A Statistical–Mechanical View on Code Ensembles and Random Coding Exponents

Neri Merhav

Communication and Information Theory Seminar
March 27, 2008
General Background

Relations between information theory and statistical physics:

- The maximum entropy principle: Jaynes, Shore & Johnson, Burg, ...
- Physics of information: Landauer, Bennet, Maroney, Plenio & Vitelli, ...
- Large deviations theory: Ellis, Oono, McAllester, ...
- Random matrix theory: Wigner, Balian, Foschini, Telatar, Tse, Hanly, Shamai, Verdú, Tulino, ...
- Coding and spin glasses: Sourlas, Kabashima, Saad, Kanter, Mézard, Montanari, Nishimori, Tanaka, ...

Physical insights and analysis tools are ‘imported’ to IT.
In This Talk We:

- Briefly review basic background in statistical physics.
- Describe relationships between coding and spin glasses.
- Relate performance measures in coding to physical quantities.
- Develop an analysis technique inspired by stat–mech.
- Discuss extensions of the basic models.
Background in Statistical Physics

Consider a system with \( n \gg 1 \) particles which can lie in various microstates, \( \{x = (x_1, \ldots, x_n)\} \), e.g., a combination of locations, momenta, angular momenta, spins, ...

For every \( x \), \( \exists \) energy \( E(x) \) – Hamiltonian.

Example: For \( x_i = (p_i, r_i) \),

\[
E(x) = \sum_{i=1}^{n} \left( \frac{\|p_i\|^2}{2m} + mgh_i \right).
\]
In thermal equilibrium, \( x \sim \text{Boltzmann–Gibbs distribution} \):

\[
P(x) = \frac{e^{-\beta E(x)}}{Z(\beta)}
\]

where \( \beta = \frac{1}{kT} \), \( k \) – Boltzmann’s constant, \( T \) – temperature, and

\[
Z(\beta) = \sum_x e^{-\beta E(x)}, \quad \text{a normalization factor} = \text{partition function}
\]

\[
\phi(\beta) = \ln Z(\beta) \quad \Rightarrow \text{many physical quantities:}
\]

free energy: \( F = -\frac{\phi}{\beta} \); mean internal energy: \( E = -\frac{d\phi}{d\beta} \);

entropy: \( S = \phi - \beta \frac{d\phi}{d\beta} \); heat capacity: \( C = -\beta^2 \frac{d^2\phi}{d\beta^2} \);...

From now on: \( T \leftarrow kT \quad \Rightarrow \quad \beta = \frac{1}{T}. \)
Example: magnetic material – each particle has a magnetic moment (spin) – a 3D vector which tends to align with the net magnetic field = external field + effective fields of other particles.

Quantum mechanics: each spin ∈ discrete set of values, e.g., for spin $\frac{1}{2}$:

$$x_i = +\frac{1}{2} \Rightarrow +1$$

$$x_i = -\frac{1}{2} \Rightarrow -1$$
Background Cont’d: The Ising Model

\[ \mathcal{E}(x) = -H \cdot \sum_{i=1}^{n} x_i - J \cdot \sum_{\langle i,j \rangle} x_i x_j \]

\( J = 0 \) – paramagnetic: no interactions \( \Rightarrow \) spins are independent:

magnetization

\[ m \triangleq \mathbb{E} \left\{ \frac{1}{n} \sum_i X_i \right\} \]

\[ = ( +1 ) \cdot \frac{e^{\beta H}}{2 \cosh(\beta H)} + ( -1 ) \cdot \frac{e^{-\beta H}}{2 \cosh(\beta H)} = \tanh(\beta H) \]

\( J > 0 \) – ferromagnetic; \( J < 0 \) – antiferromagnetic.
The Ising Model (Cont’d)

Strong interaction $\Rightarrow$ two conflicting effects:

- 2nd law $\Rightarrow$ entropy ↑ $\Rightarrow$ disorder ↑
- Interaction energy ↓ $\Rightarrow$ order ↑.

Q: Who wins?
A: Depends on temperature:

$$Z = \sum_{\mathbf{x}} e^{-\beta \mathcal{E}(\mathbf{x})} = \sum_{E} N(E) e^{-\beta E} = \sum_{E} \exp\{S(E) - \beta E\}$$

- High temperature – disorder (paramagnetism).
- Low temperature – order: magnetization (sometimes spontaneous).

