Can ‘finite’ be more than ‘infinite’ in distributed source coding

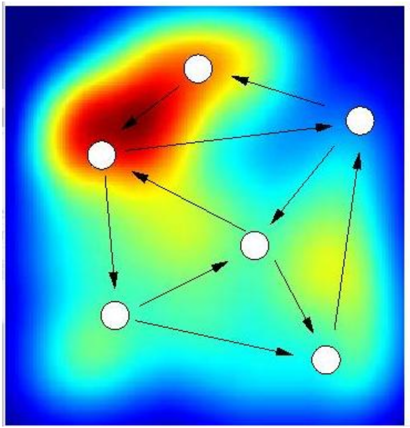
S. Sandeep Pradhan
(Joint work with Farhad Shirani)
University of Michigan, Ann Arbor

U of M 2014

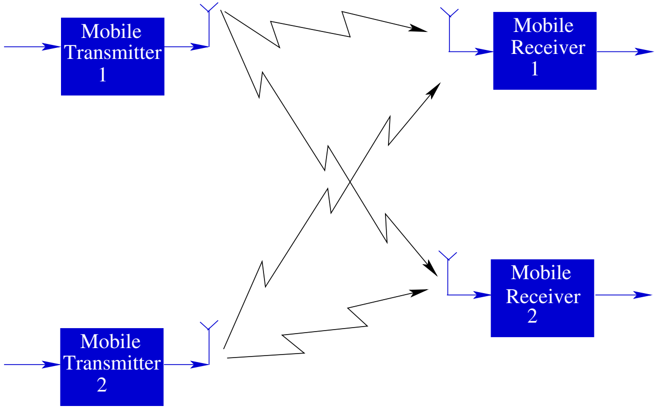
Distributed Information Coding

- Proliferation of Internet, wireless and sensor network applications
- Supported by distributed information processing
- Information-theoretic perspective

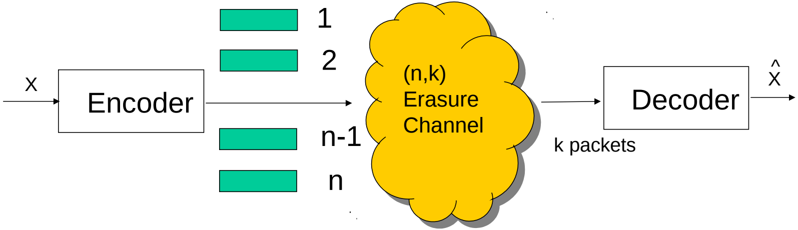
1: Distributed Field Gathering



2: Broadcast and Interference Networks



3: Streaming over the Internet



Information and Coding theory: Tradition

Information Theory:

- Develop efficient communication strategies
- No constraints on memory/computation for encoding/decoding
- Obtain performance limits that are independent of technology

Information and Coding theory: Tradition

Information Theory:

- Develop efficient communication strategies
- No constraints on memory/computation for encoding/decoding
- Obtain performance limits that are independent of technology

Coding Theory:

- Approach these limits using algebraic codes (Ex: linear codes)
- Fast encoding and decoding algorithms
- Objective: practical implementability of optimal communication systems

Point-to-point Data Compression (lossy)

Start with Binary Symmetric Source: X is IID Be(1/2)

- Wish to compress with Hamming distortion: $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and equals 0 otherwise.

Point-to-point Data Compression (lossy)

Start with Binary Symmetric Source: X is IID Be(1/2)

- Wish to compress with Hamming distortion: $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and equals 0 otherwise.
- Minimum Rate of Compression for a given distortion δ:

Point-to-point Data Compression (lossy)

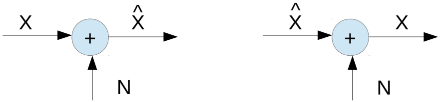
Start with Binary Symmetric Source: X is IID Be(1/2)

- Wish to compress with Hamming distortion: $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and equals 0 otherwise.
- Minimum Rate of Compression for a given distortion δ:
- $$R(D) = \min_{P(\hat{X}|X)} I(X; \hat{X}) = 1 - h(\delta)$$
- $E(d_H(X, \hat{X})) \leq \delta$.
- Single-letter relation between source and its quantized version

