Signal Estimation in Gaussian Noise: A Statistical Physics Perspective

Neri Merhav
Electrical Engineering Dept.
Technion–Israel Inst. of Tech.
Haifa 32000, Israel
Email: merhav@ee.technion.ac.il

Dongning Guo
Electrical Engineering & Computer Science Dept.
Northwestern University
Evanston, IL 60208, USA
Email: dGuo@Northwestern.edu

Shlomo Shamai (Shitz)
Electrical Engineering Dept.
Technion–Israel Inst. of Tech.
Haifa 32000, Israel
Email: sshlomo@ee.technion.ac.il

***Abstract*— We consider the problem of signal estimation from a statistical mechanical perspective, using a relationship between the minimum mean square error (MMSE), of estimating a signal, and the mutual information between this signal and its noisy version. We derive several statistical-mechanical relationships between a few important quantities in this problem area, such as the MMSE, the mutual information, the Fisher information, the free energy, and a generalized notion of temperature. We also draw analogies and differences between certain relations pertaining to the estimation problem and the parallel relations in statistical physics. Finally, we provide several examples, demonstrating how analysis tools in statistical physics prove useful in the MMSE analysis. In most of the examples, the corresponding statistical-mechanical systems turn out to exhibit phase transitions, which are reflected as irregularities in behavior of the MMSE.**

I. INTRODUCTION

Relations between information theory and statistical physics are by no means new and much of the research on them focuses on identifying ‘mappings’ between problems in information theory and certain many-particle system models in statistical physics which are analogous in their mathematical formalisms. This work relates information theory and statistical physics with yet another area – estimation theory, whose link to information theory is provided by the *I–MMSE relation* [5] (equivalent to the de Bruijn identity [2, Theorem 17.7.2]), which relates the minimum mean square error (MMSE), of estimating a signal in additive white Gaussian noise, to the mutual information between this signal and its noisy version. An earlier work where the I–MMSE relation was studied from a physical viewpoint is [13]: an attempt was made there to offer to this relation an alternative proof which was rooted in thermodynamics, but this required a generalization of the theory of thermodynamics.

Our work is triggered by [13], but it is on a substantially different route: We simply use the I–MMSE relation together with analysis tools of statistical physics. The underlying idea is that quite often, the mutual information can be evaluated using these tools and then related to the MMSE using the I–MMSE relation. This combination proves rather powerful, because it enables us to identify situations where discontinuities in the mean square error (as a function of the signal-to-noise ratio (SNR)) are inherent to the model in the sense that they are apparent even for the optimum, MMSE estimator, and not just artifacts of some ad hoc estimator. In these situations, these discontinuities are intimately related to *phase transitions* in the parallel statistical-mechanical systems.

We derive several statistical-mechanical relationships between a few quantities: the MMSE, the mutual information, the Fisher information, the free energy, and a generalized notion of temperature. We also draw analogies and differences between certain relations pertaining to the estimation problem and the parallel relations in statistical physics. Finally, we offer several examples demonstrating how statistical-mechanical analysis tools prove useful in the analysis of the MMSE. In most of the examples, the corresponding statistical-mechanical systems turn out to exhibit phase transitions, which are reflected as discontinuities in the MMSE. For reasons of space limitation, we do not provide here the full detailed derivations. These appear in the full version [9].

II. STATISTICAL PHYSICS BACKGROUND

Consider a physical system with n particles, which can be in a variety of microscopic states (e.g., positions and momenta). For each microstate $\boldsymbol{x} = (x_1, \ldots, x_n)$, there is an energy, given by a *Hamiltonian* (energy function), $\mathcal{E}(\boldsymbol{x})$. One of the most fundamental results in statistical physics is that, in thermal equilibrium, the probability of finding the system at state $\boldsymbol{x}$ is given by the *Boltzmann–Gibbs* distribution

$$P(\boldsymbol{x}) = \frac{e^{-\beta\mathcal{E}(\boldsymbol{x})}}{Z(\beta)} \tag{1}$$

where $\beta = 1/(kT)$, k being Boltzmann’s constant and T being temperature, and $Z(\beta)$ is the normalization constant, called the *partition function*, which is given by

$$Z(\beta) = \sum_{\boldsymbol{x}} e^{-\beta\mathcal{E}(\boldsymbol{x})}, \text{ or } Z(\beta) = \int \mathrm{d}\boldsymbol{x}\, e^{-\beta\mathcal{E}(\boldsymbol{x})}, \tag{2}$$

depending on whether $\boldsymbol{x}$ is discrete or continuous. The partition function is a key quantity from which many macroscopic physical quantities can be derived, for example, the average energy w.r.t. (1) is $\bar{E} = \boldsymbol{E}\{\mathcal{E}(\boldsymbol{X})\} = -(\mathrm{d}/\mathrm{d}\beta)\ln Z(\beta)$. The Shannon entropy pertaining to (1) is $S(\beta) = \ln Z(\beta) + \beta\cdot\bar{E}$, which suggests the differential equation

$$\beta\,\dot{\psi}(\beta) - \psi(\beta) = S(\beta), \tag{3}$$

The work of D. Guo is supported by the NSF under grant CCF-0644344 and DARPA under grant W911NF-07-1-0028. The work of S. Shamai is supported in part by the Israel Science Foundation.