Abrupt passage $\Rightarrow$ phase transition.
Background Cont’d: Other Models

Interactions between remote pairs:

\[ \mathcal{E}_I(\mathbf{x}) = - \sum_{i,j} J_{ij} x_i x_j \]

\{J_{ij}\} with mixed signs \(\Rightarrow\) spin glass.

Disorder: \(\{J_{ij}\} = \) quenched random variables.

- Edwards–Anderson (EA): \(J_{ij} \sim\) i.i.d. Gaussian; neighbors only.
- Sherrington–Kirkpatrick (SK): \(J_{ij}\) same, but all pairs.
- \(p\)-spin glass model: Like SK, but products of \(p\) spins.
- Random Energy model (REM): \(p \to \infty \Rightarrow \{\mathcal{E}_I(\mathbf{x})\} =\) i.i.d. Gaussian.
Background Cont’d: The REM (Derrida, 1980,81)

Very simple, but rich enough for phase transitions.

\[ Z(\beta) = \sum_{x=1}^{2^n} e^{-\beta \mathcal{E}_I(x)} = \int \text{d}E \cdot N(E) e^{-\beta E} \quad \mathcal{E}_I(x) \sim \mathcal{N}(0, nJ^2/2) \]

\[ \overline{N(E)} \approx 2^n \cdot \Pr\{ \mathcal{E}_I(x) \approx E \} = 2^n \cdot e^{-E^2/(nJ^2)} = \exp\{n[\ln 2 - (E/nJ)^2]\} \]

- \( \overline{N(E)} \) with a negative exponent \( \iff |E| \geq E_0 \overset{\Delta}{=} nJ\sqrt{\ln 2} \Rightarrow N(E) \sim 0. \)
- \( |E| < E_0 \Rightarrow N(E) \) concentrates rapidly around \( \overline{N(E)}. \)

**Typical realization:**

\[ Z(\beta) \approx \int_{-E_0}^{E_0} \text{d}E \cdot \overline{N(E)} \cdot e^{-\beta E}. \]
The REM (Cont’d)

\[
Z(\beta) \approx \int_{-E_0}^{E_0} dE \cdot \exp \left\{ n \left[ \ln 2 - \left( \frac{E}{nJ} \right)^2 \right] \right\} \cdot e^{-\beta E}
\]

\[
\phi(\beta) = -\beta F(\beta) = \begin{cases} 
\ln 2 + \frac{\beta^2 J^2}{4} & \beta < \frac{2}{J} \sqrt{\ln 2} \\
\beta J \sqrt{\ln 2} & \beta \geq \frac{2}{J} \sqrt{\ln 2}
\end{cases}
\]

Phase transition at \( \beta = \beta_0 \triangleq \frac{2}{J} \sqrt{\ln 2} \):

- High temp. (\( \beta < \beta_0 \)) – paramagnetic phase: entropy > 0; \( Z(\beta) \) dominated by exponentially many \( x \)'s at \( E = -n\beta J^2 / 2 \).

- Low temp. (\( \beta \geq \beta_0 \)) – spin–glass phase: \( \phi = \) linear, entropy = 0, frozen at ground–state \( E = -E_0 \) with sub–exponentially few dominant \( x \)'s.
REM & Random Coding (Mézard & Montanari, 2008)

BSC(p) + a random code $\mathcal{C} = \{x_0, x_1, \ldots, x_{M-1}\}$, $M = e^{nR}$, (fair coin tossing).