Point-to-point Data Compression (lossy)

Start with Binary Symmetric Source: X is IID Be(1/2)

- Wish to compress with Hamming distortion: $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and equals 0 otherwise.
- Minimum Rate of Compression for a given distortion δ:
- $$R(D) = \min_{P(\hat{X}|X)} I(X; \hat{X}) = 1 - h(\delta)$$
- $E(d_H(X, \hat{X})) \leq \delta$.
- Single-letter relation between source and its quantized version

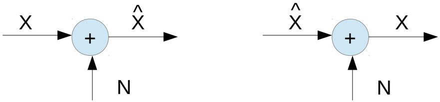


- Test Channel: $N \sim Be(\delta)$, and $+$ is addition modulo 2

Point-to-point Data Compression (lossy)

Start with Binary Symmetric Source: X is IID Be(1/2)

- Wish to compress with Hamming distortion: $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and equals 0 otherwise.
- Minimum Rate of Compression for a given distortion δ:
- $$R(D) = \min_{P(\hat{X}|X)} I(X; \hat{X}) = 1 - h(\delta)$$
- $E(d_H(X, \hat{X})) \leq \delta$.
- Single-letter relation between source and its quantized version



- Test Channel: $N \sim Be(\delta)$, and $+$ is addition modulo 2
- Achieved using Random, Infinite-dimensional-vector Quantization

Point-to-point communication: Biased Source

Binary Source with bias (BBS) p: X is IID, Be(p)

- Wish to compress with Hamming distortion

Point-to-point communication: Biased Source

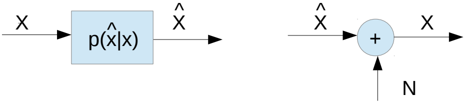
Binary Source with bias (BBS) p: X is IID, Be(p)

- Wish to compress with Hamming distortion
- Minimum Rate of Compression for a given distortion δ:
- $R(D) = h(p) - h(\delta)$.

Point-to-point communication: Biased Source

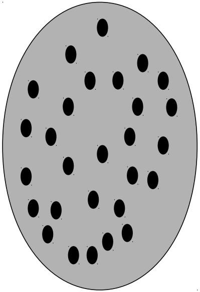
Binary Source with bias (BBS) p: X is IID, Be(p)

- Wish to compress with Hamming distortion
- Minimum Rate of Compression for a given distortion δ:
- $R(D) = h(p) - h(\delta)$.

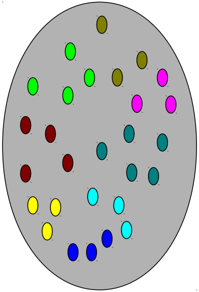


- $N \sim Be(\delta)$, and $+$ is addition modulo 2
- Achieved using Random, Infinite-dimensional-vector Quantization

Many-to-one transformation: Quantization



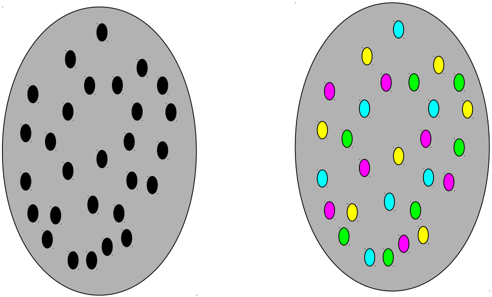
Set of all n-length sequences



3-bit quantization

- Sequences that get the same color are *NEARBY*
- $\hat{X}^n = f(X^n)$, i.e., deterministically related
- But $\hat{X}_i$ is related to X_i probabilistically: $P(\hat{X}_i|X_i)$.

Many-to-one transformation: Binning



Set of all n-length sequences   2-bit binning

- Sequences that get the same color are *FAR APART*

More is Better

- More is thought be better in information theory and coding theory
    - BOON of dimensionality: there is more space in higher dimensions

More is Better

- More is thought be better in information theory and coding theory
  - BOON of dimensionality: there is more space in higher dimensions
- Proven to be true for point-to-point communication

More is Better

- More is thought be better in information theory and coding theory
  - BOON of dimensionality: there is more space in higher dimensions
- Proven to be true for point-to-point communication
- A lot of effort in constructing codes of large block-lengths
- Even more effort in trying to encode and decode

More is Better

- More is thought be better in information theory and coding theory
  - BOON of dimensionality: there is more space in higher dimensions
- Proven to be true for point-to-point communication
- A lot of effort in constructing codes of large block-lengths
- Even more effort in trying to encode and decode
- Where else more is better?

