Phase Transitions for Mutual Information

Towards Capacity Achieving Codes over General Binary Channels

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

algo lma
laboratoire d'algorithmique
laboratoire de mathématiques algorithmiques

A Brief History of Coding Theory!

- Sender needs to transmit information to a receiver over a noisy channel that introduces errors.
- [Shannon, 1948]: Can transmit information almost error-free across noisy channels, as long as the rate of transmission (bits/channel use) is lesser than the information theoretic *capacity* of the channel.
- Need to introduce redundancy (error correcting codes) to combat noise.
- Shannon's result only proves the existence of "good" codes, does not provide explicit schemes.
- Classical coding theory: Use algebraic structures to construct codes having good minimum-distance properties.
- Classical coding theorist's lament:

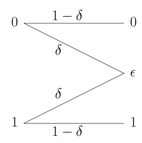
"All codes are good codes, except those we know about!"

Modern Codes - LDPC, Turbo, Raptor, ...

- Design codes based on sparse matrices:
  - Sacrifice on minimum distance, but gain because there are fewer neighbours for each codeword
  - Long codes: optimal decoding infeasible, use suboptimal belief propagation techniques
  - Suboptimality is minimal $\rightarrow$ Obtain good codes operating very close to capacity

However, only limited theoretical results known:

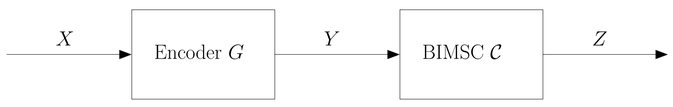
- LDPC and Raptor codes are capacity achieving over the binary erasure channel (BEC)



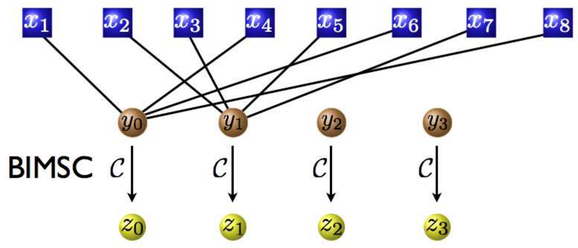
- Other practical binary channels that involve errors: very few analytical results exist!

How do modern codes perform over general binary-input memoryless symmetric channels (BIMSC)?

Our Framework



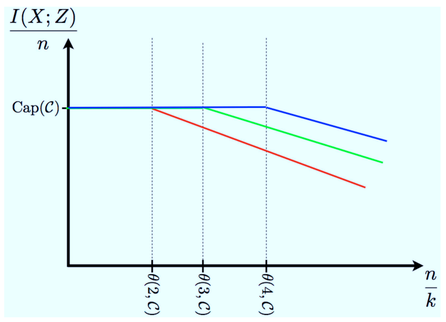
- Information symbols $X \in \mathbb{F}_2^k$ is linearly encoded into $Y \in \mathbb{F}_2^n$: $Y = XG$
- Encoding scheme: Each column of G has d ones, at locations chosen uniformly at random
  - Each element of Y is the XOR or a random d-subset of elements of X
  - Models a family of LT/Raptor-like codes



How Close are we to the Capacity $\mathrm{Cap}(\mathcal{C})$?

- How close is the mutual information $I(X;Z)$ to $n\mathrm{Cap}(\mathcal{C})$?
- **Conjecture 1** *For any BIMSC $\mathcal{C}$ and any integer d, there exists a positive real number $\theta(d,\mathcal{C})$ such that if n/k converges to a value η as $n \rightarrow \infty$, then*

$$\Pr\{I(X;Z) < n\mathrm{Cap}(\mathcal{C}) - o(n)\} \rightarrow \begin{cases} 0 \text{ if } \eta < \theta(d,\mathcal{C}) \\ 1 \text{ if } \eta > \theta(d,\mathcal{C}) \end{cases}$$



Special Case: Trivial Channel

- $\mathcal{C} = \mathcal{I}$ is the trivial (error-free) channel, i.e., $Z = Y$.
- In this case, $I(X;Z) = \mathrm{rank}(G)$. Can show that

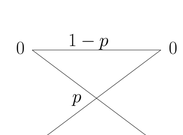
$$\Pr\{\mathrm{rank}(G) < n\} \leq \mathbb{E}\left[|\mathrm{lker}(G)|\right] - 1,$$

  where $\mathrm{lker}(G) \rightarrow$ left-kernel of G.
- Known to the mathematics community [Kolchin, 1992] that there exists a positive real $\alpha(d,\mathcal{I})$ such that if n/k converges to a value η as $n \rightarrow \infty$, then

$$\mathbb{E}\left[|\mathrm{lker}(G)|\right] \rightarrow \begin{cases} 1 \text{ if } \eta < \alpha(d,\mathcal{I}) \\ \infty \text{ if } \eta > \alpha(d,\mathcal{I}) \end{cases}$$

- Hence, $\Pr\{\mathrm{rank}(G) < n\} \rightarrow 0$ for $\eta < \alpha(d,\mathcal{I})$
- Was shown later by [Pakzad & Shokrollahi, 2007] using exit charts that Conjecture 1 holds for $\mathcal{C} = \mathcal{I}$.
- What can be said for more general channels?