Posterior:

$$P(x|y) = \frac{P(y|x)}{\sum_{x' \in \mathcal{C}} P(y|x')} = \frac{e^{-\ln[1/P(y|x)]}}{\sum_{x' \in \mathcal{C}} e^{-\ln[1/P(y|x')]}}.$$

Suggests a Boltzmann family:

$$P_\beta(x|y) = \frac{e^{-\beta \ln[1/P(y|x)]}}{\sum_{x' \in \mathcal{C}} e^{-\beta \ln[1/P(y|x')]}} = \frac{P_\beta(y|x)}{\sum_{x' \in \mathcal{C}} P_\beta(y|x')}.$$
REM & Random Coding (Cont’d)

Motivations:

- $\beta = \text{degree of freedom for channel uncertainty.}$
- Annealing: find ground–state by ‘cooling’.
- Finite–temperature decoding (Ruján 1993):

$$\hat{x}_t = \arg\max_a P_\beta(x_t = a|y)$$

- $\beta = 1 \Rightarrow \text{minimum bit–error probability}$
- $\beta = \infty \Rightarrow \text{minimum block–error probability.}$

- $Z(\beta|y) = \sum_{x \in \mathcal{C}} P_\beta(y|x) \exists \text{ in bounds on } P_e.$ Random $\mathcal{C} \iff \text{REM: Phase transitions ‘inherited’ from REM.}$
\( x_0 = \text{correct codeword}; \ B = \ln \frac{1-p}{p} : \)

\[
Z(\beta|y) = (1 - p)^{n\beta} \sum_{x \in C} e^{-\beta B d(x, y)}
\]

\[
= (1 - p)^{n\beta} e^{-\beta B d(x_0, y)} + (1 - p)^{n\beta} \sum_{x \in C \setminus \{x_0\}} e^{-\beta B d(x, y)}
\]

\[
\triangleq Z_c(\beta|y) + Z_e(\beta|y).
\]

\[
d(x_0, y) \approx np \Rightarrow Z_c(\beta|y) \approx (1 - p)^{n\beta} e^{-\beta B np}.
\]

\[
Z_e(\beta|y) = (1 - p)^{n\beta} \sum_{\delta=0}^{n} N_y(n\delta) e^{-\beta B n\delta}
\]

with \( N_y(n\delta) = e^{nR} . e^{n[h(\delta) - \ln 2]} \).

\( R + h(\delta) - \ln 2 < 0 \Rightarrow N_y(n\delta) \sim 0. \) Happens for \( \delta < \delta_{GV}(R) \) and \( \delta > 1 - \delta_{GV}(R) \), where \( \delta_{GV}(R) = \text{solution } \delta \leq 1/2 \) of \( R + h(\delta) - \ln 2 = 0. \)
Stat. Phys. of Code Ensembles (Cont’d)

Similar to the REM:

\[ Z_e(\beta|y) \doteq \exp\left\{ n \max_{\delta \in [\delta_{GV}(R),1-\delta_{GV}(R)]} [R + h(\delta) - \ln 2 - \beta B \delta] \right\} \overset{\Delta}{=} e^{n \phi(\beta,R)} \]

\[ \phi(\beta,R) = \begin{cases} R + \ln[p^\beta + (1-p)^\beta] - \ln 2 & \beta < \beta_c(R) \quad \text{paramagnetic} \\ \beta[\delta_{GV}(R) \ln p + (1 - \delta_{GV}(R)) \ln(1 - p)] & \beta \geq \beta_c(R) \quad \text{spin–glass} \end{cases} \]

\[ \beta_c(R) = \frac{\ln[(1 - \delta_{GV}(R))/\delta_{GV}(R)]}{\ln[(1-p)/p]} \]

\[ Z_c(\beta|y) \Rightarrow \text{ordered phase} = \text{ferromagnetic phase}. \]

Ferro–glassy boundary: \( R = C \).
Ferro–para boundary: \( T = T_0(R) = 1/\beta_0(R) \), solution to:

\[ \beta h(p) = \ln 2 - R - \ln[p^\beta + (1-p)^\beta]. \]
Phase diagram of finite-temperature decoding (Mezard & Montanari, 2008).
The Correct Decoding Exponent (M. 2007)

\[
\overline{P_c} = E \left\{ \frac{1}{M} \sum_y \max_m P(y|X_m) \right\}
\]

\[
= E \left\{ \frac{1}{M} \sum_y \lim_{\beta \to \infty} \left[ \sum_{m=0}^{M-1} P^\beta(y|X_m) \right]^{1/\beta} \right\}
\]

\[
= \frac{1}{M} \sum_y \lim_{\beta \to \infty} E \left\{ Z_e(\beta|y)^{1/\beta} \right\}
\]