More is Better

More is Better

More is Better

More is Better

More is Better

More is Better

More is Better

More is Better

More is Better

More is Better

Four Basic Problems in Information Theory

- Multiple access channel: Ahlswede-Liao region, 71 (wireless uplink)
- Distributed source coding: Berger-Tung region, 77 (sensor networks)

Four Basic Problems in Information Theory

- Multiple access channel: Ahlswede-Liao region, 71 (wireless uplink)
- Distributed source coding: Berger-Tung region, 77 (sensor networks)
- Broadcast channel: Marton's region, 79 (wirless downlink)
- Multiple description Coding: Zhang-Berger region, 87 (streaming)

Four Basic Problems in Information Theory

- Multiple access channel: Ahlswede-Liao region, 71 (wireless uplink)
- Distributed source coding: Berger-Tung region, 77 (sensor networks)
- Broadcast channel: Marton's region, 79 (wirless downlink)
- Multiple description Coding: Zhang-Berger region, 87 (streaming)
- Till recently we did not know whether these regions are tight or not.

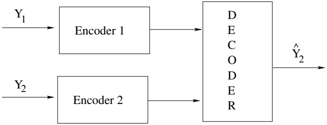
Four Basic Problems in Information Theory

- Multiple access channel: Ahlswede-Liao region, 71 (wireless uplink)
- Distributed source coding: Berger-Tung region, 77 (sensor networks)
- Broadcast channel: Marton's region, 79 (wirless downlink)
- Multiple description Coding: Zhang-Berger region, 87 (streaming)

- Till recently we did not know whether these regions are tight or not.
- Wagner et al ['11] proved that Berger-Tung region is not tight using a continuity argument.

Lossy Distributed Source Coding

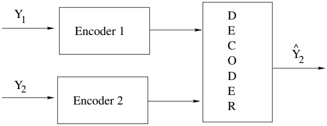
- Compression of correlated sources in a distributed setting



- Restrict to reconstruction of one source with distortion

Lossy Distributed Source Coding

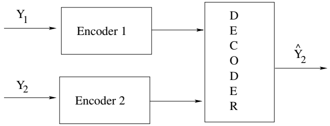
- Compression of correlated sources in a distributed setting



- Restrict to reconstruction of one source with distortion
- Window into the world of network information theory

Lossy Distributed Source Coding

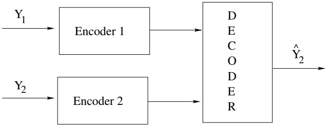
- Compression of correlated sources in a distributed setting



- Restrict to reconstruction of one source with distortion
- Window into the world of network information theory
- Single-letter Achievable Rate Distortion Region [Berger-Tung 77]
- Independent, Random (unstructured), Infinite-dimensional quantization

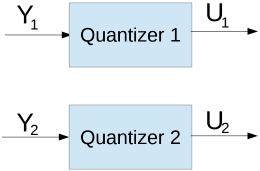
Lossy Distributed Source Coding

- Compression of correlated sources in a distributed setting

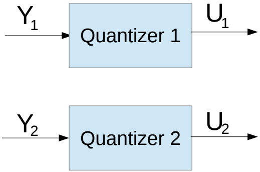


- Restrict to reconstruction of one source with distortion
- Window into the world of network information theory
- Single-letter Achievable Rate Distortion Region [Berger-Tung 77]
- Independent, Random (unstructured), Infinite-dimensional quantization
- Let U_i denote the quantized version of Y_i
- Curse: Long Markov chain: $U_1 - Y_1 - Y_2 - U_2$

Berger Tung: Double Quantization

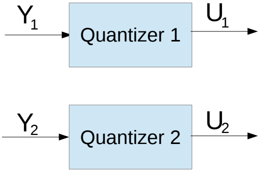


Berger Tung: Double Quantization



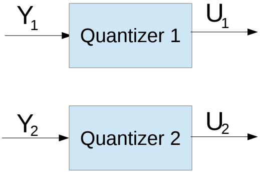
- Quantization Rate: $I(Y_1; U_1)$ and $I(Y_2; U_2)$.