The Binary Symmertic Channel $\mathcal{C} = \mathrm{BSC}(p)$



- Theorem: Let B_w denote the number of words of weight w in the right kernel of the matrix G. Then

$$\Pr[I(X;Z) < n\mathrm{Cap}(\mathcal{C})] \leq \log_2\left(\sum_{w=0}^{n} \mathbb{E}[B_w](1-2p)^{2w}\right).$$

  Proof technique: Prove a more general lower bound on the entropy of any distribution D on $\mathbb{F}_2^n$, by using the Hadamard transform and concavity arguments.
- Determine the region where the above summation vanishes, using properties of binary random matrices to analyze B_w, along the lines of [Kolchin, 1999].
- Can show: For $d \geq 3$, suppose $n, k \rightarrow \infty$ such that $n/k \rightarrow \eta$. Then $\exists$ a positive $\alpha(d,\mathcal{C})$ such that

$$\Pr\{I(X;Z) < n\mathrm{Cap}(\mathcal{C})\} \rightarrow 0 \text{ if } \eta < \alpha(d,\mathcal{C}).$$

- Numerical evaluation of $\mathrm{Cap}(\mathrm{BSC}(p))\alpha(d,\mathrm{BSC}(p))$:

$d \setminus p$	10^{-4}	10^{-3}	0.01	0.1	0.2	0.4	0.45
3	0.889	0.881	0.837	0.680	0.590	0.496	0.488
5	0.979	0.979	0.928	0.738	0.623	0.512	0.503
8	0.989	0.989	0.938	0.743	0.626	0.513	0.503
10	0.999	0.989	0.938	0.743	0.626	0.513	0.503

- Bounds tight for low p, room for improvement in other regions

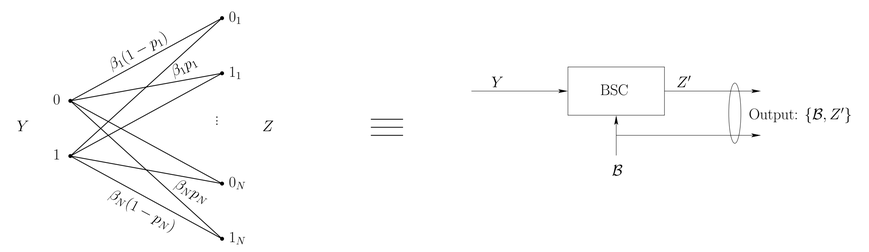
Convex Combination of BSCs

- Assuming that the all zero codeword was transmitted, the pdf of the log-likelihood ratio (LLR) of the output of BSC(p) is given by

$$\phi_p = p\Delta_{\log\frac{p}{1-p}} + (1-p)\Delta_{-\log\frac{p}{1-p}},$$

  where Δ_x denotes a Delta function at the value x.
- The pdf of LLRs of any discrete BIMSC can be written as a convex combination

$$\phi = \sum_{i=1}^{N} \beta_i \phi_{p_i}.$$



- Information theoretic arguments showing equivalence between the above two channels: allows us to extend the BSC result to arbitrary convex combinations of BSCs!

The Gaussian Channel $\mathcal{C} = \mathrm{AWGN}(\rho)$

- Binary input additive white Gaussian noise (AWGN) channel, with BPSK modulation and SNR ρ.
- AWGN channel: convex combination of infinitely many BSCs, hence may use the previous analysis (assuming that the summations converge to integrals).
- Additional result: For an LDPC ensemble with the parity-check matrix having uniformly random weight-d columns, can show that there exists an $\alpha(d,\mathcal{C})$ such that for all $n, k \rightarrow \infty$ with $n/k \rightarrow \eta$,

$$\Pr\{I(X;Z) < n\mathrm{Cap}(\mathcal{C}) - o(n)\} \rightarrow 0 \text{ if } \eta < \alpha(d,\mathcal{C}).$$

Supported by Grant 228021-ECCSciEng of the European Research Council.