What Physics has to do with Information Theory?

Boltzmann, Gibbs (19th century):
statistical physics: macroscopic behavior from microscopic interactions

E.T. Jaynes (1957):
same principles can be derived from maximizing Shannon entropy

H. Bethe (1935), R. Gallager (1963), J. Pearl (1982):
similar ideas for exploiting sparsity in spin glasses, channel coding and AI

late 90s – unified perspective: it's all about inference
computing avg. physical properties, decoding ECCs, learning in neural networks, denoising images, reconstructing signals ...

(M. Mezard, A. Montanari, F. Krzakala, R. Urbanke, D. Saad, Y. Kabashima, H. Nishimori, M. Opper, M. Jordan, M. Wainwright, ...)

Example: decoding ECCs

$$\boldsymbol{y} \xrightarrow{Q(\boldsymbol{y}|\boldsymbol{x})} \boldsymbol{x}$$

transmitter noisy channel receiver

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{\mathcal{Z}(\boldsymbol{y})} \underbrace{\prod_i Q(y_i|x_i)}_{\text{likelihood}} \underbrace{\prod_a \mathbb{I}\left(x_{i_1^a} \oplus \cdots \oplus x_{i_p^a} = 0\right)}_{\text{prior}}$$

single instance

compute marginals $P(x_i|\boldsymbol{y})$,
using belief propagation;

assign symbol MAP estimate

$$\hat{x}_i(\boldsymbol{y}) = \arg\max P(x_i|\boldsymbol{y})$$

typical case

compute free energy $\mathbb{E}_{\boldsymbol{y}} \log \mathcal{Z}(\boldsymbol{y})$,
using replica/cavity methods;

obtain avg. estimator performance

$$\varepsilon = \mathbb{E}_{\boldsymbol{y}} \sum_i \mathbb{I}\left(\hat{x}_i(\boldsymbol{y}) \neq x_i\right)$$

What I've worked with in the recent past ...

Compressed sensing: sampling signals at sub-Nyquist rates

$$\boldsymbol{y} = F\boldsymbol{x}, \qquad \boldsymbol{y} \in \mathbb{R}^M, \boldsymbol{x} \in \mathbb{R}^N, F \in \mathbb{R}^{M\times N}$$

Solve for sparse $\boldsymbol{x}$ when $M \ll N$. Probabilistic approach:

$$P(\boldsymbol{x}|\boldsymbol{y}, F) \propto P(\boldsymbol{y}|F, \boldsymbol{x})\, P_0(\boldsymbol{x})$$

Compute marginals using *approximate message-passing* (AMP)

Been working on ...

variational approximations [arXiv:1402.1384]
converging AMP algorithms [arXiv:1406.4311]

(joint work w/ F. Krzakala, E. Tramel and L. Zdeborová)

... and what I'm working on right now

From optimal source coding ...

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & D(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{C} \end{aligned}$$

e.g. using LDPC, $\boldsymbol{x} \in \mathcal{C} \;\Leftrightarrow\; \mathbb{H}\boldsymbol{x} = 0$, and BP for optimization

... to information hiding

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & D(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} \quad & \mathbb{H}^{(\kappa)}\boldsymbol{x} = \boldsymbol{m} \end{aligned} \quad \Longleftrightarrow \quad \text{embed message } \boldsymbol{m} \text{ in image } \boldsymbol{y}$$

(joint work with Renato Vicente)