Berger Tung: Double Quantization



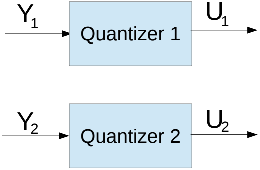
- Quantization Rate: $I(Y_1; U_1)$ and $I(Y_2; U_2)$.
- Sources are correlated $\implies$ quantized vectors are correlated

Berger Tung: Double Quantization

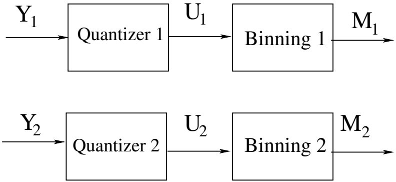


- Quantization Rate: $I(Y_1; U_1)$ and $I(Y_2; U_2)$.
- Sources are correlated $\Longrightarrow$ quantized vectors are correlated
- Bin the quantized vectors to exploit correlation
- Boon: rate rebate of $I(U_1; U_2)$.

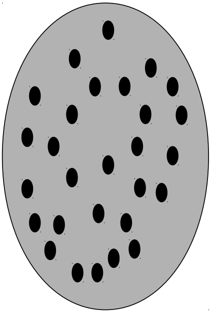
Berger Tung: Double Quantization



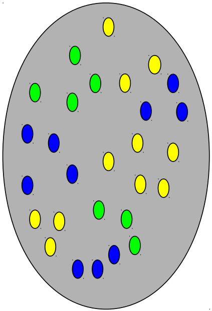
- Quantization Rate: $I(Y_1; U_1)$ and $I(Y_2; U_2)$.
- Sources are correlated $\implies$ quantized vectors are correlated
- Bin the quantized vectors to exploit correlation
- Boon: rate rebate of $I(U_1; U_2)$.



Berger Tung: Quantization + Binning

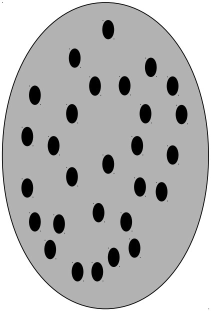


Set of all n-length sequences

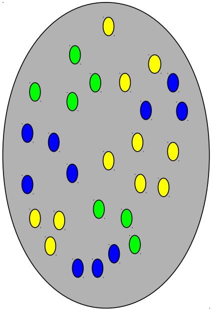


3-bit quantization+2-bit binning

Berger Tung: Quantization + Binning



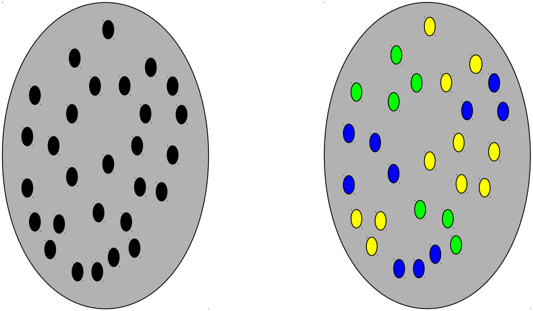
Set of all n-length sequences



3-bit quantization+2-bit binning

- Total Rate: $I(Y_1;U_1) + I(Y_2;U_2) - I(U_1;U_2) = I(Y_1,Y_2;U_1,U_2)$.

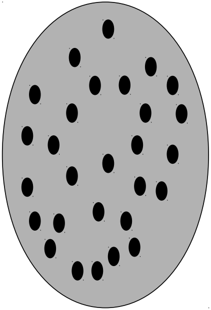
Berger Tung: Quantization + Binning



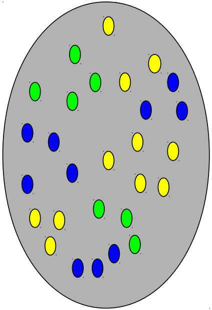
Set of all n-length sequences   3-bit quantization+2-bit binning