\[ R > C \text{ and } \beta \to \infty \Rightarrow \text{calculating } E\{Z_e(\beta|y)^{1/\beta}\} \text{ in the spin-glass phase.} \]
The Correct Decoding Exponent (Cont’d)

\[
E\{Z_e(\beta|y)^{1/\beta}\} = E \left\{ \left[ (1 - p)^n \beta \sum_{\delta} N_y(n\delta)e^{-\beta Bn\delta} \right]^{1/\beta} \right\}
\]

\[
= (1 - p)^n E \left\{ \sum_{\delta} N_{y}^{1/\beta}(n\delta)e^{-Bn\delta} \right\}
\]

\[
= (1 - p)^n \sum_{\delta} E \left\{ N_{y}^{1/\beta}(n\delta) \right\} \cdot e^{-Bn\delta}
\]

\[
E \left\{ N_{y}^{1/\beta}(n\delta) \right\} = \begin{cases} 
\exp\{n[R + h(\delta) - \ln 2]\} & \delta \leq \delta_{GV}(R) \text{ or } \delta \geq 1 - \delta_{GV}(R) \\
\exp\{n[R + h(\delta) - \ln 2]/\beta\} & \delta_{GV}(R) < \delta < 1 - \delta_{GV}(R)
\end{cases}
\]
The Correct Decoding Exponent (Cont’d)

Intuition: Below $\delta_{GV}(R)$

$$E \left\{ N_y^{1/\beta}(n\delta) \right\} \approx 0^{1/\beta} \cdot \Pr\{ N_y(n\delta) = 0 \} + 1^{1/\beta} \cdot \Pr\{ N_y(n\delta) = 1 \} \approx \exp\{n[R + h(\delta) - \ln 2]\}$$

Above $\delta_{GV}(R) \Rightarrow$ double–exponentially fast concentration:

$$E \left\{ N_y^{1/\beta}(n\delta) \right\} \approx \left[ E\{ N_y(n\delta) \} \right]^{1/\beta} \approx \left( e^{n[R + h(\delta) - \ln 2]} \right)^{1/\beta}$$

Putting into $E\{ Z_e^{1/\beta}(\beta|y) \}$ & taking the dominant $\delta$:

$$\overline{P_c} \approx \exp\{-n[R - \ln 2 - F_g]\}$$

$F_g = \text{free energy of glassy phase}$:

$$F_g = \delta_{GV}(R) \ln \frac{1}{p} + (1 - \delta_{GV}(R)) \ln \frac{1}{1-p}.$$
Correct Decoding Exponent (Cont’d)

Alternative expression: use $\ln 2 - R \equiv h(\delta_{GV}(R))$:

\[
\overline{P_c} = \exp\{-nD(\delta_{GV}(R)||p)\} = \Pr\{x_0 \text{ at distance } < \delta_{GV}(R)\}
\]

$\delta_{GV}(R)$ = typical distance of wrong codewords dominating the spin–glass phase.

Main ideas of the analysis technique:

- Summations over exponentially many codewords ⇒ summations over polynomially few terms of distance enumerators, $\{N_y(\cdot)\}$.
- Power of $\sum \equiv \sum$ of powers.
- Moments of $\{N_y(n\delta)\}$: treated differently depending on whether or not $\delta \in [\delta_{GV}(R), 1 - \delta_{GV}(R)]$. 
The Random Coding Error Exponent

Gallager’s bound:

\[
P_{e|m=0} \leq \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}_0)^{1/(1+\rho)} \left[ \sum_{m \geq 1} P(\mathbf{y}|\mathbf{x}_m)^{1/(1+\rho)} \right]^\rho
\]

\[
= \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}_0)^{1/(1+\rho)} \cdot Z_e^\rho \left( \frac{1}{1+\rho}|\mathbf{y} \right)
\]

Jensen ⇒ \( E\{Z_e^\rho (1/(1+\rho)|\mathbf{y})\} \leq [EZ_e(1/(1+\rho)|\mathbf{y})]^\rho \). Calculation in paramagnetic regime ⇒ \( E_r(R) \) is related to paramagnetic \( F' \):

\[
\bar{P}_e \leq \exp \left\{ -n \left[ \frac{\rho}{1+\rho} F_p \left( \frac{1}{1+\rho} \right) - \ln(p^{1/(1+\rho)} + (1-p)^{1/(1+\rho)}) \right] \right\}
\]
Another Application: Decoding with Erasures

