{\substack{Q_{XY}: Q_X = P_X \\ I_Q(X;Y) > R}} F(Q_{XY}), \tag{26}$$

captures the contribution of *sparse-type* codewords: those whose joint type Q_{XY} with the output y^n satisfies $I_Q(X;Y) > R$. For fixed y^n, such codewords are rarely encountered in a typical codebook of the ensemble.

Definition 3.4 (The annealed optimizer Q_β^{s} and its mutual information $I^{\mathrm{s}}(\beta)$)**.** *For each $\beta > 0$, define*

$$Q_\beta^{\mathrm{s}}(y|x) = \frac{[W(y|x)]^\beta}{\sum_{y' \in \mathcal{Y}} [W(y'|x)]^\beta} \quad x \in \mathcal{X}. \tag{27}$$

It can be readily shown that Q_β^{s} maximizes $F(Q_{XY}) = H_Q(Y|X) - \beta \ell(Q_{XY})$ over all Q_{XY} with $Q_X = P_X$. Accordingly, define

$$I^{\mathrm{s}}(\beta) := I_{Q_\beta^{\mathrm{s}}}(X;Y). \tag{28}$$

The curve $R = I^{\mathrm{s}}(\beta)$ marks the point where Q_β^{s} crosses from sparse-feasible ($I_{Q_\beta^{\mathrm{s}}}(X;Y) > R$, and so, Q_β^{s} optimizes the sparse branch) to bulk-feasible ($I_{Q_\beta^{\mathrm{s}}}(X;Y) \le R$, so Q_β^{s} optimizes the bulk branch).

Closed-form formula. By Lemma 3.2, when $I_Q(X;Y) > R$ the βth moment of the enumerator satisfies $\mathbb{E}\{[N(Q_{XY}|y^n)]^\beta\} \doteq e^{n[R - I_Q(X;Y)]}$ regardless of β. The resulting annealed calculation gives

$$\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta) \quad \text{whenever } R < I^{\mathrm{s}}(\beta), \tag{29}$$

where

$$C(\beta) = \sum_{x \in \mathcal{X}} P_X(x) \log \left(\sum_{y \in \mathcal{Y}} [W(y|x)]^\beta \right). \tag{30}$$

The formula ceases to hold when $R > I^{\mathrm{s}}(\beta)$ as Q_β^{s} is then bulk-feasible ($I_{Q_\beta^{\mathrm{s}}}(X;Y) \le R$) and hence does not comply with the rate constraint of the sparse branch.

Phase structure. The sparse branch has its own phase transition at $R = I^s(\beta)$:

- For $R < I^{\mathrm{s}}(\beta)$: $\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta)$, the closed-form linear formula. The sparse branch is maximized by Q_β^s.
- For $R > I^s(\beta)$: Q_β^{s} is not sparse-feasible; the closed-form $R(1-\beta) + C(\beta)$ does not apply and the sparse branch must be evaluated directly from (26).

Operational meaning. The sparse branch governs the behavior of the ensemble average $\mathbb{E}[Z_n(\beta|\mathcal{C})]$ when the latter is inflated by atypical codebooks. Specifically, when $\psi_{\mathrm{s}}(\beta, R) > \psi_{\mathrm{b}}(\beta, R)$ (which occurs for sufficiently small R, as characterized by the phase boundary $R^\star(\beta)$ derived in Theorem 3.9 below), the annealed free energy satisfies $\psi(\beta, R) = \psi_{\mathrm{s}}(\beta, R)$, meaning that $\mathbb{E}[Z_n(\beta|\mathcal{C})]$ is dominated by a sub-exponential fraction of codebooks with atypical codewords (those containing a sparse-type codeword with $I_Q(X;Y) > R$). A typical codebook has $\frac{1}{n} \log Z_n(\beta|\mathcal{C}) \approx \psi_{\mathrm{b}}(\beta, R)$.

3.2 The Bulk Branch: Operational Meaning

We now motivate why the bulk branch $\psi_{\mathrm{b}}(\beta, R)$ deserves a separate study within the analysis of $\psi(\beta, R)$. As said before, the bulk branch, $\psi_{\mathrm{b}}(\beta, R)$, is the component of $\psi(\beta, R)$ driven by *typical* codewords — those whose joint type with the output satisfies $I_Q(X;Y) \leq R$, i.e., codewords that are plausible given the code rate. A typical random codebook contains no sparse-type codewords (those with $I_Q(X;Y) > R$) with high probability, so for a typical fixed codebook $Z_n(\beta|\mathcal{C}) \doteq e^{n\psi_{\mathrm{b}}(\beta,R)}$ at the exponential scale. The annealed free energy $\psi(\beta, R) = \max\{\psi_{\mathrm{b}}(\beta, R), \psi_{\mathrm{s}}(\beta, R)\}$ can exceed $\psi_{\mathrm{b}}(\beta, R)$ (when the sparse branch dominates) because it is inflated by the rare codebooks that happen to contain a sparse-type codeword; In the language of statistical physics, $\psi_{\mathrm{b}}(\beta, R)$ is the *quenched* free energy, which is the free energy of a typical random code, while $\psi(\beta, R)$ is the *annealed* free energy (the free energy of the disorder-averaged partition function). To support this observation, we now state the self-averaging property of $Z_n(\beta|\mathcal{C})$, for $\beta \geq 1$ and $R \geq R^\star(\beta)$. The proof appears in the Appendix.

Theorem 3.5 (self-averaging and quenched free energy)**.** *For $\beta \geq 1$ and $R > R^\star(\beta)$:*

(i) $\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big\{\log Z_n(\beta|\mathcal{C})\big\} = \psi_b(\beta, R).$

(ii) $\frac{1}{n} \log Z_n(\beta|\mathcal{C}) \xrightarrow{\text{a.s.}} \psi_b(\beta, R).$

The condition $R > R^\star(\beta)$ is precisely the regime (regions A and B in the phase diagram) where the bulk branch dominates the annealed free energy: $\psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R)$ (Theorem 3.9). This is needed for the upper bound in the proof: Markov's inequality on $\mathbb{E}[Z_n]$ gives $(1/n) \log Z_n \leq \psi(\beta, R)$, and this equals $\psi_{\mathrm{b}}(\beta, R)$ only when $R > R^\star(\beta)$. The lower bound of the proof holds under the weaker condition $R > I^{\mathrm{b}}(\beta)$, so region C ($I^{\mathrm{b}}(\beta) < R < R^\star(\beta)$), where a typical codebook still lives in the bulk phase but the annealed average is inflated by rare sparse-type codebooks, remains open. Whether the result also extends to the condensed phase $R \leq I^{\mathrm{b}}(\beta)$, or to $\beta < 1$, likewise remains open.

By the definition of Rényi entropy, $H_\beta(Y) = \frac{1}{1-\beta} \log \sum_y P(y)^\beta$, so $\log Z_n(\beta|\mathcal{C}) = \log \sum_{y^n} P_{Y^n|\mathcal{C}}(y^n)^\beta = (1-\beta)\, H_\beta(P_{Y^n|\mathcal{C}})$ exactly, for every codebook. Since $\frac{1}{n} \log Z_n(\beta|\mathcal{C}) \to \psi_{\mathrm{b}}(\beta, R)$ a.s. (Theorem 3.5), the bulk branch encodes the Rényi entropy rate of the output mixture at order β:

$$\psi_{\mathrm{b}}(\beta, R) \doteq \frac{1-\beta}{n}\, H_\beta(P_{Y^n|\mathcal{C}}) \tag{31}$$

for a typical fixed codebook.