- Total Rate: $I(Y_1;U_1) + I(Y_2;U_2) - I(U_1;U_2) = I(Y_1,Y_2;U_1,U_2)$.
- BT: $R_1 + R_2 = I(Y_1,Y_2;U_1,U_2)$ optmized with $U_1 - Y_1 - Y_2 - U_2$

Berger Tung: Quantization + Binning



Set of all n-length sequences



3-bit quantization+2-bit binning

- Total Rate: $I(Y_1;U_1) + I(Y_2;U_2) - I(U_1;U_2) = I(Y_1,Y_2;U_1,U_2)$.
- BT: $R_1 + R_2 = I(Y_1,Y_2;U_1,U_2)$ optmized with $U_1 - Y_1 - Y_2 - U_2$
- Centralized: $R_1 + R_2 = I(Y_1,Y_2;U_1,U_2)$ optimized over everything

Quantize+Bin

- Quantize+Bin is ubiquitous in communications, signal processing

BT: Independent, Random, Infinite-dimensional quantizers

Why BT rate region is not optimal?

BT: Independent, Random, Infinite-dimensional quantizers

Why BT rate region is not optimal?

Is it because of independent and random (unstructured) quantization?

BT: Independent, Random, Infinite-dimensional quantizers

Why BT rate region is not optimal?

Is it because of independent and random (unstructured) quantization?

- What if we quantize the two sources using identical linear codes?

BT: Independent, Random, Infinite-dimensional quantizers

Why BT rate region is not optimal?

Is it because of independent and random (unstructured) quantization?

- What if we quantize the two sources using identical linear codes?
- Lemma: If $Y_1 \neq Y_2$, as $n \to \infty$, the two quantization noises become independent.

BT: Independent, Random, Infinite-dimensional quantizers

Why BT rate region is not optimal?

Is it because of independent and random (unstructured) quantization?

- What if we quantize the two sources using identical linear codes?
- Lemma: If $Y_1 \neq Y_2$, as $n \to \infty$, the two quantization noises become independent.
- So, as long as dimension (block-length) $\to \infty$, it does not matter whether the quantizers are
  - (i) independent or identical,
  - (ii) unstructured or linear.

BT: Independent, Random, Infinite-dimensional quantizers

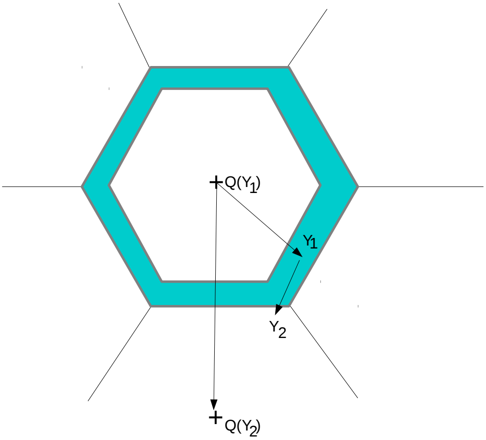
Why BT rate region is not optimal?

Is it because of independent and random (unstructured) quantization?

- What if we quantize the two sources using identical linear codes?
- Lemma: If $Y_1 \neq Y_2$, as $n \to \infty$, the two quantization noises become independent.
- So, as long as dimension (block-length) $\to \infty$, it does not matter whether the quantizers are
  - (i) independent or identical,
  - (ii) unstructured or linear.
- You cannot escape the curse with the wand of linear codes.

BT: Independent, random infinite-length quantizers

- As block-length becomes large, most volume is inside the walls
- infinitesimal perturbation will take you to the next voronoi region



Quantize and Bin: Why BT rate region is not optimal?

- Is it because of infinite block-length quantization?

Quantize and Bin: Why BT rate region is not optimal?

- Is it because of infinite block-length quantization?
- Answer: YES in the following sense.

Quantize and Bin: Why BT rate region is not optimal?

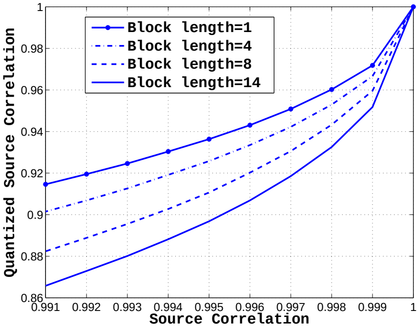
- Is it because of infinite block-length quantization?
- Answer: YES in the following sense.
- When the quantizers are identical, for a fixed distortion:
  - As block-length increases, the correlation between quantized versions becomes weaker

Quantize and Bin: Why BT rate region is not optimal?