Decoder with an option not to decide (erasure): Decision rule = partition into $(M + 1)$ regions:

$$y \in \mathcal{R}_0 \text{ erase}$$

$$y \in \mathcal{R}_m \ (m \geq 1) \text{ decide } x_m.$$  

Performance – tradeoff between

$$\Pr\{\mathcal{E}_1\} = \frac{1}{M} \sum_m \sum_{y \in \mathcal{R}_m^c} P(y|x_m) \text{ erasure + undetected error}$$

$$\Pr\{\mathcal{E}_2\} = \frac{1}{M} \sum_m \sum_{y \in \mathcal{R}_m} \sum_{m' \neq m} P(y|x_{m'}) \text{ undetected error}$$
Decoding with Erasures (Cont’d)

Optimum decoder: decide message $m$ iff

$$\frac{P(y|x_m)}{\sum_{m' \neq m} P(y|x_{m'})} \geq e^{nT} \quad (T \geq 0 \text{ for the erasure case}).$$

Erasure: if this holds for no $m$.

Forney’s lower bounds on err. exponents of $E_1$ and $E_2$:

$$E_1(R, T) = \max_{0 \leq s \leq \rho \leq 1} [E_0(s, \rho) - \rho R - sT] \quad \text{where}$$

$$E_0(s, \rho) = -\ln \left[ \sum_y \left( \sum_x P(x) P^{1-s}(y|x) \right) \cdot \left( \sum_{x'} P(x') P^{s/\rho}(y|x') \right)^{\rho} \right],$$

$$E_2(R, T) = E_1(R, T) + T.$$

and $P(x) = \text{random coding distribution}$. 
Decoding with Erasures (Cont’d)

Main step in [Forney68]:

\[
E \left\{ \left( \sum_{m' \neq m} P(y | X_{m'}) \right)^s \right\} \text{ upper bounded by}
\]

\[
E \left\{ \left( \sum_{m' \neq m} P(y | X_{m'})^{s/\rho} \right)^\rho \right\}, \quad \rho \geq s,
\]

and then Jensen.

Our technique: 1st expression \textit{exponentially tight, no need for }\rho.

- A simpler bound (under some symmetry condition), at least as tight.
- Sometimes (e.g., BSC): optimum \(s\) in closed form.

Also: \textit{exact} exponent (complicated) – joint work with A. Somekh–Baruch.
Two Extensions

- The REM in a uniform magnetic field and joint source–channel coding.
- The generalized REM (GREM) and hierarchical coding structures.
Earlier we modeled only interaction energies, \( \{\mathcal{E}_I(x)\} \) as \( \mathcal{N}(0, nJ^2/2) \).

When an external magnetic field \( H \) is applied

\[
\mathcal{E}(x) = \mathcal{E}_I(x) - H \cdot \sum_{i=1}^{n} x_i = \mathcal{E}_I(x) - n \cdot m(x)H
\]

where \( m(x) = \frac{1}{n} \sum_{i=1}^{n} x_i \) = magnetization of \( x \).

\[
Z(\beta, H) = \sum_{x} e^{-\beta[\mathcal{E}_I(x) - nm(x)H]}
\]

\[
= \sum_{m} \left[ \sum_{x: m(x)=m} e^{-\beta\mathcal{E}_I(x)} \right] \cdot e^{n\beta mH}
\]

\[
\Delta = \sum_{m} \zeta(\beta, m)e^{n\beta mH}
\]
The REM in a Magnetic Field (Cont’d)

\[ \zeta(\beta, m) = \sum_{x: m(x) = m} e^{-\beta \mathcal{E}_I(x)}: \text{ similar to REM with } H = 0 \text{ with only } \exp[nh((1 + m)/2)] \text{ terms.} \]

Using the same technique, we compute \( \zeta(\beta, m) = e^{n\psi(\beta, m)} \) and

\[ \phi(\beta, H) = \max_m [\psi(\beta, m) + \beta m H], \]

where \( m^* = m(\beta, H) = \text{mean (typical) magnetization.} \)
The REM in a Magnetic Field (Cont’d)