where $\psi(\beta) = -\ln Z(\beta)$ and $\dot{\psi}$ means the derivative of ψ. Equivalently, eq. (3) can be rewritten as:

$$\beta \frac{\mathrm{d}}{\mathrm{d}\beta}\left[\frac{\psi(\beta)}{\beta}\right] = \frac{S(\beta)}{\beta}, \tag{4}$$

whose solution is easily found to be

$$\psi(\beta) = \beta E_0 - \beta \int_\beta^\infty \frac{d\hat{\beta} S(\hat{\beta})}{\hat{\beta}^2}, \tag{5}$$

where E_0 is a constant of integration obtained from the limit $\beta \to \infty$. Thus, the log-partition function at a given temperature is expressed as a *heat integral* of $S(\beta)$, i.e., an integral of the entropy at all lower temperatures. Taking the derivative of (5), we obtain:

$$\bar{E} = \dot{\psi}(\beta) = E_0 - \int_\beta^\infty \frac{d\hat{\beta} S(\hat{\beta})}{\hat{\beta}^2} + \frac{S(\beta)}{\beta}, \tag{6}$$

where the first two terms form the free energy.

III. THEORETICAL DERIVATIONS

Let $(\boldsymbol{X}, \boldsymbol{Y})$ be random vectors in $\mathbb{R}^n$, related by $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}$, where the noise $\boldsymbol{N}$ consists of i.i.d., zero-mean, Gaussian random variables with variance $1/\beta$, where $\beta > 0$ denotes the signal-to-noise ratio, or the inverse temperature in the statistical-mechanical viewpoint. It is assumed that $\boldsymbol{X}$ and $\boldsymbol{N}$ are independent. As is well known, the conditional mean estimator $\boldsymbol{E}(\boldsymbol{X}|\boldsymbol{Y})$ achieves the MMSE, $\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) \overset{\triangle}{=} \boldsymbol{E}\|\boldsymbol{X} - \boldsymbol{E}\{\boldsymbol{X}|\boldsymbol{Y}\}\|^2$. Theorem 2 in [5] establishes the I–MMSE relation:[1]

$$\frac{\mathrm{d}}{\mathrm{d}\beta} I(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}). \tag{7}$$

Consider next the posterior distribution. Let $Q(\boldsymbol{x})$ denote the probability mass function of $\boldsymbol{x}$ and $P(\boldsymbol{y}|\boldsymbol{x})$ denote the channel induced by $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}$, then

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{Q(\boldsymbol{x})P(\boldsymbol{y}|\boldsymbol{x})}{P_\beta(\boldsymbol{y})} = \frac{Q(\boldsymbol{x})\exp[-\beta\|\boldsymbol{y} - \boldsymbol{x}\|^2/2]}{Z(\beta|\boldsymbol{y})}, \tag{8}$$

where $P_\beta(\boldsymbol{y})$ is the channel output density and we have defined

$$Z(\beta|\boldsymbol{y}) \overset{\triangle}{=} \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) e^{-\frac{\beta}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2} = \left(\frac{2\pi}{\beta}\right)^{\frac{n}{2}} P_\beta(\boldsymbol{y}) \tag{9}$$

where we have assumed that $\boldsymbol{x}$ is discrete. The function $Z(\beta|\boldsymbol{y})$ is similar to a partition function but the exponentials are weighted by $\{Q(\boldsymbol{x})\}$, which are independent of β. As explained in [6, p. 3713], this is not an issue because $\{Q(\boldsymbol{x})\}$ can be interpreted as multiple states whose number is proportional to $Q(\boldsymbol{x})$ of the same energy.