3.3 The Bulk Branch: Formula and Phase Structure

Definition 3.6 (The bulk optimizer Q^{b}_β and its mutual information $I^b(\beta)$)**.** *For each $\beta > 0$, define*

$$Q^b_\beta = \arg\max_{Q_{XY}:\, Q_X = P_X} \big[(1-\beta) I_Q(X;Y) + F(Q_{XY})\big], \tag{32}$$

the unconstrained maximizer of the bulk objective, and set

$$I^b(\beta) := I_{Q^b_\beta}(X;Y). \tag{33}$$

Obviously, for $R \geq I^{\mathrm{b}}(\beta)$,

$$\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{b,u}}(\beta) := \sup_{Q_{XY}:\ Q_X = P_X} \big[(1-\beta) I_Q(X;Y) + F(Q_{XY})\big]. \tag{34}$$

The curve $R = I^{\mathrm{b}}(\beta)$ is therefore the *bulk phase boundary*: for $R \geq I^{\mathrm{b}}(\beta)$ the rate constraint $I_Q(X;Y) \leq R$ is inactive (bulk phase), while for $R < I^{\mathrm{b}}(\beta)$ it is active (condensed phase).

The next lemma establishes an inequality relation between $I^{\mathrm{b}}(\beta)$ and $I^{\mathrm{s}}(\beta)$.

Lemma 3.7 (Ordering of the two optimizers)**.** *For $\beta \geq 1$:*

$$I^{b}(\beta) \leq I^{s}(\beta), \tag{35}$$

with strict inequality for $\beta > 1$. At $\beta = 1$, $I^{b}(1) = I^{s}(1) = I(X;Y)$.

Proof. By the optimality of Q_β^{b} in the bulk objective:

$$\begin{aligned} F(Q_\beta^{\mathrm{b}}) + (1-\beta) I^{\mathrm{b}}(\beta) &\geq F(Q_\beta^{\mathrm{s}}) + (1-\beta) I^{\mathrm{s}}(\beta) \\ &= C(\beta) + (1-\beta) I^{\mathrm{s}}(\beta), \end{aligned} \tag{36}$$

which for $\beta > 1$ is equivalent to

$$I^{\mathrm{s}}(\beta) - I^{\mathrm{b}}(\beta) \geq \frac{C(\beta) - F(Q_\beta^{\mathrm{b}})}{\beta - 1} \tag{37}$$

where the right-hand side is clearly non-negative since the denominator $\beta - 1$ is positive and the numerator is non-negative as $C(\beta)$ is the global maximum of $F(Q_{XY})$, which cannot be smaller than $F(Q_\beta^{\mathrm{b}})$. Therefore, $I^{\mathrm{s}}(\beta) - I^{\mathrm{b}}(\beta)$ is non-negative as well. □

3.4 Phase Structure of the Annealed Free Energy

The annealed free energy formula (Theorem 3.1) holds for all $\beta > 0$. However, the analysis of the phase boundary between the bulk and sparse branches is restricted here to the range $\beta \geq 1$, for two reasons. The first reason is motivational: In the statistical-mechanics language, $\beta \geq 1$ is the low-temperature regime where $Z_n(\beta)$ is closely related to Rényi entropies of order at least one. This range captures the operationally most relevant quantities: $\beta = 1$ (Shannon entropy, resolvability threshold), $\beta = 2$ (collision entropy, birthday attacks), $\beta \to \infty$ (minimum entropy, $-\log \max_{y^n} P_{Y^n}(y^n)$), and general $\beta > 1$ (guessing moments $\mathbb{E}[G(Y^n)^s] \doteq e^{nsH_{1/(1+s)}}$ [9]. The second reason is that the range $0 < \beta < 1$ does not appear to lend itself to closed-form analysis, the main issue being the lack of guarantee concerning the uniqueness of the solution R to the equation $\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{s}}(\beta, R)$, which is the phase boundary of the total annealed free energy, as will be defined next. In particular, for $0 < \beta < 1$, the branches $\psi_{\mathrm{b}}(\beta, R)$ and $\psi_{\mathrm{s}}(\beta, R)$ may or may not cross, and existence and uniqueness of a crossing are not established in general.

Definition 3.8 (Annealed Phase Boundary)**.** *For $\beta \geq 1$, the* annealed phase boundary *$R^\star(\beta)$ is the unique rate at which the bulk and sparse branches are equal:*

$$\psi_b\big(\beta, R^\star(\beta)\big) = \psi_s\big(\beta, R^\star(\beta)\big). \tag{38}$$

Existence and uniqueness of $R^\star(\beta) \in (I^{\mathrm{b}}(\beta), I^{\mathrm{s}}(\beta))$ are established in Theorem 3.9(b) below, together with an explicit formula.

To facilitate the need to keep track of the various phase boundaries and the quantities associated with them so far, the following table summarizes those ingredients.

Summary of the two optimizers.

	Q^{b}_β	Q^{s}_β
Definition	Unconstrained maximizer of $\psi_{\mathrm{b}}(\beta, R)$ objective (eq. (32))	Gibbs distribution: $Q^{\mathrm{s}}_\beta(y\|x) \propto [W(y\|x)]^\beta$ per letter (eq. (27))
Maximizes	$H_Q(Y) - \beta\big[I_Q(X;Y) + \ell(Q_{XY})\big]$	$H_Q(Y\|X) - \beta\ell(Q_{XY}) = F(Q_{XY})$
MI at optimum	$I^{\mathrm{b}}(\beta)$	$I^{\mathrm{s}}(\beta)$
Governs	Bulk condensation boundary $R = I^{\mathrm{b}}(\beta)$	Bulk/sparse crossover $R = R^\star(\beta)$
At $\beta = 1$	True channel W, $I^{\mathrm{b}}(1) = I(X;Y)$	True channel W, $I^{\mathrm{s}}(1) = I(X;Y)$

The following theorem provides a characterization of the phase structure of the annealed free energy.

Theorem 3.9. *For any DMC W, any P_X, and any $\beta \geq 1$, $R > 0$:*

(a) For $R \leq I^{\mathrm{s}}(\beta)$, Q^{s}_β is sparse-feasible and

$$\psi_s(\beta, R) = R(1-\beta) + C(\beta). \tag{39}$$

(b) (Annealed phase boundary, $\beta \geq 1$.) Let

$$R^\star(\beta) := \frac{C(\beta) - \psi_{b,u}(\beta)}{\beta - 1}, \quad \beta > 1, \qquad R^\star(1) := I(X;Y). \tag{40}$$

For all $\beta \geq 1$, $R^\star(\beta) \in [I^b(\beta), I^s(\beta)]$, it is the unique rate at which $\psi_s(\beta, R) = \psi_b(\beta, R)$, and

$$\psi(\beta, R) = \begin{cases} \psi_s(\beta, R) & R < R^\star(\beta), \\ \psi_b(\beta, R) & R \geq R^\star(\beta). \end{cases} \tag{41}$$

The curve $R = R^\star(\beta)$ is the *annealed phase boundary*.

