Polarization and Source Coding

Erdal Arıkan

Electrical-Electronics Engineering Department,
Bilkent University, Ankara, Turkey

UCSD Shannon Memorial Lecture
11 June 2021

Outline

- Goals
  - Present basic ideas of polarization in a general framework
  - Discuss potential applications
- Scope limited to discrete memoryless sources

Transforms and entropy

- Consider a linear transforms over a field $\mathbb{F}$

$$U_k = \sum_{i=0}^{N} g_{k,i}\, X_i, \quad k = 0, 1, \ldots, N-1$$

where $\{X_k\}$ are source variables and $\{g_{k,i}\}$ are transform coefficients.

- $\{X_k\}$ are produced by a discrete memoryless source X with entropy $H(X) < \infty$.
- Our primary goal is to track the evolution of entropy from input to output of such transforms
- We will impose certain constraints on the transform to make the analysis tractable

Entropy conservation and polarization

- We will assume that the transform is invertible so that there is entropy conservation

$$H(U_0, \ldots, U_{N-1}) = H(X_0, \ldots, X_{N-1}) = NH(X)$$

- By the chain rule, the output entropy can be written as

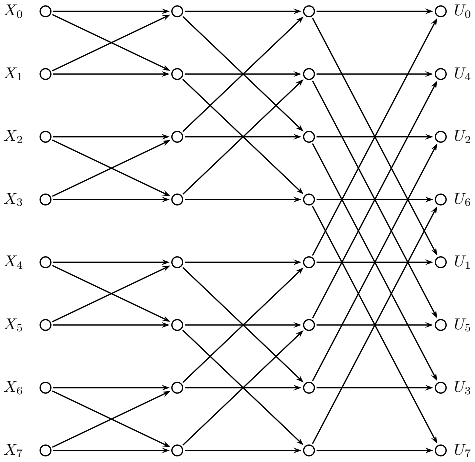
$$H(U_1, \ldots, U_{N-1}) = \sum_{i=0}^{N-1} H(U_i | U_0, \ldots, U_{i-1})$$

- We speak of entropy polarization if some of the terms $H(U_i | U_0, \ldots, U_{i-1})$ approach zero as n becomes large.

- Polarization enables compression

Butterfly transforms

- Restrict attention to butterfly transforms for low-complexity of implementation and analytical tractability
- Output read in bit-reversed order
- Examples include Discrete Fourier, Walsh-Hadamard, and Polar transforms

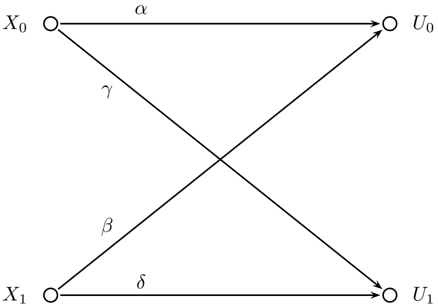


Basic butterfly operation

- Assume that each butterfly is invertible:

$$\alpha\delta - \beta\gamma \neq 0.$$

- There may be multiple types of butterflies in the overall signal flow graph but there is a tree structure



$$U_0 = \alpha X_0 + \beta X_1$$
$$U_1 = \gamma X_0 + \delta X_1$$

Entropy relations in the butterfly operation

X_0, X_1 are i.i.d. copies of X.

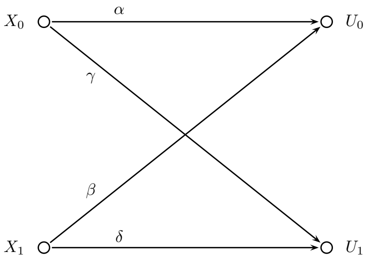
- Conservation

  $$H(U_0, U_1) = H(X_0, X_1) = 2H(X)$$

- Polarization

  $$H(U_0) \geq H(X) \geq H(U_1 | U_0)$$

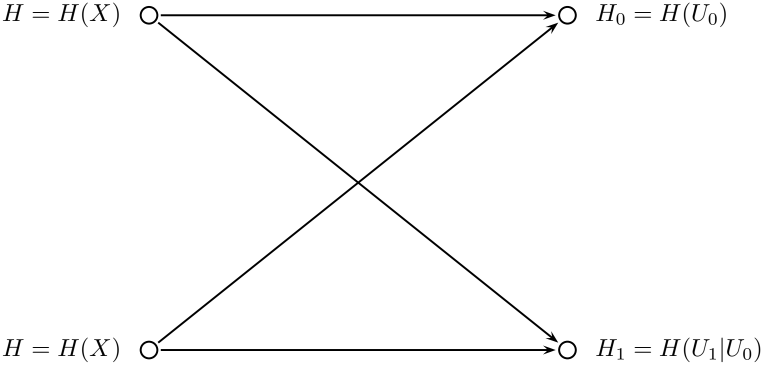
  with equality iff U_0 and U_1 are independent