Results: Let $\beta_c(H)$ solve the equation

$$\beta^2 J^2 = 4h \left( \frac{1 + \tanh(\beta H)}{2} \right).$$

Phase transition at $\beta = \beta_c(H)$:

$$m(\beta, H) = \begin{cases} 
\tanh(\beta H) & \beta < \beta_c(H) \quad \text{paramagnetic phase} \\
\tanh(\beta_c(H) \cdot H) & \beta \geq \beta_c(H) \quad \text{spin glass phase}
\end{cases}$$

Free energy: $F = -\phi/\beta$, where:

$$\phi(\beta, H) = \begin{cases} 
\frac{\beta^2 J^2}{4} + h \left( \frac{1+\tanh(\beta H)}{2} \right) + \beta H \tanh(\beta H) & \beta < \beta_c(H) \\
\beta \left[ J \sqrt{h \left( \frac{1+\tanh(\beta_c(H)H)}{2} \right)} + H \cdot \tanh(\beta_c(H)H) \right] & \beta \geq \beta_c(H)
\end{cases}$$
$T = T_c(H)$
REM in a Magnetic Field & JSC Coding

- Binary source: \( U_1, U_2, \ldots, U_i \in \{-1, +1\} \), \( q = \text{Pr}\{U_i = 1\} \).
- Source-rate/channel-rate = \( \theta \).
- JSC code: \( u = (u_1, \ldots, u_{n\theta}) \Rightarrow x(u) \) of length \( n \).
- Random coding: Draw \( 2^{n\theta} \) binary \( n \)-vectors \( \{x(u)\} \) by fair coin tossing.

Finite-temperature decoder:

\[
\hat{u}_i = \arg\max_{u \in \{-1, +1\}} \sum_{u: u_i = u} [P(u)P(y|x(u))]^\beta, \quad i = 1, 2, \ldots, n\theta.
\]

\[
Z = \sum_u [P(u)P(y|x(u))]^\beta \\
= [P(u_0)P(y|x(u_0))]^\beta + \sum_{u \neq u_0} [P(u)P(y|x(u))]^\beta \\
\triangleq Z_c + Z_e
\]  

(1)
REM in a Magnetic Field & JSC Coding (Cont’d)

\[ P(u) = [q(1 - q)]^{n\theta/2} e^{n\theta m(u)H} \text{ where } H = \frac{1}{2} \ln \frac{q}{1-q}. \text{ Thus,} \]

\[ Z_e = [q(1 - q)]^{n\beta \theta/2} \sum_m \left[ \sum_{x(u): m(u)=m} e^{-\beta \ln[1/P(y|x(u))]} \right] e^{n\beta mH} \]

\[ = [q(1 - q)]^{n\beta \theta/2} (1 - p)^{n\beta} \sum_m \left[ \sum_{x(u): m(u)=m} e^{-\beta Bd(x(u),y)} \right] e^{n\beta \theta mH} \]

\[ \Delta = [q(1 - q)]^{n\beta \theta/2} (1 - p)^{n\beta} \sum_m \zeta(\beta,m) e^{n\beta \theta mH} \]

Statistical physics of \( Z_e \sim \text{REM in a magnetic field}. \) Similar analysis \( \Rightarrow \):

Let \( \beta_{pg}(H) \) solve:

\[ \ln 2 - h(p_\beta) = \theta h \left( \frac{1 + \tanh(\beta H)}{2} \right), \quad p_\beta \Delta \frac{p^\beta}{p^\beta + (1 - p)^\beta}. \]
Magnetization of $Z_e$ (incorrectly decoded patterns):

$$m(\beta, H) = \begin{cases} \tanh(\beta H) & \beta < \beta_{pg}(H) \\ \tanh(\beta_{pg}(H) \cdot H) & \beta \geq \beta_{pg}(H) \end{cases}$$

$Z_e \Rightarrow 3$rd phase.

The ferro–glassy boundary is $H = H_{fg}$ where

$$H_{fg} = \frac{1}{2} \ln \frac{q^*}{1 - q^*} \quad \theta h(q^*) = \ln 2 - h(p).$$
Discussion

Correct decoding for large $|H|$.