- Is it because of infinite block-length quantization?
- Answer: YES in the following sense.
- When the quantizers are identical, for a fixed distortion:
  - As block-length increases, the correlation between quantized versions becomes weaker
  - For example: Y_1 and Y_2 jointly Gaussian, distortion $= 0.363$.

Quantize and Bin: Why BT rate region is not optimal?

- Is it because of infinite block-length quantization?
- Answer: YES in the following sense.
- When the quantizers are identical, for a fixed distortion:
  - As block-length increases, the correlation between quantized versions becomes weaker
  - For example: Y_1 and Y_2 jointly Gaussian, distortion $= 0.363$.



Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Finite-length quantizer is better than Infinite-length quantizer

Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Finite-length quantizer is better than Infinite-length quantizer

- Able to transfer the correlation from the sources to the quantized versions more efficiently,

Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Finite-length quantizer is better than Infinite-length quantizer

- Able to transfer the correlation from the sources to the quantized versions more efficiently,
- a.k.a free from the curse of long Markov chain

Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Finite-length quantizer is better than Infinite-length quantizer

- Able to transfer the correlation from the sources to the quantized versions more efficiently,
- a.k.a free from the curse of long Markov chain
- As block-length increases, the efficiency goes down, until you hit long Markov chain

Inverse Correlation Effect

- Correlation is inversely related to block-length for a fixed distortion

Finite-length quantizer is better than Infinite-length quantizer

- Able to transfer the correlation from the sources to the quantized versions more efficiently,
- a.k.a free from the curse of long Markov chain
- As block-length increases, the efficiency goes down, until you hit long Markov chain
- Short block-length: Correlation Transfer Efficiency $\uparrow$, Source Representation Efficiency $\downarrow$

Inverse Correlation Effect

Infinite-length quantizer is better than finite-length quantizer

- Large block-length: Correlation Transfer Efficiency $\downarrow$, Source Representation Efficiency $\uparrow$

Inverse Correlation Effect

Infinite-length quantizer is better than finite-length quantizer

- Large block-length: Correlation Transfer Efficiency $\downarrow$, Source Representation Efficiency $\uparrow$
- **There is a sweet-spot for the block-length where overall efficiency is maximum**

Inverse Correlation Effect

Infinite-length quantizer is better than finite-length quantizer

- Large block-length: Correlation Transfer Efficiency $\downarrow$, Source Representation Efficiency $\uparrow$
- **There is a sweet-spot for the block-length where overall efficiency is maximum**
- This is an artifact of quantize and bin strategy

Common Information (Gacs-Korner-Witsenhausen)

- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.

Common Information (Gacs-Korner-Witsenhausen)

- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.
- Example of CI: $Y_1 = (X, \tilde{Y}_1), Y_2 = (X, \tilde{Y}_2)$

Common Information (Gacs-Korner-Witsenhausen)

- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.
- Example of CI: $Y_1 = (X, \tilde{Y}_1), Y_2 = (X, \tilde{Y}_2)$
- CI-based Coding [Wagner-Kelly-Altug 09]
  - Quantize the CI at both encoders using the same code.

Common Information (Gacs-Korner-Witsenhausen)

- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.
- Example of CI: $Y_1 = (X, \tilde{Y}_1), Y_2 = (X, \tilde{Y}_2)$
- CI-based Coding [Wagner-Kelly-Altug 09]
  - Quantize the CI at both encoders using the same code.
  - Encoders cooperate at sending the quantized version.