Proof. Part (a) was already shown earlier (but we include it here as part of the theorem for completeness since it is used in part (b)): Observe that for $R \leq I^{\mathrm{s}}(\beta)$, we have $I_{Q^{\mathrm{s}}_\beta}(X;Y) \geq R$, so it is sparse-feasible. Since Q^{s}_β achieves the global maximum, $F(Q^{\mathrm{s}}_\beta) = \sup_{I_Q(X;Y) > R} F(Q_{XY}) = C(\beta)$, and so, $\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta)$.

As for part (b), observe that for $R \in (I^b(\beta), I^s(\beta))$, both simplified formulas hold simultaneously: $\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{b,u}}(\beta)$, since $R > I^{\mathrm{b}}(\beta))$ and $\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta)$ (by part (a), since $R < I^{\mathrm{s}}(\beta)$). Setting $\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{s}}(\beta, R)$ and solving gives immediately

$$R^\star(\beta) = \frac{C(\beta) - \psi_{\mathrm{b,u}}(\beta)}{\beta - 1}. \tag{42}$$

Substituting $\psi_{\mathrm{b,u}}(\beta) = F(Q_\beta^{\mathrm{b}}) + (1-\beta)I^{\mathrm{b}}(\beta)$, we end up with:

$$R^\star(\beta) = I^{\mathrm{b}}(\beta) + \frac{C(\beta) - F(Q_\beta^{\mathrm{b}})}{\beta - 1}. \tag{43}$$

Since $\beta - 1 > 0$ and $C(\beta) \geq F(Q_\beta^{\mathrm{b}})$, the second term is non-negative, giving $R^\star(\beta) \geq I^{\mathrm{b}}(\beta)$, with strict inequality since $F(Q_\beta^{\mathrm{b}}) < C(\beta)$ strictly for $\beta > 1$ (Lemma 3.7).

For the upper bound, recall from the proof of Lemma 3.7 that

$$C(\beta) - F(Q_\beta^{\mathrm{b}}) \leq (\beta - 1)\big(I^s(\beta) - I^b(\beta)\big), \tag{44}$$

which gives

$$\frac{C(\beta) - F(Q_\beta^{\mathrm{b}})}{\beta - 1} \leq I^s(\beta) - I^b(\beta). \tag{45}$$

Substituting into (43): $R^\star(\beta) \leq I^{\mathrm{s}}(\beta)$, again with strict inequality for $\beta > 1$. Hence $R^\star(\beta) \in (I^{\mathrm{b}}(\beta), I^{\mathrm{s}}(\beta))$, confirming the formula is valid.

To establish the uniqueness of the solution $R^\star(\beta)$ to the equation $\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{s}}(\beta, R)$, it remains to rule out any additional solutions outside the interval $[I^{\mathrm{b}}(\beta), I^{\mathrm{s}}(\beta)]$. For $R \leq I^{\mathrm{b}}(\beta)$, Q_β^{s} is still sparse-feasible (since $R \leq I^{\mathrm{b}}(\beta) < I^{\mathrm{s}}(\beta)$), so part (a) gives $\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta)$. The bulk constraint is active, so $\psi_{\mathrm{b}}(\beta, R) \leq \psi_{\mathrm{b,u}}(\beta)$, hence:

$$\begin{aligned} \Delta(\beta, R) &:= \psi_{\mathrm{s}}(\beta, R) - \psi_{\mathrm{b}}(\beta, R) \\ &= R(1-\beta) + C(\beta) - \psi_{\mathrm{b}}(\beta, R) \\ &\geq R(1-\beta) + C(\beta) - \psi_{\mathrm{b,u}}(\beta). \end{aligned} \tag{46}$$

The right side is positive at $R = I^{\mathrm{b}}(\beta)$ (as shown above) and increasing as R decreases (since $1 - \beta < 0$). Hence $\Delta(\beta, R) > 0$ for all $R \leq I^{\mathrm{b}}(\beta)$. For $R \geq I^{\mathrm{s}}(\beta)$, consider the following. Since $I^{\mathrm{b}}(\beta) \leq I^{\mathrm{s}}(\beta) \leq R$, the bulk constraint is inactive and $\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{b,u}}(\beta)$. Since $R \geq I^{\mathrm{s}}(\beta) > R^\star(\beta)$, and $\psi_{\mathrm{b,u}}(\beta) = R^\star(\beta)(1-\beta) + C(\beta)$ (from the formula of $R^\star(\beta)$), we have:

$$\psi_{\mathrm{b}}(\beta, R) = \psi_{\mathrm{b,u}}(\beta) = R^\star(\beta)(1-\beta) + C(\beta) > R(1-\beta) + C(\beta) \geq \psi_{\mathrm{s}}(\beta, R), \tag{47}$$

where the strict inequality uses $R > R^\star(\beta)$ and $1 - \beta < 0$, and the last inequality is because $\psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + \sup_{\{Q:\ Q_X = P_X,\ I_Q(X;Y) > R\}} F(Q_{XY}) \leq R(1-\beta) + C(\beta)$. Hence $\Delta(\beta, R) < 0$ for all $R \geq I^{\mathrm{s}}(\beta)$. □

4 A Numerical Example

In this section, we provide a numerical example and illustrate the behavior the two branches of the annealed free energy as well as the phase boundary curves.

All phase diagrams are computed for an example of a Z-channel. The details are as follows. The input and output alphabets are $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and the crossover

probability is $p = 0.45$, i.e., $W(0|0) = 1$, $W(1|0) = 0$, $W(0|1) = 0.45$, and $W(1|1) = 0.55$. The input type is given by $P_X = (0.5, 0.5)$. Accordingly, the resulting channel mutual information is $I(X;Y) = 0.2441$.

Figure 1 plots all phase boundaries together in the relevant quadrant of the (β, R) plane.