Low temp.: (sub–exponentially few) typical patterns of erroneously decoded $\{u\}$ have $m$ dictated by the frozen phase, i.e.,

$$m_g(H) = \tanh(\beta_{pg}(H) \cdot H)$$

independently of temp.

For $|H| < H_{fg}$, $\beta_{pg}(H) > 1$, means that $m$ of a typical erroneously decoded $u$ is $> m$ of a typical (correct) $u$, $m_f = \tanh(H)$.

- If $T < T_{pg}(0)$, remains true no matter small $|H|$ is.
- If $T_{pg}(0) < T < 1$, then when $|H| \downarrow$ the $m$ of (exponentially many) erroneously decoded $\{u\}$ is $m_p(\beta, H) = \tanh(\beta H)$: still $> m$ of typical $u$, but now temperature–dependent.

Analysis of $P_e$ and $P_c$ – similar as before.


\[ T = T_{pf}(H) \]

\[ T = T_{pg}(H) \]

\[ H_{fg} \]

\[ -H_{fg} \]
GREM (Derrida, ‘85) and Hierarchical Code Ensembles

Allowing correlations between \( \{\mathcal{E}_I(x)\} \) in an hierarchical (tree) structure.

Features:
- More realistic model of dependencies.
- Still (relatively) easy analysis.
- May have \( > 1 \) phase transition.
- Analogies with code ensembles with a tree structure.
\[ R_1 + R_2 = \ln 2 \]
\[ a_1 + a_2 = 1 \]
\[ \epsilon_i \sim \mathcal{N}(0, nJ^2 a_1/2) \]
\[ \epsilon_{i,j} \sim \mathcal{N}(0, nJ^2 a_2/2) \]
\[ E_{i,j} = \epsilon_i + \epsilon_{i,j} \]
\[ M_1 = e^{nR_1} \text{ branches} \]
\[ M_2 = e^{nR_2} \text{ leaves} \]

\[ M_2 = e^{nR_2} \text{ leaves} \]
\[ M_2 = e^{nR_2} \text{ leaves} \]
\[ M_2 = e^{nR_2} \text{ leaves} \]
Sketch of Analysis for GREM

As before,

\[ Z(\beta) = \sum_x e^{-\beta \mathcal{E}_I(x)} \approx \int dE \cdot N(E) e^{-\beta E} \]

Estimating \( N(E) \doteq e^{nS(E)} \) for a typical realization: \( \forall x \) with energy \( E \): 1st branch \(-\epsilon\), 2nd branch: \( E - \epsilon \).

\[ N_1(\epsilon) \doteq e^{nR_1} \cdot \exp \left\{ -\frac{\epsilon^2}{nJ^2 a_1} \right\} = \exp \left\{ n \left[ R_1 - \frac{1}{a_1} \left( \frac{\epsilon}{nJ} \right)^2 \right] \right\}, \]

"alive" for \( |\epsilon| \leq \epsilon_0 \triangleq nJ \sqrt{a_1 R_1} \). Thus,

\[ N(E) \doteq \int_{-\epsilon_0}^{+\epsilon_0} d\epsilon \cdot N_1(\epsilon) \cdot \exp \left\{ n \left[ R_2 - \frac{1}{a_2} \left( \frac{E - \epsilon}{nJ} \right)^2 \right] \right\}. \]
Sketch of Analysis for the GREM (Cont’d)

\[ S(E) = \lim_{n \to \infty} \frac{\ln N(E)}{n} = \max_{|\epsilon| \leq \epsilon_0} \left[ R_1 - \frac{1}{a_1} \left( \frac{\epsilon}{nJ} \right)^2 + R_2 - \frac{1}{a_2} \left( \frac{E - \epsilon}{nJ} \right)^2 \right] \]

\[ \phi(\beta) = \lim_{n \to \infty} \frac{1}{n} \ln \left[ \int dE \cdot e^{nS(E)} \cdot e^{-\beta E} \right] = \max_{E} [S(E) - \beta E] \]

Two cases:
If \( R_1/a_1 > R_2/a_2 \) ⇒ behavior exactly like in the REM.