Denote by $\boldsymbol{E}_\beta$ the expectation operator w.r.t. joint pdf of $(\boldsymbol{X}, \boldsymbol{Y})$ induced by β. It is clear that $-\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\}$ differs from the output differential entropy $h(\boldsymbol{Y})$ by a constant only. The following results are all proved in [9]:

Proposition 1: The mutual information, MMSE and average free energy are related by

$$I(\boldsymbol{X};\boldsymbol{Y}) = -(n/2) - \boldsymbol{E}_\beta\left\{\ln Z(\beta|\boldsymbol{Y})\right\}, \tag{10}$$

$$\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = -2\frac{\partial}{\partial\beta}\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\}, \tag{11}$$

$$\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = (n/\beta) + \mathrm{Cov}\{\|\boldsymbol{Y} - \boldsymbol{X}\|^2, \ln Z(\beta|\boldsymbol{Y})\}. \tag{12}$$

The first term in (12), $n/\beta = \boldsymbol{E}_\beta\|\boldsymbol{Y} - \boldsymbol{X}\|^2$, is the amount of noise in the raw data $\boldsymbol{Y}$. The second term, which is always negative, designates the *noise suppression level* due to MMSE estimation relative to the raw data. We henceforth denote $\Delta = -\mathrm{Cov}\{\|\boldsymbol{Y} - \boldsymbol{X}\|^2, \ln Z(\beta|\boldsymbol{Y})\}$. In particular, when the actual channel input $\boldsymbol{x}$ dominates (9), the correlation cancels the first term and the MMSE vanishes; otherwise, the correlation is weaker. In view of the following relation due to the orthogonality principle,

$$\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{E}\|\boldsymbol{Y} - \boldsymbol{X}\|^2 - \boldsymbol{E}\|\boldsymbol{Y} - \boldsymbol{E}(\boldsymbol{X}|\boldsymbol{Y})\|^2, \tag{13}$$

it is clear that Δ is also related to the free energy. We next relate Δ to the Fisher information (FI) and then to a new generalized notion of temperature [11] via the de Bruijn identity. By de Bruijn's identity, if $\boldsymbol{W}$ is a vector of i.i.d. standard normal components, independent of $\boldsymbol{X}$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} h(\boldsymbol{X} + \sqrt{t}\boldsymbol{W}) = \frac{1}{2}\mathrm{tr}\{J(\boldsymbol{X} + \sqrt{t}\boldsymbol{W})\} \tag{14}$$

where $h(\boldsymbol{Y})$ is differential entropy and $J(\boldsymbol{Y})$ is the FI matrix associated with $\boldsymbol{Y}$ w.r.t. a translation parameter. Due to the relation between $Z(\beta|\boldsymbol{y})$ and the $P(\boldsymbol{y})$ of (9), the FI can also be related to the free energy by

$$\mathrm{tr}\{J(\boldsymbol{Y})\} = \sum_{i=1}^n \boldsymbol{E}\left\{\left[\left.\frac{\partial \ln Z(\beta|\boldsymbol{y})}{\partial y_i}\right|_{\boldsymbol{y}=\boldsymbol{Y}}\right]^2\right\}. \tag{15}$$

By comparing the I-MMSE relation (7) and de Bruijn's identity, we find $\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{X}+\boldsymbol{N}) = -\beta^{-2}\mathrm{tr}\{J(\boldsymbol{Y})\} + n/\beta$, which is also shown in [5], where the factor $-\beta^{-2}$ accounts for the passage from t to $\beta = 1/t$. Combining this with the previous relations yields the following:

Proposition 2: $\Delta = \beta^{-2}\,\mathrm{tr}\{J(\boldsymbol{Y})\}$.

In [11, Theorem 3.1], a generalized definition of the inverse temperature is proposed, as the response of the entropy to small energy perturbations, using de Bruijn's identity. As a consequence of that definition, the generalized inverse temperature in [11] turns out to be proportional to the FI of $\boldsymbol{Y}$, and thus, in our setting, it is also proportional to $\beta^2\Delta$. It is noted that whenever the system undergoes a phase transition, then Δ, and hence also the effective temperature, may exhibit a discontinuity.

Additional relationships can be obtained in analogy to certain relations in statistical thermodynamics that were mentioned in Section II. Using (10)–(12), we obtain:

$$\begin{aligned} &\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\} - \beta\frac{\mathrm{d}}{\mathrm{d}\beta}\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y}) \\ &\quad = \frac{\beta}{2}\mathrm{Cov}\{\|\boldsymbol{Y} - \boldsymbol{X}\|^2, \ln Z(\beta|\boldsymbol{Y})\} - I(\boldsymbol{X};\boldsymbol{Y}). \end{aligned} \tag{16}$$

---

[1] The units of all information measures in the paper are nats and all logarithms are natural.

