Statistical Physics of Random Binning

Neri Merhav

Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 3200004, Israel

ISIT 2015, Hong Kong, June 2015.
Some Background

- Slepian & Wolf (‘73) – (almost) lossless compression with SI @ decoder.
- Gallager (‘76) – random coding error exponents.
- Csiszár, Körner & Marton (‘77,’80) – universal achievability.
- Csiszár, Körner (‘81) – same with linear codes + expurgated bounds.
- Csiszár (‘82) + Oohama & Han (‘94) – coded SI.
- Kelly & Wagner (‘11) – improvements for high rates.
This Work

- Exponential error bounds for random binning.
- A statistical–mechanical perspective – finite temperature decoding.
- Phase diagram in the rate vs. temperature plane.
- Similarities and differences relative to channel coding.
- Exact random coding error exponent – phase transitions.
- Extensions: mismatch, universality, variable rates, joint coding.
Problem Setup and Preliminaries

- $\{(X_i, Y_i)\}_{i=1}^{N}$ – $N$ independent copies of $(X, Y) \sim P(x, y)$.
- $X = (X_1, \ldots, X_N)$ – source to be compressed.
- $Y = (Y_1, \ldots, Y_N)$ – side info @ decoder.
- Random binning – $e^{NR}$ bins.
- Encoder: $u = f(x)$, $f : \mathcal{X}^N \rightarrow \{1, \ldots, e^{NR}\}$.
- Inverse image (= bin) of $x$: $f^{-1}(u) = \{x : f(x) = u\}$.

Block–level MAP decoder:

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

Symbol–level MAP decoder:

$$\hat{x}_i = \arg\max_{x \in \mathcal{X}} P(x_i = x, y) = \arg\max_{x \in \mathcal{X}} \sum_{x : x_i = x} P(x, y), \quad i = 1, 2, \ldots, N.$$
Finite–Temperature Decoding (Ruján, ‘93)

\[ \hat{x}_i = \arg\max_{x \in X} \sum_{x: x_i = x} P^\beta(x, y), \quad \beta > 0. \]

Motivation:

- Common framework for both SL and BL MAP ($\beta = 1$, $\beta \to \infty$, resp.).
- Mismatch due to uncertainty (e.g., double BSS):
  - $\beta < 1$ – pessimistic decoder
  - $\beta > 1$ – optimistic decoder
- Gallager–style bounds include probabilities raised to some power.
The Finite–Temperature Posterior

Define

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

or, the Boltzmann distribution:

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

with energy function: \( \mathcal{E}(x, y) \triangleq -\ln P(x, y) \) and partition function \( Z(\beta|y, u) \).

We first study the phase diagram of the corresponding “physical system”.

- Ordinary random coding ↔ Random Energy Model (REM).
- Random binning ↔ Random Dilution Model (RDM).
Some Quick Background on the REM and RDM

In channel coding, the analogous posterior is

\[ P_\beta(x|y) = \begin{cases} \frac{P_\beta(y|x)}{\sum_{x' \in C} P_\beta(y|x')} & x \in C \\ 0 & \text{elsewhere} \end{cases} \]

which is the Boltzmann distribution with \( E(x, y) \triangleq -\ln P(y|x) \).

In random coding the \( x \)'s are drawn independently at random:

- \( E(X, y) \) are independent given \( y \).
- Analogous to the REM, which has random i.i.d. energy levels.
- The REM (and hence also \( P_\beta(\cdot|y) \)) undergoes a \( \phi \)–transition:
  - Below a critical temperature (\( \beta > \beta_c \)) – zero–entropy – glassy phase.
  - Above critical temperature – positive entropy – paramagnetic phase.
The RDM

Consider $Z(\beta) = \sum_x e^{-\beta \mathcal{E}(x)}$. $\beta = \frac{1}{kT}$ inverse temperature

The randomly diluted version is

$$Z_D(\beta) = \sum_x I(x) e^{-\beta \mathcal{E}(x)} = \sum_x e^{-\beta [\mathcal{E}(x) + \Psi(x)]}$$

where $\{I(x)\}$ are i.i.d. Bernoulli RV's with

$$\Pr\{I(x) = 1\} = \Pr\{\Psi(x) = 0\} = 1 - \Pr\{I(x) = 0\} = 1 - \Pr\{\Psi(x) = \infty\} = e^{-NR}.$$ 

The RDM also has a glassy phase transition, similar to the REM.

Relevance to random binning:

$$Z(\beta | y, u) = \sum_x I[x \in f^{-1}(u)] \cdot P^\beta(x, y).$$

However, for the correct $x$, $I[x \in f^{-1}(u)] \equiv 1$. 
Phase Diagram of $Z(\beta|y, u)$

\[ T = \frac{1}{\beta} \]

\[ T = \frac{1}{1 - \Gamma(R)} \]

paramagnetic

1

ferromagnetic

\[ T = T_c(R) \]

glassy

\[ \Gamma(\beta) = \beta H(X, Y) + \sum_y P(y) \ln \left[ \sum_x P^\beta(x, y) \right] \]

\[ \beta_c(R) = s'[s^{-1}(R)] \quad s(\epsilon) = \max \{ H_Q(X|Y) : \mathbf{E}_Q \ln[1/P(X, Y)] \leq \epsilon \}. \]
Discussion

Phase diagram $\sim$ mirror image of channel coding (Mézard & Montanari, '09).

**Reason:** SW coding at rate $R \leftrightarrow$ channel coding at rate $(H - R)$.

But there are a few non-trivial differences:

- Typical $|f^{-1}(u)|$ (RV) $\sim |\mathcal{X}|^N e^{-NR}$. Only $e^{N(H-R)}$ are in $\mathcal{T}(P)$.
  - Different from channel coding – fixed codebook size.
  - A–typical bin members may affect large deviations behavior.

- Prior – not necessarily uniform.

- Compositions of “codewords” are random.
Extensions and Variations

- Variable–rate SW coding (type–dependent rate).
- Mismatched decoding: decoding according to $\tilde{P}(x|y)$.
- Universal decoding: minimum empirical conditional entropy decoding.
- Full SW problem: separate codings & joint decoding of $(X, Y)$. 
Universal Decoding

\[ T = \frac{1}{\beta} \]

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

paramagnetic

\[ T_c = 1 \]

ferromagnetic

glassy

\[ H(X|Y) \]

\[ \ln |\mathcal{X}| \]
Full SW Coding: Encoding and Decoding Both $X$ and $Y$

$$Z(\beta|u,v) = \sum_{x,y} I[x \in f^{-1}(u)] \cdot I[y \in g^{-1}(u)] \cdot P^\beta(x,y).$$

- 4 partial partition functions, corresponding to in/correct decoding of $X$, $Y$.
- Each partial function has 3 phases ...

For $\beta \leq 1$, reliable decoding occurs if:

$$R_X > \beta H(X,Y) + \mathbb{E} \ln \left[ \sum_x P^\beta(x,Y) \right]$$

$$R_Y > \beta H(X,Y) + \mathbb{E} \ln \left[ \sum_y P^\beta(X,y) \right]$$

$$R_X + R_Y > \beta H(X,Y) + \ln \left[ \sum_{x,y} P^\beta(x,y) \right]$$
Exact Random Binning Error Exponent $E(R, \beta)$

The single-letter expression of $E(R, \beta)$ is available in the paper.