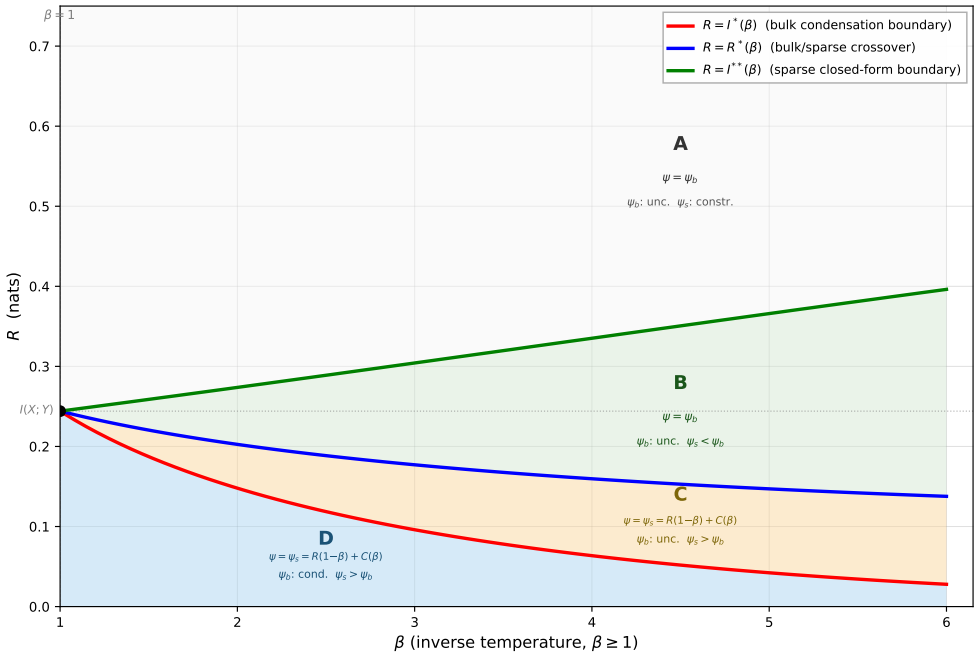

Figure 1: Phase boundaries for the Z-channel, $\beta \geq 1$. **Red:** $R = I^{\mathrm{b}}(\beta)$, the bulk branch boundary (bulk rate constraint activates). **Blue:** $R = R^{\star}(\beta)$, the annealed boundary where bulk = sparse (closed-form formula, Theorem 3.9(b)). **Green:** $R = I^{\mathrm{s}}(\beta)$, the sparse branch boundary. All three meet at $(\beta, R) = (1, I(X;Y))$ (filled dot). Four regions (see text): **A** $(R > I^{\mathrm{s}}(\beta))$: bulk and sparse both unconstrained; **B** $(R^{\star}(\beta) < R < I^{\mathrm{s}}(\beta))$: sparse branch dominates $(\psi(\beta, R) = \psi_{\mathrm{s}}(\beta, R))$; **C** $(I^{\mathrm{b}}(\beta) < R < R^{\star}(\beta))$: bulk branch dominates $(\psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R))$; **D** $(R < I^{\mathrm{b}}(\beta))$: bulk branch dominates, bulk condensed.

Three phase boundary curves are visible in Figure 1:

- $R = I^{\mathrm{b}}(\beta)$ (red): the bulk branch boundary, where the unconstrained bulk optimizer crosses the rate constraint. At $\beta = 1$: $I^{\mathrm{b}}(1) = I(X;Y)$.
- $R = R^{\star}(\beta)$ (blue): the annealed boundary where the bulk and sparse branches are equal, given by the closed-form formula (40). At $\beta = 1$: $R^{\star}(1) = I(X;Y)$. Decreasing in β, staying above $I^{b}(\beta)$.
- $R = I^{\mathrm{s}}(\beta)$ (green): the sparse branch boundary, where Q^{s}_{β} (which maximizes $F(Q_{XY}) = H_Q(Y|X) - \beta\ell(Q_{XY})$) has $I_{Q^{\mathrm{s}}_{\beta}}(X;Y) = R$. For $R > I^{\mathrm{s}}(\beta)$ the sparse closed-form ceases to hold. At $\beta = 1$: $I^{\mathrm{s}}(1) = I(X;Y)$.

All three curves meet at the point $(\beta, R) = (1, I(X;Y))$, confirming that the soft-covering threshold is the unique point where all regions collapse to a single point.

For $\beta \geq 1$, the three phase boundaries divide the quadrant $\{\beta \geq 1, R > 0\}$ into four regions:

- **A** $(R > I^{\mathrm{s}}(\beta)$, above all curves): $\psi_{\mathrm{b}}(\beta, R)$ is unconstrained $(I^{\mathrm{b}}(\beta) < R)$ and $\psi_{\mathrm{s}}(\beta, R)$ is at its optimum $(I^{\mathrm{s}}(\beta) < R)$. Both branches are unconstrained; $\psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R)$ (bulk dominates since $R^{\star}(\beta) < I^{\mathrm{s}}(\beta)$).

- **B** ($R^\star(\beta) < R < I^{\mathrm{s}}(\beta)$): Sparse branch dominates: $\psi(\beta, R) = \psi_{\mathrm{s}}(\beta, R) = R(1-\beta) + C(\beta)$ (closed-form linear formula). Rare-event codewords inflate the ensemble average; a typical codebook is in the bulk phase.
- **C** ($I^{\mathrm{b}}(\beta) < R < R^\star(\beta)$): Bulk branch dominates: $\psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R)$. The bulk rate constraint is inactive ($R > I^{\mathrm{b}}(\beta)$); no condensation.
- **D** ($R < I^{\mathrm{b}}(\beta)$, below the red curve): Bulk branch dominates with an active rate constraint: condensed phase. The bulk optimizer is constrained to $I_Q(X;Y) = R$.

For the very same z-channel example, figures 2, 3, and 4 display $\psi_{\mathrm{b}}(\beta, R)$ and $\psi_{\mathrm{s}}(\beta, R)$ as functions of $\beta \geq 1$ for three representative values of R (one below $I(X;Y)$, another one equal to $I(X;Y)$, and yet another one above $I(X;Y)$) and contrasts them with the *i.i.d. reference free energy*

$$\psi_{\mathrm{iid}}(\beta) := \frac{1}{n} \log \left(\sum_{y^n} [P_Y(y^n)]^\beta \right) = \log \left(\sum_{y \in \mathcal{Y}} [P_Y(y)]^\beta \right), \tag{48}$$

where P_Y is the output distribution induced by P_X and W. This is the free energy of the product distribution $P_Y^{\otimes n}$, i.e., the distribution to which P_{Y^n} converges for $R > I(X;Y)$ based on well known soft-covering results. Clearly, $\psi_{\mathrm{iid}}(\beta)$ is a purely smooth function with no phase transitions whatsoever.