Thus, redefining the function $\psi(\beta) = -\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\}$, we obtain a differential equation similar to (3):

$$\beta\,\dot{\psi}(\beta) - \psi(\beta) = \Sigma(\beta) \tag{17}$$

where $\Sigma(\beta) = \frac{\beta}{2}\mathrm{Cov}\{\|\boldsymbol{Y}-\boldsymbol{X}\|^2, \ln Z(\beta|\boldsymbol{Y})\} - I(\boldsymbol{X};\boldsymbol{Y})$. The solution to this equation is the same as (5), except that $S(\beta)$ is replaced by $\Sigma(\beta)$ and the ground-state energy E_0 is redefined as $E_0 = \boldsymbol{E}_\beta\{\min_{\boldsymbol{x}} \|\boldsymbol{Y}-\boldsymbol{x}\|^2\}$. Consequently, $\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = 2\dot{\psi}(\beta)$, where

$$\dot{\psi}(\beta) = E_0 - \int_\beta^\infty \frac{\mathrm{d}\hat{\beta}\Sigma(\hat{\beta})}{\hat{\beta}^2} + \frac{\Sigma(\beta)}{\beta} \tag{18}$$

and one can easily identify the contributions of the free energy and the internal energy (heat), as was done in Section II.

IV. EXAMPLES

In the following examples, the asymptotic MMSE is calculated using the I–MMSE relation in conjunction with statistical-mechanical techniques for evaluating the mutual information, or the partition function pertaining to the posterior distribution. In some of the examples, the mutual information can also be obtained through existing channel capacity results from information theory. In the last example, however, we are not aware of any alternative to the calculation.

A. Gaussian I.I.D. Inputs: Let the components of $\boldsymbol{X}$ be i.i.d. standard Gaussian. In this case, $Z(\beta|\boldsymbol{y})$ is proportional to a Gaussian density, so $\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\} = -(n/2)\ln(1+\beta) - n/2$ and its negative derivative is $n/[2(1+\beta)]$, which is indeed half of the MMSE. Here, we have: $\Delta = n/(\beta(1+\beta))$ and the relation $\mathrm{tr}\{J(\boldsymbol{Y})\} = \beta^2\Delta$ can be easily verified. There is no phase transition in the MMSE.

B. Random Codebook on a Sphere Surface. Let $\boldsymbol{X}$ assume a uniform distribution over a codebook $\mathcal{C} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$, $M = e^{nR}$, where each codeword $\boldsymbol{x}_i$ is drawn independently under the uniform distribution over the surface of the n-dimensional sphere, which is centered at the origin, and whose radius is $\sqrt{n}$. The code is capacity achieving (the input becomes essentially i.i.d. Gaussian as $n \to \infty$).

Without loss of generality, assume $\boldsymbol{x}_0$ to be the transmitted codeword. Here, for a given $\boldsymbol{y}$, we have: $Z(\beta|\boldsymbol{y}) = Z_c(\beta|\boldsymbol{y}) + Z_e(\beta|\boldsymbol{y})$ where

$$Z_c(\beta|\boldsymbol{y}) = e^{-nR}\exp[-\beta\|\boldsymbol{y}-\boldsymbol{x}_0\|^2/2] \tag{19}$$

is typically about $e^{-nR}e^{-\beta\cdot n/(2\beta)} = e^{-n(R+1/2)}$ and

$$Z_e(\beta|\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\mathcal{C}\setminus\{\boldsymbol{x}_0\}} e^{-nR}\exp[-\beta\|\boldsymbol{y}-\boldsymbol{x}\|^2/2]\ . \tag{20}$$

Let $N(\epsilon)$ denote the number of codewords $\{\boldsymbol{x}\}$ in $\mathcal{C}\setminus\{\boldsymbol{x}_0\}$ for which $\|\boldsymbol{y}-\boldsymbol{x}\|^2/2$ lies between $n\epsilon$ and $n(\epsilon+\mathrm{d}\epsilon)$. Then

$$Z_e(\beta|\boldsymbol{y}) \doteq e^{-nR}\int_{\mathbb{R}} \mathrm{d}\epsilon N(\epsilon)e^{-\beta n\epsilon} \tag{21}$$

where "$\doteq$" denotes that the two sides of the equation are exponential in n with asymptotically identical exponent. The integral can be treated as in the random energy model (REM) of spin glasses [3], a model of disordered magnetic materials where the energy levels pertaining to the various configurations of the system $\{\mathcal{E}(\boldsymbol{x})\}$ are i.i.d. (see [10, Chapters 5,6], and [7] for a detailed analysis of random code ensemble performance).

Using large deviations theory, it can be shown that

$$\psi(\beta) = -\lim_{n\to\infty}\frac{\ln Z(\beta|\boldsymbol{y})}{n} = \begin{cases} \frac{1}{2}\ln(1+\beta)+\frac{1}{2}, & \beta < \beta_R \\ R+\frac{1}{2} & \beta \geq \beta_R \end{cases}$$

where $\beta_R = e^{2R}-1$. Note that $\psi(\beta)$ is continuous but not smooth at $\beta = \beta_R$. Thus, there is a phae transition at β_R. Now,

$$\lim_{n\to\infty}\frac{\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y})}{n} = 2\frac{\mathrm{d}\psi(\beta)}{\mathrm{d}\beta} = \begin{cases} \frac{1}{1+\beta}, & \beta < \beta_R \\ 0, & \beta \geq \beta_R \end{cases} \tag{22}$$

which means that this is a first order phase transition in the MMSE: As long as $\beta \geq \beta_R$, which means that the code rate R is below the capacity, the MMSE essentially vanishes since $\boldsymbol{x}_0$ is reliably decodable; but for code rates above the capacity, the MMSE is as if the input were i.i.d. standard Gaussian. Such a phase transition has previously been shown for good binary codes in general in [12] using the I-MMSE relation.