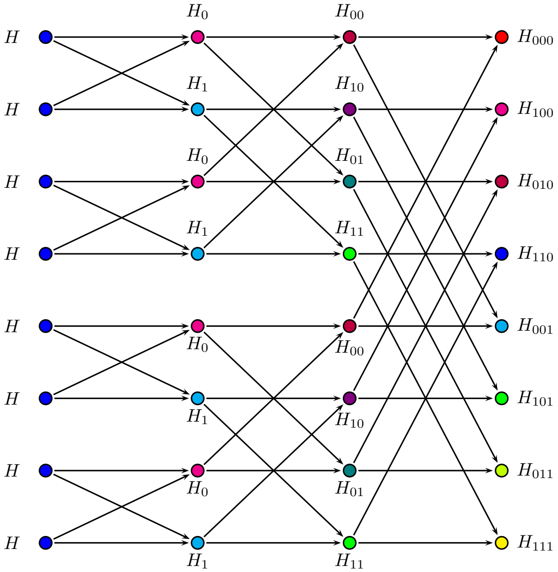
Notation

From two copies of H we obtain a pair (H_0, H_1) such that

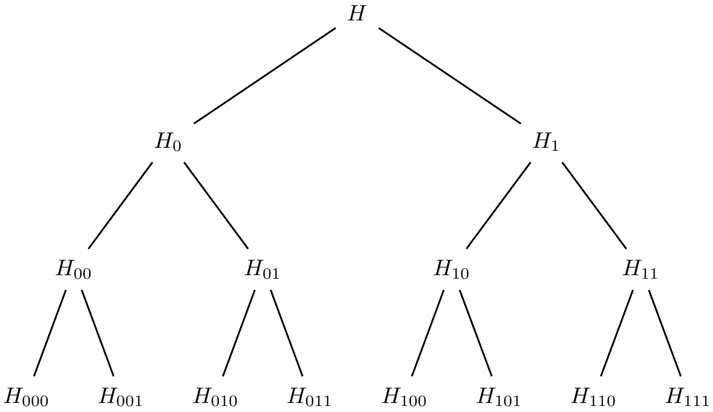
$$H_0 + H_1 = 2H, \qquad H_0 \geq H \geq H_1$$



Entropy evolution



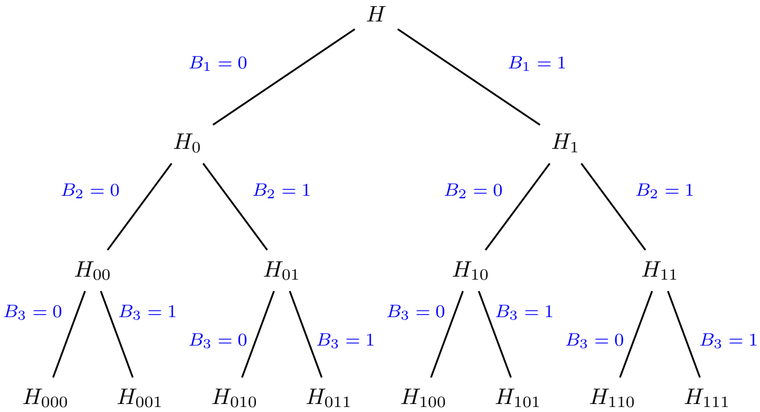
Entropy evolution tree



The same butterfly at each node, but possibly different butterflies at different nodes