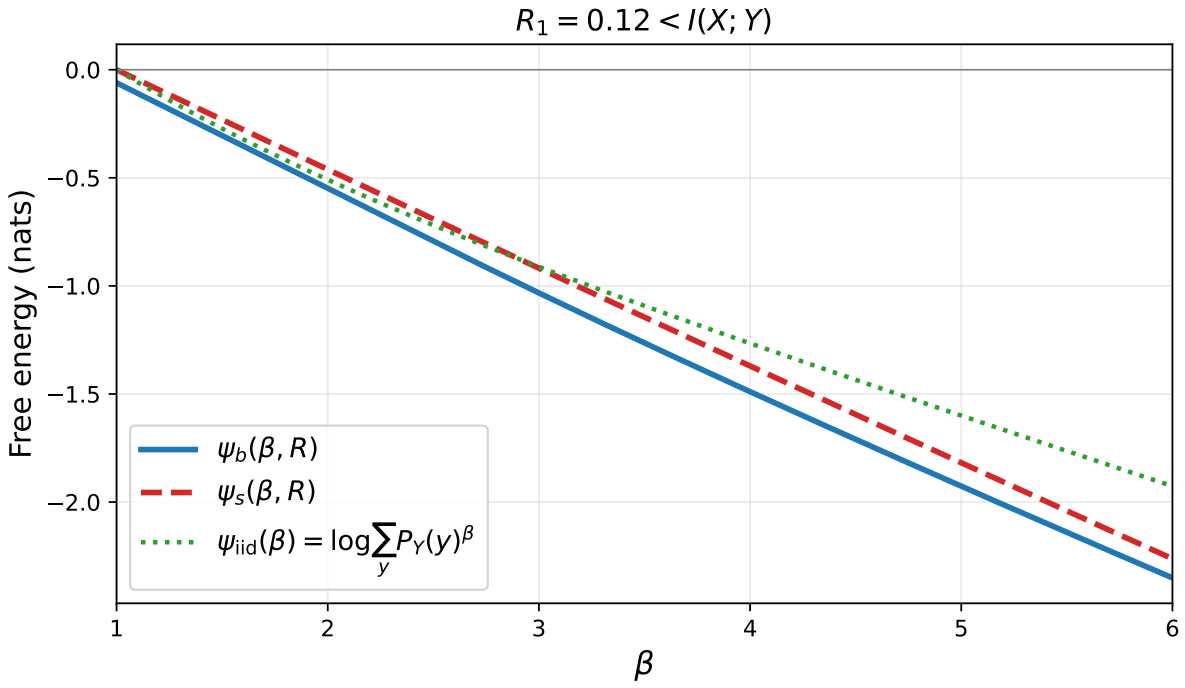

Figure 2: $R_1 = 0.12 < I(X;Y)$. The sparse branch $\psi_{\mathrm{s}}(\beta, R_1)$ lies above the bulk branch for all $\beta \geq 1$, so the annealed free energy is dominated by the sparse branch; both lie well above the i.i.d. reference, reflecting ensemble inflation by rare codebooks.

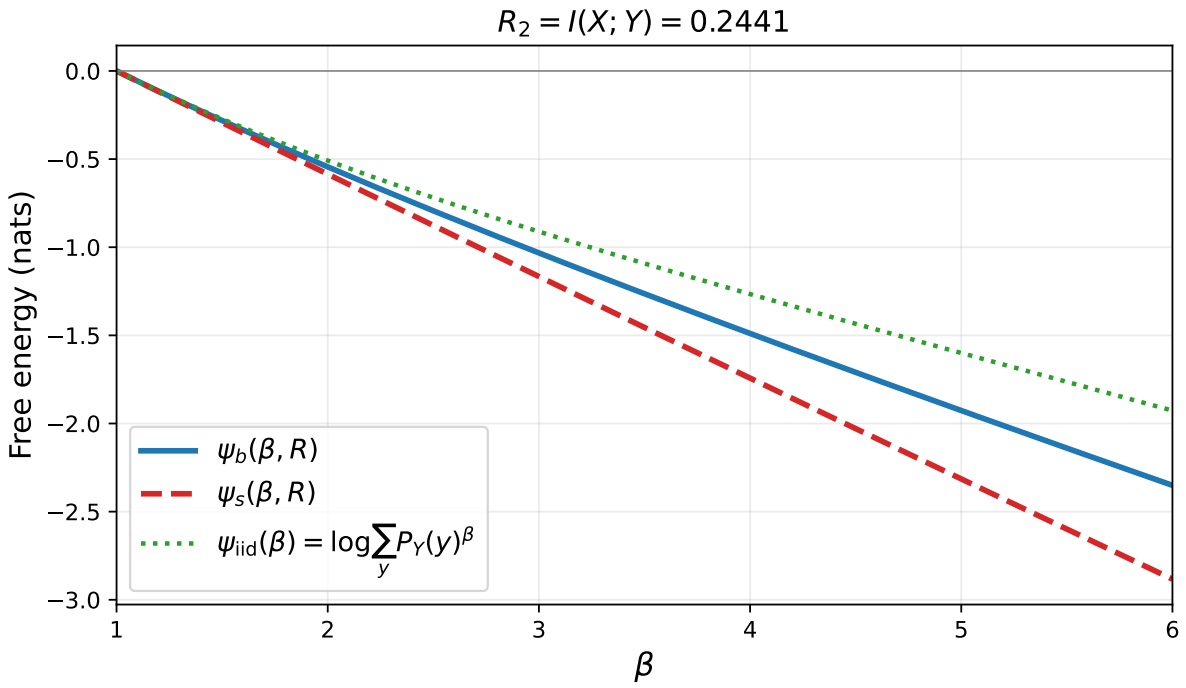

Figure 3: $R_2 = I(X;Y) = 0.2441$ nats (the soft-covering threshold). Both branches and the i.i.d. reference all vanish at $\beta = 1$ (since $Z_n(1) = 1$), and diverge differently for $\beta > 1$.

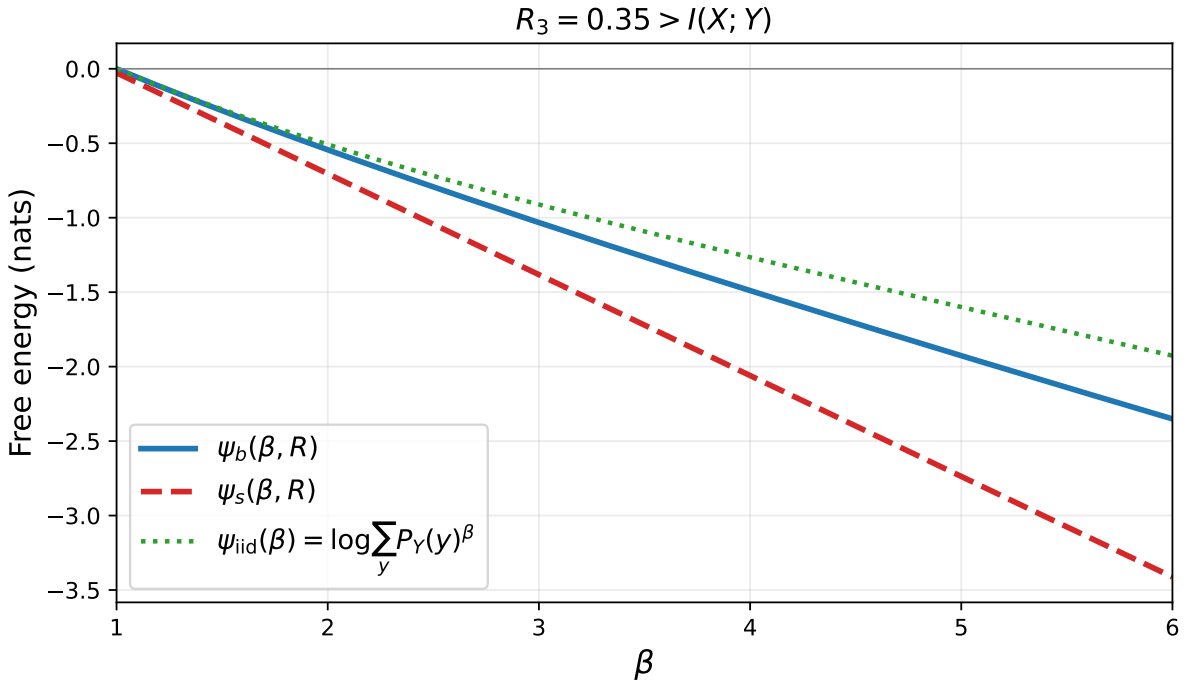

Figure 4: $R_3 = 0.35 > I(X;Y)$. The bulk branch dominates throughout and tracks the i.i.d. reference closely, confirming that above the soft-covering threshold the output mixture approximates $P_Y^{\otimes n}$ in the Rényi sense. In all three figures, $\psi_{\text{iid}}(\beta) \leq \psi(\beta, R)$, with strict inequality below $I(X;Y)$: the random-coding mixture has strictly higher Rényi entropy than the i.i.d. output distribution at every order $\beta \geq 1$.