C. Hierarchical Code for the Degraded Broadcast Channel. Consider the following code ensemble: First, randomly draw $M_1 = e^{nR_1}$ cloud-center vectors $\{\boldsymbol{u}_i\}$ on the $\sqrt{n}$-sphere. Then, for each $\boldsymbol{u}_i$, randomly draw $M_2 = e^{nR_2}$ codewords $\{\boldsymbol{x}_{i,j}\}$ according to $\boldsymbol{x}_{i,j} = \alpha\boldsymbol{u}_i + \sqrt{1-\alpha^2}\,\boldsymbol{v}_{i,j}$, where $\alpha \in (0,1)$ and $\{\boldsymbol{v}_{i,j}\}$ are randomly drawn uniformly and independently on the $\sqrt{n}$-sphere. This means that $\|\boldsymbol{x}_{i,j} - \alpha\boldsymbol{u}_i\|^2 = n(1-\alpha^2)$.

For high enough SNR, the codeword $\boldsymbol{x}_{i,j}$ can be decoded; while at certain lower SNR only the cloud center $\boldsymbol{u}_i$ can be decoded but not $\boldsymbol{v}_{i,j}$. In the following we show the corrsponding phase transitions of the MMSE.

Without loss of generality, let $\boldsymbol{x}_{0,0}$, belonging to cloud center $\boldsymbol{u}_0$, be the input to the Gaussian channel. We decompose the partition function as follows:

$$Z(\beta|\boldsymbol{y}) = e^{-nR}\sum_{i,j}\exp(-\beta\|\boldsymbol{y}-\boldsymbol{x}_{i,j}\|^2/2) \tag{23}$$

$$\begin{aligned} &= e^{-nR}\exp(-\beta\|\boldsymbol{y}-\boldsymbol{x}_{0,0}\|^2/2) \\ &\quad + e^{-nR}\sum_{j\geq 1}\exp(-\beta\|\boldsymbol{y}-\boldsymbol{x}_{0,j}\|^2/2) \\ &\quad + e^{-nR}\sum_{i\geq 1}\sum_{j}\exp(-\beta\|\boldsymbol{y}-\boldsymbol{x}_{i,j}\|^2/2) \end{aligned} \tag{24}$$

$$\stackrel{\triangle}{=} Z_c(\beta|\boldsymbol{y}) + Z_{e1}(\beta|\boldsymbol{y}) + Z_{e2}(\beta|\boldsymbol{y}) \tag{25}$$

where once again, $Z_c(\beta|\boldsymbol{y})$ – the contribution of the correct codeword, is typically about $e^{-n(R+1/2)}$. The other two terms $Z_{e1}(\beta|\boldsymbol{y})$ and $Z_{e2}(\beta|\boldsymbol{y})$ correspond to contributions of incorrect codewords from the same cloud and from other clouds, respectively.

For typical codes, we can use large deviations theory to show that

$$Z_c + Z_{e1} \doteq \exp\left\{-n\left[R_1 + \min\left\{R_2, \frac{1}{2}\ln(1+b\beta)\right\} + \frac{1}{2}\right]\right\}$$

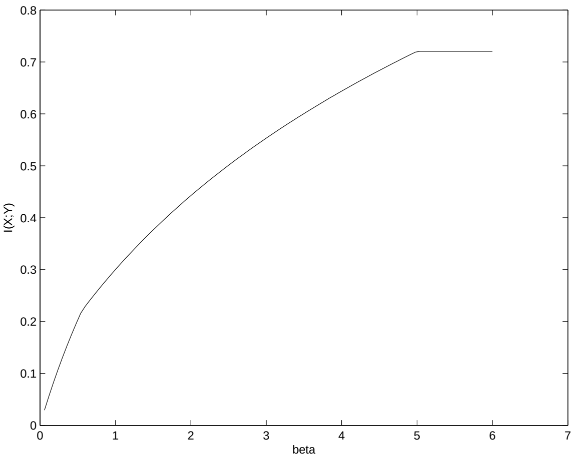

Fig. 1. Graph of $\lim_{n\to\infty} I(\boldsymbol{X};\boldsymbol{Y})/n = -\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\}/n - 1/2$ as a function of β for $R_1 = 0.1$, $R_2 = 0.6206$, and $\alpha = 0.7129$, which result in $\beta_1 = 0.5545$ and $\beta_2 = 5.001$. As can be seen quite clearly, there are phase transitions at these values of β.