Common Information (Gacs-Korner-Witsenhausen)

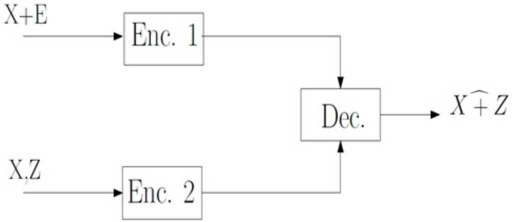
- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.
- Example of CI: $Y_1 = (X, \tilde{Y}_1), Y_2 = (X, \tilde{Y}_2)$
- CI-based Coding [Wagner-Kelly-Altug 09]
  - Quantize the CI at both encoders using the same code.
  - Encoders cooperate at sending the quantized version.
  - Treat the quantized version as side-information + BT strategy.
  - Break the long Markov chain using CI

Common Information (Gacs-Korner-Witsenhausen)

- A random variable X such that $X = f_i(Y_i)$ for $i \in \{1, 2\}$ such that $H(X) > 0$.
- Example of CI: $Y_1 = (X, \tilde{Y}_1), Y_2 = (X, \tilde{Y}_2)$
- CI-based Coding [Wagner-Kelly-Altug 09]
  - Quantize the CI at both encoders using the same code.
  - Encoders cooperate at sending the quantized version.
  - Treat the quantized version as side-information + BT strategy.
  - Break the long Markov chain using CI
  - Reduces to BT strategy when CI is trivial

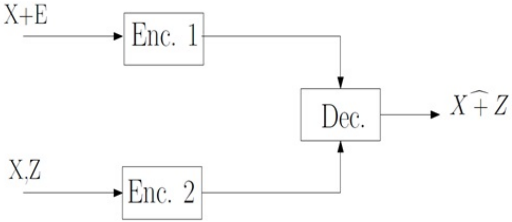
Example [Wagner-Kelly-Altug 09]

- Let $Y_1 = X + E$ and $Y_2 = (X, Z)$. Where $X \sim Be(\frac{1}{2}), E \sim Be(\epsilon), Z \sim Be(p)$.



Example [Wagner-Kelly-Altug 09]

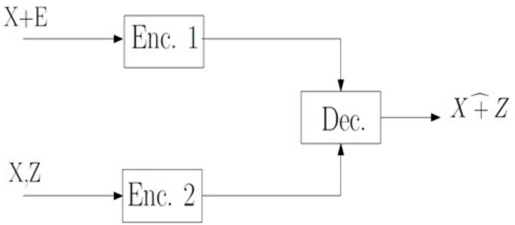
- Let $Y_1 = X + E$ and $Y_2 = (X, Z)$. Where $X \sim Be(\frac{1}{2}), E \sim Be(\epsilon), Z \sim Be(p)$.



- If $\epsilon = 0$ then X is a CI.

Example [Wagner-Kelly-Altug 09]

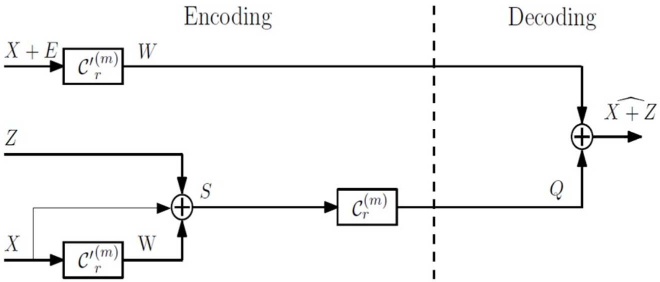
- Let $Y_1 = X + E$ and $Y_2 = (X, Z)$. Where $X \sim Be(\frac{1}{2}), E \sim Be(\epsilon), Z \sim Be(p)$.



- If $\epsilon = 0$ then X is a CI.
- The decoder wants to reconstruct $X + Z$ with distortion D.

CI is present: $\epsilon = 0$

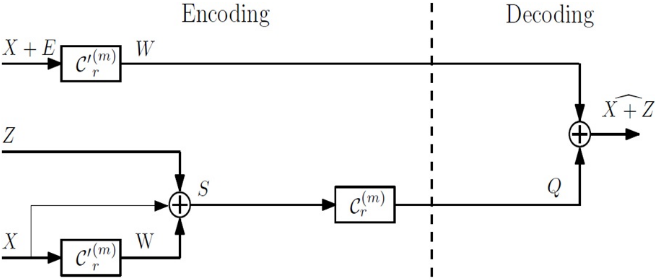
- Here is a block diagram of the CI scheme: $\epsilon = 0$