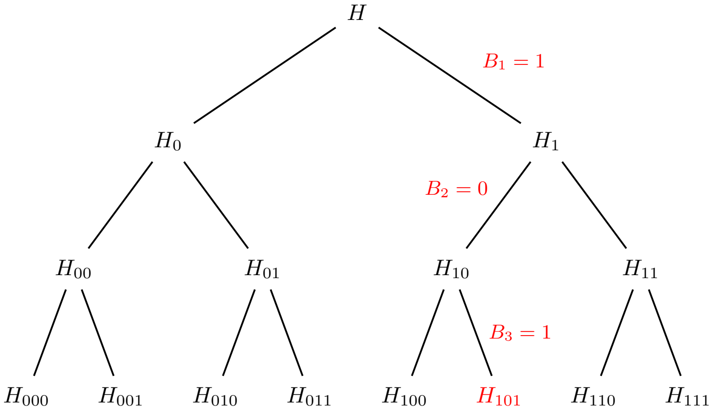
Random walk down the entropy tree

Select a path by a sequence $B_1, B_2, \ldots$ of coin flips



Random walk down the entropy tree

(B_1, B_2, B_3) selects the path to the leaf $H_{B_1 B_2 B_3}$



Entropy martingale

- Define a sequence $\{S_n; n \geq 0\}$ of entropy random variables

$$S_0 = H, \quad S_1 = H_{B_1}, \quad S_2 = H_{B_1 B_2}, \quad \cdots$$

$$S_n = H_{B_1 B_2 \cdots B_n}, \quad n \geq 1$$

- Claim: $\{S_n; n \geq 0\}$ is a martingale w.r.t. $\{B_n\}$, i.e.,

$$\mathbf{E}[S_{n+1} \mid B_1, \ldots, B_n] = S_n, \quad n \geq 0.$$

Proof.
$\mathbf{E}[S_{n+1} \mid B_1, \ldots, B_n] = \frac{1}{2} H_{B_1 \cdots B_n 0} + \frac{1}{2} H_{B_1 \cdots B_n 1} = H_{B_1 \cdots B_n} = S_n.$

Martingale convergence

- The entropy martingale $\{S_n; n \geq 0\}$ converges a.s.

$$S_n \to S_\infty \quad \text{with} \quad \mathbf{E}[S_\infty] \leq S_0$$

- $\{S_n; n \geq 0\}$ converges in $\mathcal{L}^1$, i.e.,

$$\lim_{n\to\infty} \mathbf{E}[|S_n - S_\infty|] = 0,$$

iff it is uniformly integrable, i.e.,

$$\lim_{M\to\infty} \sup_n \sum_{s \geq M} p_{S_n}(s)\, s = 0.$$

(Contribution of the tail to the expectation must die off.)

Uniform integrability examples

- Example 1. If $S_n < K$ for all n, then $\{S_n\}$ is u.i. and converges in $\mathcal{L}^1$
- Example 2. Consider the sequence of random variables

$$S_n = \begin{cases} n & \text{with prob. } 1/n \\ 0 & \text{with prob. } 1 - 1/n \end{cases}$$

This sequence is not u.i.; it converges a.s.

$$S_n \to S_\infty = 0, \quad \text{a.s.},$$

but not in $\mathcal{L}^1$

$$\mathbf{E}[S_n] = 1 \neq \mathbf{E}[S_\infty] = 0.$$

Entropy martingales for Bernoulli sources

We will consider a number of examples.

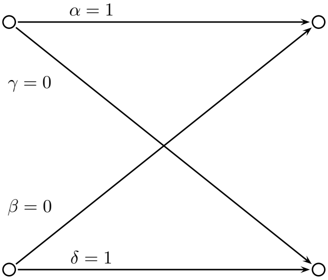
In all examples, the source X will be a memoryless Bernoulli source,

$$X = \begin{cases} 1, & \text{with prob. } p \\ 0, & \text{with prob. } 1-p \end{cases}$$

Entropy $H(X) = \mathcal{H}(p) = -p \log p - (1-p) \log(1-p)$

Identity transform

- $S_n = S_0 = \mathcal{H}(p)$ for all $n \geq 0$.
- $\{S_n\}$ converges a.s. and in $\mathcal{L}^1$
- Shows that martingale convergence does not imply polarization



$$U_0 = X_0, \quad U_1 = X_1$$

Cantor mapping

- There exists a butterfly transform of length $N = 2^n$ with

$$U_0 = 2^{-1}X_0 + 2^{-2}X_1 + 2^{-3}X_2 + \cdots + 2^{-(N-1)}X_{N-1}$$

- Clearly, $\{X_k\}$ can be recovered from just U_0, so the compression ratio is 2^n, and

$$S_n = \begin{cases} 2^n\mathcal{H}(p) & \text{with prob. } 2^{-n} \\ 0 & \text{with prob. } 1 - 2^{-n} \end{cases}$$

- Hence, $S_n \to S_\infty = 0$ a.s., but $\mathbf{E}[S_n] = \mathcal{H}(p) \neq \mathbf{E}[S_\infty] = 0$ (no $\mathcal{L}^1$ convergence).

- Shannon[1] discussed this mapping as a compression scheme

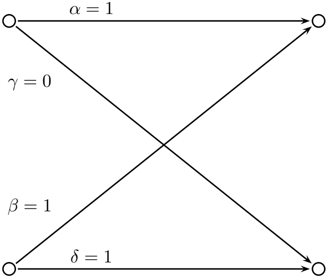
[1] C. E. Shannon, "Communication in the presence of noise," Proc. IRE, 1949.