where $b = 1 - \alpha^2$ and

$$\lim_{n\to\infty} \frac{\ln Z_{e2}(\beta|\boldsymbol{y})}{n} = \max_{|r_1|\le\rho_1} \max_{|r_2|\le\rho_2(r_1)} \left\{\frac{1}{2}\ln(1-r_1^2) + \frac{1}{2}\ln(1-r_2^2) - \beta\left[\frac{b}{2} + P_a - r_1 P_g - r_2\sqrt{2b(P_a - r_1 P_g)}\right]\right\}, \tag{26}$$

where $\rho_1 = \sqrt{1-e^{-2R_1}}$, $\rho_2(r_1) = \sqrt{1-e^{-2R}/(1-r_1^2)}$, $P_a = (1+1/\beta+\alpha^2)/2$, and $P_g = \alpha\sqrt{1+1/\beta}$. The above expression does not seem to lend itself to closed form analysis in an easy manner. Numerical results (cf. Fig. 1) show a reasonable match (within the order of magnitude of 1×10^{-5}) between values of $\lim_{n\to\infty} I(\boldsymbol{X};\boldsymbol{Y})/n$ obtained numerically from the asymptotic exponent of $\boldsymbol{E}_\beta\{\ln Z(\beta|\boldsymbol{Y})\}$ and those that are obtained from the expected behavior in this case:

$$\lim_{n\to\infty} \frac{I(\boldsymbol{X};\boldsymbol{Y})}{n} = \begin{cases} \frac{1}{2}\ln(1+\beta), & \beta < \beta_1 \\ R_1 + \frac{1}{2}\ln(1+\beta b), & \beta_1 \le \beta < \beta_2 \\ R = R_1 + R_2, & \beta \ge \beta_2 \end{cases}$$

where

$$\beta_1 \triangleq \frac{e^{2R_1}-1}{1-be^{2R_1}}, \quad \beta_2 \triangleq \frac{e^{2R_2}-1}{1-b},$$

and it is assumed that the parameters of the model (R_1, R_2 and α) are chosen such that $\beta_1 < \beta_2$. Accordingly, the MMSE undergoes two phase transitions, where it behaves as if the input were: (i) Gaussian i.i.d. with unit variance for $\beta < \beta_1$ (where no information can be decoded), (ii) Gaussian input of a smaller variance (corresponding to the cloud), in the intermediate range (where the cloud center is decodable, but the refined message is not), and (iii) the MMSE altogether vanishes for $\beta > \beta_2$, where both messages are reliably decodable.

The hierarchical code ensemble takes the superposition code structure which achieves the capacity region of the Gaussian broadcast channel. Consider two receivers, referred to as receiver 1 and receiver 2, with SNR equal to β_1 and β_2 respectively. Receiver 1 can decode the cloud center, whereas receiver 2 can decode the entire codeword. In other words, suppose the hierarchical code ensemble with rate pair (R_1, R_2) and parameter α is sent to two receivers with fixed SNR of γ_1 and γ_2 respectively. Then the minimum decoding error probability vanishes as long as (R_1, R_2, α) are such that

$$R_1 < \frac{1}{2}\log\left(1 + \frac{\alpha^2\gamma_1}{1+(1-\alpha^2)\gamma_1}\right), \tag{27}$$

$$R_2 < \frac{1}{2}\log\left(1+\alpha^2\gamma_2\right). \tag{28}$$

In particular, all boundary points of the capacity region can be achieved by varying the power distribution coefficient α. This capacity region result also leads to the fact that if only the cloud center is decodable, then the MMSE for the codeword $\boldsymbol{v}_{i,j}$ is no different to that if the elements of $\boldsymbol{v}_{i,j}$ were i.i.d. standard Gaussian. Knowledge of the codebook structure of $\{\boldsymbol{v}_{i,j}\}$ does not reduce the MMSE because otherwise the code cannot achieve the capacity region.

D. Hierarchical Tree-Structured Code. Let n be partitioned into two segments of length $n_1 = \lambda_1 n$ and length $n_2 = n - n_1 = \lambda_2 n$, respectively, $\lambda_1 \in (0,1)$. We randomly draw $M_1 = e^{n_1 R_1}$ first-segment codewords $\{\boldsymbol{x}_i\}$ on the surface of the $\sqrt{n_1}$-sphere, and then, for each $\boldsymbol{x}_i$, we randomly draw $M_2 = e^{n_2 R_2}$ second-segment codewords $\{\boldsymbol{x}'_{i,j}\}$ on the surface of the $\sqrt{n_2}$-sphere. The total message of $nR = n_1 R_1 + n_2 R_2$ nats ($R = \lambda_1 R_1 + \lambda_2 R_2$) is encoded in two parts: The first-segment codeword depends only on the first $n_1 R_1$ nats of the message whereas the second-segment codeword depends on the entire message.