A central message of this paper is that the classical soft-covering threshold $R = I(X;Y)$ is far from the end of the story, and as this numerical example illustrates, we refine and deepen those results in this work. As mentioned in the Introduction, classical results, like those of [1], [3], [7] and [8], tell us that $\|P_{Y^n} - P_Y^{\otimes n}\|_{\mathrm{TV}} \to 0$ (and in other metrics) if and only if $R > I(X;Y)$, and provide exponential rates for this convergence. Our statistical-physics analysis reveals that, even well above this threshold, P_{Y^n} retains non-trivial internal structure that is invisible to total variation. Figures 2, 3, and 4 make this visible. The three plots of $\psi_{\mathrm{b}}(\beta, R)$, $\psi_{\mathrm{s}}(\beta, R)$, and the i.i.d. reference $\psi_{\mathrm{iid}}(\beta) = \log\left(\sum_y [P_Y(y)]^\beta\right)$ against β for $R_1 < I(X;Y) = R_2 < R_3$. At the classical threshold $R_2 = I(X;Y)$ (second diagram), all three quantities vanish together at $\beta = 1$ as required, but diverge for $\beta > 1$: the free energy $\psi(\beta, R_2)$ lies strictly above $\psi_{\mathrm{iid}}(\beta)$ for all $\beta > 1$, meaning that P_{Y^n} has strictly higher Rényi entropy than $P_Y^{\otimes n}$ at every order. Above the threshold at $R_3 > I(X;Y)$ (third diagram), the bulk branch closely tracks $\psi_{\mathrm{iid}}(\beta)$, confirming the soft-covering intuition that the output mixture approaches the i.i.d. law — but the gap, while small, remains nonzero at finite β, and quantifying it requires $\psi_{\mathrm{b}}(\beta, R_3)$. Below the threshold at $R_1 < I(X;Y)$ (first diagram), the gap $\psi(\beta, R_1) - \psi_{\mathrm{iid}}(\beta)$ is large and grows with β: the sparse branch dominates the ensemble average, and the random-coding output is far from i.i.d. in the Rényi sense even though TV-distance arguments do not apply here at all.

Specifically, $\psi_{\mathrm{b}}(\beta, R)$ is the Rényi entropy rate of P_{Y^n} at order β for a typical codebook, and its phase transition at $R = I^{\mathrm{b}}(\beta)$ implies a qualitative change in how probability mass is distributed over output sequences. In the *bulk phase* ($R > I^{\mathrm{b}}(\beta)$), a typical output sequence is supported by exponentially many codewords—the output distribution is diffuse and "spread out" in a way consistent with the intuition behind soft-covering. In the *condensed phase* ($R < I^{\mathrm{b}}(\beta)$), a typical output is supported by only sub-exponentially few codewords, meaning the mass of P_{Y^n} concentrates on a sparse set of sequences, even though the TV distance to $P_Y^{\otimes n}$ may already be negligible. This condensation is visible in Figure 2: the first diagram ($R_1 < I(X;Y)$) shows $\psi_{\mathrm{s}}(\beta, R_1) > \psi_{\mathrm{b}}(\beta, R_1)$ for all $\beta \geq 1$, so the ensemble average is dominated by the rare sparse-branch codebooks, while the third diagram ($R_3 > I(X;Y)$) shows $\psi_{\mathrm{b}}(\beta, R_3) > \psi_{\mathrm{s}}(\beta, R_3)$ confirming typical-codebook dominance throughout.

The gap $\psi(\beta, R) - \psi_{\mathrm{iid}}(\beta) \geq 0$ visible in all three diagrams has a direct operational meaning: it measures how much the random-coding output mixture exceeds the i.i.d. distribution in Rényi entropy at order β. Classical soft-covering results establish that this gap vanishes in total variation for $R > I(X;Y)$; The figures show that in Rényi entropy the gap is nonzero for every $\beta > 1$ and every finite R, decaying to zero only as $R \to \infty$.

5 Applications and Implications

In this section, we summarize several applications and implications of the results in this work. Further investigation of these applications is deferred to future work.

1. *Refined output statistics.* Resolvability shows $\|P_{Y^n} - (P_Y)^n\|_{\mathrm{TV}} \to 0$ for $R > I(X;Y)$. Our phase diagram reveals that even in this regime, P_{Y^n} may be in the condensed phase (below $I^{\mathrm{b}}(\beta)$ for large β), where mass concentrates on outputs supported by a sub-exponential number of codewords. This is invisible to the total variation distance and other distances.

2. *Guesswork.* $\mathbb{E}[G(Y^n)^s] \doteq e^{nsH_{1/(1+s)}(P_{Y^n})}$ [9]. Since $H_\beta(P_{Y^n|\mathcal{C}}) = \frac{1}{(1-\beta)n}\log Z_n(\beta|\mathcal{C})$ and $\frac{1}{n}\log Z_n(\beta|\mathcal{C}) \doteq \psi(\beta, R)$ (the full annealed free energy, including both bulk and sparse branches), the guessing exponent at $\beta = 1/(1+s)$ is $\frac{s}{n(1-\beta)}\log Z_n(\beta|\mathcal{C}) \doteq \frac{s}{1-\beta}\psi(\beta, R)$. The phase structure of $\psi(\beta, R)$ — both its bulk phase transition at $R = I^{\mathrm{b}}(\beta)$ and its bulk-sparse crossover at $R = R^\star(\beta)$ — therefore produces regime changes in guessing complexity absent in the i.i.d. case.

3. *Security.* In the condensed phase, even when $P_{Y^n} \approx (P_Y)^n$ in total variation, only a sub-exponential number of codewords generate a typical output. A computationally powerful adversary can identify the message by examining which codewords are consistent with the observed output. Semantic security may therefore require operating strictly in the bulk phase for all β relevant to the adversary's test.

4. *Hypothesis testing and the Chernoff exponent.* A natural question is how well one can distinguish the random-code output mixture $P_{Y^n|\mathcal{C}}$ from the i.i.d. reference $P_Y^{\otimes n}$. With $P_Y = \mathrm{Unif}(\mathcal{Y})$ (achieved by choosing P_X appropriately), this is a direct test of whether soft covering has succeeded. The annealed Chernoff exponent for this binary hypothesis test equals

$$\xi(R) = \max_{0\le\beta\le 1}\big[(1-\beta)\log|\mathcal{Y}| - \psi(\beta, R)\big], \tag{49}$$

where $\psi(\beta, R)$ is the annealed free energy of Theorem 3.1. Since both branches of $\psi(\beta, R)$ are fully characterized by our results, (49) is completely determined. The optimizer $\beta^\star \in [0, 1]$ satisfies $\psi'(\beta^\star, R) = -\log|\mathcal{Y}|$, so the phase structure of $\psi(\beta, R)$ directly governs the behavior of $\xi(R)$: the bulk/sparse phase boundaries produce regime changes in the Chernoff exponent as R varies.