Polar transform over $\mathbb{F}_2$

- The entropy martingale is bounded: $0 \leq S_n \leq 1$
- $S_n \to S_\infty$ a.s. and in $\mathcal{L}^1$ with

$$S_\infty = \begin{cases} 1 & \text{with prob. } \mathcal{H}(p) \\ 0 & \text{with prob. } 1 - \mathcal{H}(p) \end{cases}$$

- Achieves the Shannon limit for source coding with practical encoding/decoding algorithms



$$U_0 = X_0 + X_1 \quad \text{mod } 2,$$
$$U_1 = X_1$$

Polar source coding over $\mathbb{F}_2$

- Given source N-tuple $(x_0, \ldots, x_{N-1})$, compute its polar transform $(u_0, \ldots, u_{N-1})$
- For source coding at rate K/N, keep the transform terms u_i such that $H(U_i | U_0, \ldots, U_{i-1})$ is in the top K
- Achives Shannon's entropy bound on coding rate with a probability of error $\approx 2^{-\sqrt{N}}$ and encoding/decoding algorithms of complexity $\mathcal{O}(N \log N)$.

Polar transform over $\mathbb{R}$

- $S_n \to S_\infty$ a.s. is still guaranteed. $\mathcal{L}^1$ convergence open.
- S_n takes values over $H(U_i|U_0,\ldots,U_{i-1})$ where U_i is a sum of at most $N = 2^n$ source variables, so

$$S_n \leq H(\text{Binom}(N,p)) \approx \frac{1}{2}\log(2\pi eNp(1-p)) \approx \frac{n}{2}$$

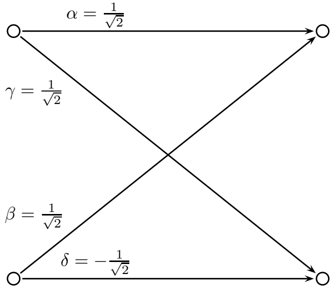
- Therefore, compression ratio cannot be larger than

$$\frac{H(\text{Binom}(N,p))}{\mathcal{H}(p)} \approx \frac{n/2}{\mathcal{H}(p)}$$

- Open problems: Is this upper bound asymptotically tight? Are there practical decoding methods?

Walsh-Hadamard transform (WHT) over $\mathbb{R}$

- $S_n \to S_\infty$ a.s.
- Open: $\mathcal{L}^1$ convergence?
- Encouraging news: Mert Pilancı[1] applied this transform successfully to compressed sensing (using "basis pursuit denoising" for recovery)



$$U_0 = \frac{1}{\sqrt{2}}(X_0 + X_1),$$
$$U_1 = \frac{1}{\sqrt{2}}(X_0 - X_1)$$

[1] M. Pilancı, Uncertain Linear Equations, M.S. Thesis, Bilkent University, 2010.

Remarks on WHT over $\mathbb{R}$

- WHT is orthonormal and conserves energy
- Each coordinate in the transform domain potentially carries $\approx n/2$ bits of information
- Recovery is limited by numerical precision problems since energy per coordinate is bounded
- Similar remarks apply to any orthonormal or unitary transform such as DFT

Possible remedy for transforms over $\mathbb{R}$

- Guido Montorsi[1] proposed using modular arithmetic over $\mathbb{R}$, such as

$$u_k = \sum_{i=0}^{N-1} g_{k,i}\, x_i \mod (M),$$

  to avoid energy growth and design low-complexity algorithms for recovering $\{x_k\}$ from a high-entropy subset of $\{u_k\}$.

- Modulus M should be small enough to provide sufficient noise margin.

- The transform viewed from $\mathbb{R}$ is now non-linear, in accordance with Shannon's recipe[2]

- In some way, it is reminiscent of lattice coding

---

[1] G. Montorsi, Polar codes over real and complex numbers, Deliverable DR.4.2 Mid-Project Report, European Union, FP7 Project 216715 NEWCOM++ (Jul. 2009).

[2] C. E. Shannon, "Communication in the presence of noise," Proc. IRE, 1949.

Concluding remarks

- Shannon's joint source-channel coding theorem implies that, N samples of a discrete source X can be carried reliably using K-dimensions in a transform domain only if

$$\frac{N}{K} < \frac{C}{H(X)}$$

  where C is the capacity per dimension (carrier).

- Polarization solves this problem when the transform is over a finite field.

- For real and complex fields, the main practical problem is finding efficient recovery methods

- Such methods may prove useful in high-order modulation schemes over $\mathbb{R}$ or $\mathbb{C}$, where $K << N$

Thank you!

Note added on June 15, 2021

After the presentation, I realized that the entropy martingale for the WHT has been studied in full detail in the PhD thesis:

S. Haghighatshoar, "Compressed Sensing of Memoryless Sources: A Deterministic Hadamard Construction," EPFL PhD Thesis, 2014.

In particular, this thesis shows that the entropy martingale for the WHT over integers does not converge in $\mathcal{L}^1$, settling some of the open problems cited in the presentation.