Otherwise: two phase transitions at \( \beta_i = \frac{2}{J} \sqrt{\frac{R_i}{a_i}}, i = 1, 2 \):

\[ \phi(\beta) = \begin{cases} 
\ln 2 + \frac{\beta^2 J^2}{4} & \beta < \beta_1 \text{ pure paramagnetic} \\
\beta J \sqrt{a_1 R_1} + R_2 + \frac{a_2 \beta^2 J^2}{4} & \beta_1 < \beta \leq \beta_2 \text{ glassy-paramagnetic} \\
\beta J (\sqrt{a_1 R_1} + \sqrt{a_2 R_2}) & \beta > \beta_2 \text{ pure glassy} 
\end{cases} \]
GREM and Hierarchical Lossy Source Coding

- BSS $X_1, X_2, \ldots, X_i \in \{0, 1\}$ and Hamming distortion measure.
- Performance measure $E\{\exp\{-s\text{distortion}\}\}$ – related to $Z$.
- Tree structured code:
  - $n = n_1 + n_2$ and $nR = n_1 R_1 + n_2 R_2$.
  - 1st–stage code: $M_1 = e^{n_1 R_1} n_1$–vectors $\{\hat{x}_i\}$.
  - 2nd–stage code: For each $i, M_2 = e^{n_2 R_2} n_2$–vectors $\{\tilde{x}_{i,j}\}$.
  - Encode $x = (x', x'')$ by $\min\{d(x', \hat{x}_i) + d(x'', \tilde{x}_{i,j})\}$.
  - Decode 1st $n_1$ symbols using 1st $n_1 R_1$ compressed bits.
  - Overall distortion $\iff$ overall energy in GREM.
- Hierarchical ensemble:
  - Draw $M_1 n_1$–vectors $\{\hat{x}_i\}$ by fair coin tossing.
  - For each $i$, draw $M_2 n_2$–vectors $\{\tilde{x}_{i,j}\}$ by fair coin tossing.
Results

Evaluate $E\{\exp\{-s\text{distortion}\}\}$, using

$$Z(\beta|x) = \sum_{y \in C} e^{-\beta d(x,y)}$$

and then $\lim_{\theta \to \infty} E\{Z^{1/\theta}(s\theta|x)\}$.

$\Rightarrow$ calculation in the glassy regime.

For $R_1 \geq R_2$,

- $\phi(\beta) = \lim_n \frac{\ln Z}{n}$ is like in the REM:

$$\phi(\beta) = \begin{cases} R - \ln 2 - \beta + \ln(1 + e^\beta) & \beta < \beta(R) \\ -\beta \delta_{GV}(R) & \beta \geq \beta(R) \end{cases}$$

where $\beta(R) = \ln[(1 - \delta(R))/\delta(R)]$.

- $E\{\exp\{-s\text{distortion}\}\}$ like in an optimum code for $s \in (0, s_0)$ with $s_0 = \infty$ when $R_1 = R_2$. 

- [p. 40/42]
Results (Cont’d)

For $R_1 < R_2$,

- **Two phase transitions**: Defining $\lambda = \lim n_n n_1/n$ and $v(\beta, R) = \ln 2 - R + \beta - \ln(1 + e^\beta)$:

$$
\phi(\beta) = \begin{cases} 
- v(\beta, R) & \beta < \beta(R_1) \\
- \beta \lambda \delta_{GV}(R_1) - (1 - \lambda) v(\beta, R_2) & \beta(R_1) \leq \beta < \beta(R_2) \\
- \beta [\lambda \delta_{GV}(R_1) + (1 - \lambda) \delta_{GV}(R_2)] & \beta \geq \beta(R_2)
\end{cases}
$$

- $E\{\exp\{-s\text{distortion}\}\}$ behaves like in two decoupled codes in the two segments.
Conclusion

- Analogies between certain mathematical models in stat. mech. and IT.
- Inspiring alternative analysis techniques of code performance (error exponents).
- Applied to error– and correct decoding exponents in channel coding, joint source channel coding, and decoding with erasures.
- Potentially applicable to other situations, e.g., the IFC (joint work with Ordentlich and Etkin).