With $P_Y = \mathrm{Unif}$, two clean consequences follow. First, $\xi(R) = 0$ if and only if $R \ge I(X;Y)$: the two distributions are not exponentially distinguishable for $R \ge I(X;Y)$, i.e., no test can achieve an exponentially small total error probability, exactly at the soft-covering threshold. (Sub-exponential error rates, e.g., polynomial in n, are not excluded.) Second, (49) is the Legendre–Fenchel transform of $\psi(\beta, R)$ with respect to β, evaluated at $\log|\mathcal{Y}|$; it therefore equals $\frac{1}{n}\log\sum_{y^n} P_{Y^n|\mathcal{C}}(y^n)^{\beta^\star}|\mathcal{Y}|^{-n(1-\beta^\star)}$, which is precisely the Rényi divergence $D_\alpha(P_{Y^n|\mathcal{C}} \,\|\, P_Y^{\otimes n})$ at order $\alpha = \beta^\star \in [0, 1]$. This connects (49) directly to the Rényi resolvability results of Yu and Tan [8], who showed that $D_\alpha(P_{Y^n|\mathcal{C}}\|P_Y^{\otimes n}) \to 0$ if and only if $R > I(X;Y)$ for $\alpha \le 1$, while for $\alpha > 1$ the threshold is $R = I^{\mathrm{b}}(\alpha)$ — the bulk condensation boundary of our phase diagram. Our analysis provides the exact *rate* of divergence at every R and α, and reveals how the two-branch phase structure of $\psi(\beta, R)$ determines qualitative changes in this rate.

6 Conclusion

We developed a statistical-mechanical framework for the output distribution of random codes, centered on the annealed free energy $\psi(\beta, R)$ of the partition function $Z_n(\beta) = \sum_{y^n}[P_{Y^n}(y^n)]^\beta$ and its two-branch phase structure. This dichotomy between the two branches has concrete operational consequences. For guesswork, the phase structure of $\psi(\beta, R)$ — both the bulk phase transition at $R = I^{\mathrm{b}}(\beta)$ and the bulk-sparse crossover

at $R = R^\star(\beta)$ — translates directly (via Arıkan's formula [9]) into regime changes in the exponential growth rate of guessing moments $\mathbb{E}[G(Y^n)^s]$. Similar comments apply to hypothesis testing and perhaps other application areas. The annealed free energy adds yet another layer: its phase boundary $R = R^\star(\beta)$ separates the regime where the ensemble average is dominated by rare, atypical codebooks (for example, the sparse branch dominates in Figure 2) from the regime where it reflects the behavior of a typical codebook (the bulk branch dominates in Figure 4). The gap $\psi(\beta, R) - \psi_{\mathrm{b}}(\beta, R) > 0$ in the sparse-dominant regime quantifies the extent to which ensemble averages can be misleading—a warning particularly relevant when analyzing the security of specific random codes rather than random coding in the mean.

Future research directions include: a rigorous derivation of exponentially tight error exponents for the hypothesis test of Application 4 via the type-class enumeration method of [15], together with a phase-diagram analysis of the Neyman–Pearson exponent tradeoff in the (R, τ) plane; a full treatment of the guesswork regime changes produced by the bulk/sparse phase boundaries; quantitative semantic security bounds in terms of $I^{\mathrm{b}}(\beta)$ and $R^\star(\beta)$; extension to Markov and mixed sources; connections to mismatched decoding exponents; and a complete treatment of the $\beta < 1$ regime. Finally, we note that the self-averaging property of $Z_n(\beta|\mathcal{C})$ has been established in Theorem 3.5: for $\beta \geq 1$ and $R > R^\star(\beta)$, the bulk branch $\psi_{\mathrm{b}}(\beta, R)$ coincides with the quenched free energy $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\{\log Z_n(\beta|\mathcal{C})\}$. Whether the same holds in region C ($I^{\mathrm{b}}(\beta) < R < R^\star(\beta)$), the condensed phase $R \leq I^{\mathrm{b}}(\beta)$, or for $\beta < 1$, remains an interesting open problem.

Appendix: Proof of Theorem 3.5

Throughout this proof, we define

$$U_n := \frac{1}{n} \log Z_n(\beta|\mathcal{C}). \tag{50}$$

We begin with part (i). To establish the upper bound, observe that by Jensen's inequality, $\mathbb{E}\{U_n\} \leq \frac{1}{n} \log \mathbb{E}[Z_n(\beta|\mathcal{C})] \to \psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R)$, where the last equality is by the postulate $R \geq R^\star(\beta)$. It follows that $\limsup_{n\to\infty} \mathbb{E}\{U_n\} \leq \psi_{\mathrm{b}}(\beta, R)$. As for the lower bound, fix any joint type Q_{XY} with $Q_X = P_X$ and $I_Q(X;Y) \leq R$, and let Q_Y be its Y-marginal. For every output sequence y^n of type Q_Y, each codeword of joint type Q_{XY} with y^n contributes exactly $e^{-n\ell(Q_{XY})}$ to $\sum_m W^n(y^n|x^n(m))$, so

$$P_{Y^n|\mathcal{C}}(y^n) \;\geq\; \frac{N(Q_{XY}|y^n)}{M} \cdot e^{-n\ell(Q_{XY})}, \tag{51}$$

and, since $\beta \geq 1$,

$$Z_n(\beta|\mathcal{C}) \;\geq\; \sum_{y^n:\, \hat{P}_{y^n} = Q_Y} \left(\frac{N(Q_{XY}|y^n)}{M}\right)^{\beta} e^{-n\beta\ell(Q_{XY})}. \tag{52}$$

As observed before, $N(Q_{XY}|y^n)$ is binomial with e^{nR} trials and a probability of a single success of the exponential order of $e^{-nI_Q(X;Y)}$, so its mean μ_n is exponentially $e^{n(R-I_Q(X;Y))} \to \infty$. By the Chernoff bound, for any $\delta \in (0, 1)$:

$$\Pr\left\{N(Q_{XY}|y^n) \leq (1-\delta)\mu_n\right\} \leq \exp\left(-\frac{\delta^2\mu_n}{2}\right) = \exp\left(-\frac{\delta^2}{2} e^{n[R-I_Q(X;Y)]}\right), \tag{53}$$

which is *doubly* exponentially small (see, also Theorem 4.1 of [15] and in particular, eqs. (D.16), (D.17) in its proof therein, which establish the doubly exponential decay). A union bound over all joint types $\{Q_{XY}\}$ and all $|\mathcal{T}(Q_Y)| \leq e^{nH_Q(Y)}$ sequences y^n gives

$$\Pr\{\mathcal{E}_n^{\mathrm{c}}\} \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} \cdot e^{nH_Q(Y)} \cdot \exp\left(-\tfrac{\delta^2}{2}\, e^{n[R-I_Q(X;Y)]}\right) \;\longrightarrow\; 0, \tag{54}$$

where $\mathcal{E}_n$ is the event that $N(Q_{XY}|y^n) \geq (1-\delta)\mu_n$ for all types Q_{XY} and all $y^n \in \mathcal{T}(Q_Y)$ simultaneously. On $\mathcal{E}_n$, the bound (52) gives

$$\begin{aligned} Z_n(\beta|\mathcal{C}) &\geq |\mathcal{T}(Q_Y)| \cdot \left(\frac{(1-\delta)\mu_n}{M}\right)^{\beta} e^{-n\beta\ell(Q_{XY})} \\ &\doteq\; (1-\delta)^{\beta} \cdot \exp\Big(n\big[H_Q(Y) - \beta\big(I_Q(X;Y) + \ell(Q_{XY})\big)\big]\Big). \end{aligned} \tag{55}$$