Let $C(\beta) = \frac{1}{2}\ln(1+\beta)$. It can be shown using large deviations theory [9] that $-\frac{\log Z}{n}$ converges to $\min\{R, C(\beta)\} + 1/2$ if $R_1 > R_2$ and $\lambda_1 \min\{R_1, C(\beta)\} + \lambda_2 \min\{R_2, C(\beta)\} + 1/2$ otherwise. The MMSE then is as in (22) when $R_1 > R_2$, and given by

$$\mathrm{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = \begin{cases} \frac{1}{1+\beta}, & \beta \le \beta_{R_1} \\ \frac{\lambda_2}{1+\beta}, & \beta_{R_1} < \beta \le \beta_{R_2} \\ 0, & \beta > \beta_{R_2} \end{cases} \tag{29}$$

when $R_1 < R_2$ (where, again, $\beta_R \triangleq e^{2R} - 1$). This dichotomy between these two types of behavior have their roots in the behavior of the GREM, a generalized version of the REM, where the random energy levels of the various system configurations are correlated (rather than being i.i.d.) in an hierarchical structure [4]. The GREM turns out to have an intimate analogy with the tree-structured code ensemble considered here (see [8] for details).

The above result on the MMSE is consistent with the analysis based solely on information theoretic considerations. In case $R_1 < R_2$, the first segment code is decodable as long as $R_1 < C(\beta)$, whereas the second segment code is decodable if also $R_2 < C(\beta)$. Hence the MMSE is given by (29). In case

$R_1 > R_2$, the second-segment code is decodable if and only if the first-segment is also decodable, i.e., the two codes can be decoded jointly. This requires $R_2 < C(\beta)$, $R_1 < C(\beta)$ and $R = \lambda_1 R_1 + \lambda_2 R_2 < C(\beta)$. The last inequality dominates, hence the MMSE is given by (22) in Example IV-B.

E. Estimation of Sparse Signals. Motivated by compressed sensing applications, we consider the a model where $X_i = S_i U_i$, where $U_i \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. and independent of the binary variables S_i, where only a small fraction of the elements of $\boldsymbol{S} = (S_1, S_2, \ldots, S_n)$ are nonzero. Let $\boldsymbol{S} \sim P(\boldsymbol{s})$, then

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{s}} P(\boldsymbol{s}) \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi s_i \sigma^2}} \exp\{-x_i^2/(2 s_i \sigma^2)\} \right] \tag{30}$$

where a zero-variance Gaussian distribution is understood as the Dirac delta function. Transforming $\boldsymbol{s}$ to "spins" $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ by $\mu_i = 1 - 2s_i \in \{-1, +1\}$, we get:

$$Z(\beta|\boldsymbol{y}) = \frac{\exp\left\{-\frac{\beta(1+q/2)}{2(1+q)} \|\boldsymbol{y}\|^2\right\}}{(1+q)^{n/4}} \sum_{\boldsymbol{\mu}} P(\boldsymbol{\mu}) \exp\left\{\sum_{i=1}^{n} \mu_i h_i\right\} \tag{31}$$

where

$$h_i = -\frac{\beta^2 \sigma^2 y_i^2}{4\big(1+\beta\sigma^2\big)} + \frac{1}{4} \ln\left(1 + \beta\sigma^2\right). \tag{32}$$

By viewing the expression $\sum_{\boldsymbol{\mu}} P(\boldsymbol{\mu}) \exp\{\sum_i \mu_i h_i\}$ as the partition function of a certain spin system with a nonuniform, random field $\{H_i\}$ (whose realization is $\{h_i\}$), we can borrow techniques from statistical physics to analyze its behavior.

Assuming certain symmetry properties on the components of $\boldsymbol{s}$, we postulate that $P(\boldsymbol{\mu})$ depends on $\boldsymbol{\mu}$ only via the magnetization $m(\boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \mu_i$. Consider in particular the form $P(\boldsymbol{\mu}) = C_n \exp\{n f(m(\boldsymbol{\mu}))\}$, where C_n is a normalization constant and $f(m)$ is twice differentiable with finite first derivative on $[-1, 1]$. On substituting this $P(\boldsymbol{\mu})$ into (31), the dominant value of m as $n \to \infty$ in the expression for $Z(\beta|\boldsymbol{y})$ is found to satisfy:

$$m^* = \boldsymbol{E}\{\tanh(f'(m^*) + H)\} \tag{33}$$

and

$$\boldsymbol{E}\{\tanh^2(f'(m^*) + H)\} > 1 - 1/f''(m^*). \tag{34}$$

Clearly, m^* is the dominant magnetization *a-posteriori*, i.e., the one that dominates the posterior of $m(\boldsymbol{\mu})$ given (a typical) $\boldsymbol{y}$. Thus the asymptotic normalized mutual information is

$$\begin{aligned} \lim_{n\to\infty} \frac{I(\boldsymbol{X};\boldsymbol{Y})}{n} &= \frac{\beta(2+q)\boldsymbol{E}\{Y^2\}}{4(1+q)} - \frac{1}{2} + \frac{1}{4}\ln(1+q) \\ &\quad - \lim_{n\to\infty} \frac{\ln C_n}{n} + f'(m^*)\, m^* - f(m^*) \\ &\quad - \boldsymbol{E}\left\{\log\left[2\cosh(f'(m^*) + H)\right]\right\}. \end{aligned} \tag{35}$$

For the model of (30), H is defined by (32) with y_i replaced by Y and the expectation over Y is w.r.t. a mixture of two Gaussians centered at 0 and 1 respectively. The MMSE is equal to twice the derivative of (35) w.r.t. β. Note that the dominant value m^* is dependent on β. The expression is omitted here but can be found in [9].

The value of m which makes (34) an equality is known as a *critical point*, beyond which the solution to (33) ceases to be a local maximum and becomes a local minimum. The dominant m^* must jump elsewhere. The MMSE may also undergo an abrupt change, and so the MMSE may be discontinuous (w.r.t. these parameters) at these points. Also, as we vary one of the other parameters of the model, it might happen that the global maximum jumps from one local maximum to another.

The special case of quadratic $f(m) = am + bm^2/2$ is similar though not identical to the *random-field Curie–Weiss model* of spin systems. Eq. (33) becomes $m = \boldsymbol{E}\{\tanh(bm + a + H)\}$, similarly as in the mean field model with a random field [1]. The case $b = 0$ corresponds to i.i.d. S_i, i.e., a system of non-interacting particles, where no phase transitions can exist. Therefore, what we learn from statistical physics here is that phase transitions in the MMSE estimator cannot be a property of the sparsity alone (because sparsity may be present also for the i.i.d. case with $P\{S_i = 1\}$ small), but rather a property of strong dependency between $\{S_i\}$, whether it comes with sparsity or not.

ACKNOWLEDGEMENT

N. Merhav thanks Yonina Eldar for useful discussions concerning Example E during the early stages of this work.

REFERENCES

[1] A. Bianchi, A. Bovier, and D. Ioffe, "Sharp asymptotics for metastability in the random field Curie–Weiss model," June 2008. Available online: http://arxiv.org/abs/arXiv:0806.4478v1.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Hoboken, New Jersey: John Wiley & Sons, 2nd edition, 2006.

[3] B. Derrida, "Random–energy model: an exactly solvable model for disordered systems," *Phys. Rev. B*, vol. 24, no. 5, pp. 2613–2626, Sept. 1981.

[4] B. Derrida, "A generalization of the random energy model which includes correlations between energies," *J. de Physique – Lettres*, vol. 46, L–401-107, May 1985.

[5] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean–square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[6] N. Merhav, "An identity of Chernoff bounds with an interpretation in statistical physics and applications in information theory," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3710–3721, Aug. 2008.

[7] 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, Jan. 2009.

[8] N. Merhav, "The generalized random energy model and its application to the statistical physics of ensembles of hierarchical codes," to appear in *IEEE Trans. Inform. Theory*, March 2009.

[9] N. Merhav, D. Guo, and S. Shamai (Shitz), "Statistical physics of signa estimation in Gaussian noise: theory and examples of phase transitions," submitted to *IEEE Trans. Inform. Theory*, Dec. 2008. Available online: http://arxiv.org/abs/0812.4889.

[10] M. Mézard and A. Montanari, *Information, Physics and Computation*, draft, Nov. 9, 2007. Available online: http://www.stanford.edu/∼montanar/BOOK/book.html.

[11] K. R. Narayanan and A. R. Srinivasa, "On the thermodynamic temperature of a general distribution," Nov. 2007. Available online: http://arxiv.org/abs/arXiv:0711.1460v2.

[12] M. Peleg, A. Sanderovich and S. Shamai (Shitz), "On extrinsic information of good codes operating over Gaussian channels," *European Trans. Telecommun.*, Vol. 18, No. 2, pp. 133-139, 2007.

[13] O. Shental and I. Kanter, "Shannon meets Carnot: generalized second thermodynamic law," June 2008. Available online: http://arxiv.org/abs/0806.3133v1.