Taking the supremum over all Q_{XY} with $Q_X = P_X$ and $I_Q(X;Y) \leq R$, we obtain

$$U_n \;\geq\; \psi_{\mathrm{b}}(\beta, R) + \frac{\log(1-\delta)}{n} \quad \text{on } \mathcal{E}_n. \tag{56}$$

Taking expectations of both sides of (56) and using $U_n \geq (1-\beta)\log|\mathcal{Y}|$ deterministically (since $Z_n(\beta) \geq |\mathcal{Y}|^{n(1-\beta)}$ by Jensen):

$$\mathbb{E}\{U_n\} \geq \left(\psi_{\mathrm{b}}(\beta, R) + \frac{\log(1-\delta)}{n}\right) \Pr\{\mathcal{E}_n\} + (1-\beta)\log|\mathcal{Y}| \cdot \Pr\{\mathcal{E}_n^{\mathrm{c}}\}. \tag{57}$$

Since $\Pr[\mathcal{E}_n] \to 1$, we get $\liminf_{n\to\infty} \mathbb{E}\{U_n\} \geq \psi_{\mathrm{b}}(\beta, R)$, which together with

$$\limsup_{n\to\infty} \mathbb{E}\{U_n\} \leq \psi_{\mathrm{b}}(\beta, R), \tag{58}$$

implies

$$\lim_{n\to\infty} E\{U_n\} = \psi_{\mathrm{b}}(\beta, R), \tag{59}$$

thus completing the proof of part (i).

Proceeding to part (ii), we begin by proving that $\psi_{\mathrm{b}}(\beta, R)$ is an almost sure upper bound. By Markov's inequality, for any $\varepsilon > 0$:

$$\Pr\left\{U_n \geq \psi_{\mathrm{b}}(\beta, R) + \varepsilon\right\} \leq \frac{\mathbb{E}[Z_n(\beta|\mathcal{C})]}{e^{n(\psi_{\mathrm{b}}(\beta,R)+\varepsilon)}} \doteq e^{-n\varepsilon}, \tag{60}$$

where we have used the fact that $\frac{1}{n}\log\mathbb{E}[Z_n(\beta|\mathcal{C})] \to \psi_{\mathrm{b}}(\beta, R)$, which holds because $\psi(\beta, R) = \psi_{\mathrm{b}}(\beta, R)$ for $R > R^\star(\beta)$. Since $\sum_n e^{-n\varepsilon} < \infty$, the Borel–Cantelli lemma yields $U_n \leq \psi_{\mathrm{b}}(\beta, R) + \varepsilon$ for every $\varepsilon > 0$ eventually a.s., or equivalently, $\limsup_{n\to\infty} U_n \leq \psi_{\mathrm{b}}(\beta, R)$ a.s. For the compatible almost sure lower bound, note that we have already proved in (56) that whenever $\mathcal{E}_n$ occurs, $U_n \geq \psi_{\mathrm{b}}(\beta, R) + \frac{\log(1-\delta)}{n}$, which is means that $\mathcal{E}_n \subseteq \mathcal{F}_n := \{U_n \geq \psi_{\mathrm{b}}(\beta, R) + \frac{\log(1-\delta)}{n}\}$, or equivalently, $\mathcal{F}_n^{\mathrm{c}} \subseteq \mathcal{E}_n^{\mathrm{c}}$. Thus,

$$\sum_{n=1}^{\infty} \Pr\left\{U_n < \psi_{\mathrm{b}}(\beta, R) + \frac{\log(1-\delta)}{n}\right\} = \sum_{n=1}^{\infty} \Pr\{\mathcal{F}_n^{\mathrm{c}}\} \leq \sum_{n=1}^{\infty} \Pr\{\mathcal{E}_n^{\mathrm{c}}\} < \infty \tag{61}$$

where the summability of $\Pr\{\mathcal{E}_n^{\mathrm{c}}\}$ is due to the fact that each term is bounded by a doubly exponentially small quantity. By the Borel-Cantelli lemma, $U_n - \frac{\log(1-\delta)}{n} \geq \psi_{\mathrm{b}}(\beta, R)$ eventually a.s., or equivalently, $\liminf_{n\to\infty} U_n \geq \psi_{\mathrm{b}}(\beta, R)$ a.s., which together with the matching upper bound yields $\lim_{n\to\infty} U_n = \psi_{\mathrm{b}}(\beta, R)$ a.s., completing the proof.

References

[1] Han, T. S.; Verdú, S. "Approximation theory of output statistics," *IEEE Trans. Inf. Theory* **1993**, vol. 39, no. 3, pp. 752–772.

[2] Cuff, P. "Soft covering with high probability," *Proc. IEEE Int. Symp. Inf. Theory (ISIT)* **2016**, Barcelona, Spain, pp. 2963–2967.

[3] Hayashi, M. "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to wire-tap channel," *IEEE Trans. Inf. Theory* **2006**, vol. 52, no. 4, pp. 1562–1575.

[4] Hou, J.; Kramer, G. "Effective secrecy: Reliability, confusion and stealth," *Proc. IEEE Int. Symp. Inf. Theory (ISIT)* **2014**, pp. 601–605.

[5] Mézard, M.; Montanari, A. *Information, Physics, and Computation*, Oxford University Press, New York, 2009.

[6] Merhav, N. "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory* **2010**, vol. 6, no. 1–2, pp. 1–212.

[7] Yu, L.; Tan, V. Y.-F. "Exact channel resolvability exponents for the soft-covering problem," *IEEE Trans. Inf. Theory* **2020**, vol. 66, no. 4, pp. 2098–2112.

[8] Yu, L.; Tan, V. Y.-F. "Rényi resolvability and its applications to the wiretap channel," *IEEE Trans. Inf. Theory* **2019**, vol. 65, no. 3, pp. 1862–1897.

[9] Arıkan, E. "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory* **1996**, vol. 42, no. 1, pp. 99–105.

[10] Blahut, R. E. "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory* **1974**, vol. 20, no. 4, pp. 405–417.

[11] Sourlas, N. "Spin-glass models as error-correcting codes," *Nature* **1989**, vol. 339, pp. 693–695.

[12] Montanari, A. "Turbo codes: The phase transition," *European Physical Journal B* **2001**, vol. 18, pp. 321–333.

[13] Csiszár, I.; Körner, J, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[14] Csiszár, I. "The method of types," *IEEE Trans. Inf. Theory* **1998**, vol. 44, no. 6, pp. 2505–2523.

[15] Merhav, N.; Weinberger, N. "A toolbox for refined information-theoretic analyses with applications," *Foundations and Trends in Communications and Information Theory* **2025**, vol. 22, no. 1, pp. 1–184.