- Encoder 1: $X = W + N_\delta$

CI is present: $\epsilon = 0$

- Here is a block diagram of the CI scheme: $\epsilon = 0$

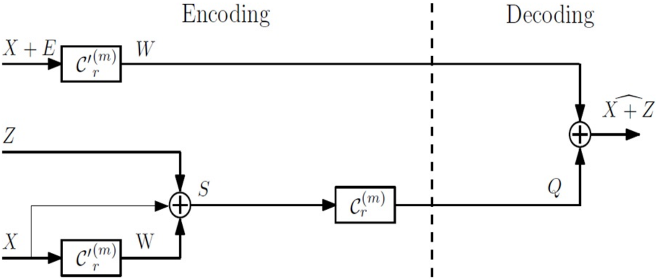


- Encoder 1: $X = W + N_\delta$

$$\Longrightarrow \quad X + Z = W + N_\delta + Z, \quad R_1 = 1 - h_b(\delta)$$

CI is present: $\epsilon = 0$

- Here is a block diagram of the CI scheme: $\epsilon = 0$



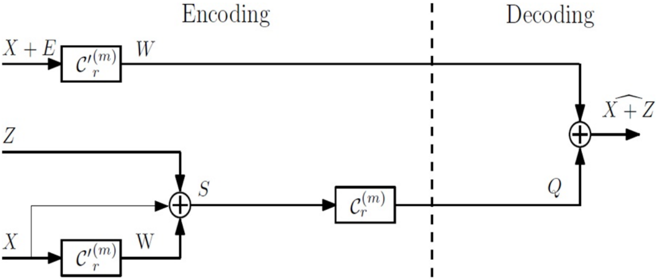
- Encoder 1: $X = W + N_\delta$

$$\implies \quad X + Z = W + N_\delta + Z, \quad R_1 = 1 - h_b(\delta)$$

- Uncertainty at decoder: $N_\delta + Z$,

CI is present: $\epsilon = 0$

- Here is a block diagram of the CI scheme: $\epsilon = 0$



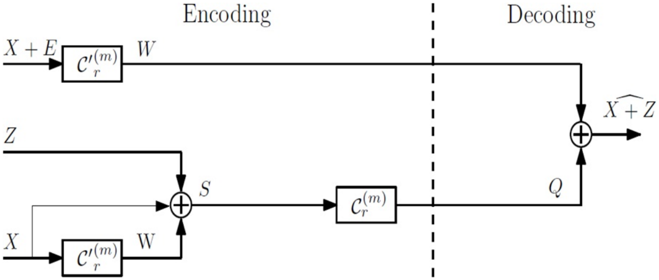
- Encoder 1: $X = W + N_\delta$

$$\implies \quad X + Z = W + N_\delta + Z, \quad R_1 = 1 - h_b(\delta)$$

- Uncertainty at decoder: $N_\delta + Z$,
- Encoder 2: $N_\delta + Z = Q + N_{\delta_1}$

CI is present: $\epsilon = 0$

- Here is a block diagram of the CI scheme: $\epsilon = 0$



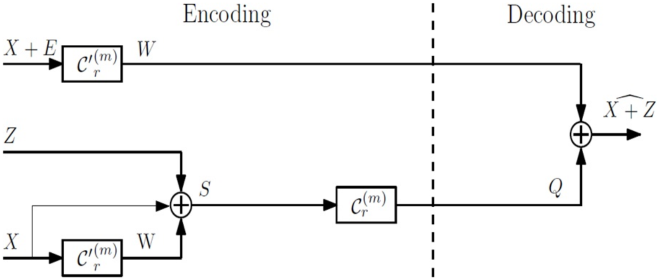
- Encoder 1: $X = W + N_\delta$

$\Longrightarrow \quad X + Z = W + N_\delta + Z, \quad R_1 = 1 - h_b(\delta)$

- Uncertainty at decoder: $N_\delta + Z$,
- Encoder 2: $N_\delta + Z = Q + N_{\delta_1} \quad \Longrightarrow \quad R_2 = h_b(\delta * p) - h_b(\delta_1)$,

CI is present: $\epsilon = 0$

- Here is a block diagram of the CI scheme: $\epsilon = 0$



- Encoder 1: $X = W + N_\delta$

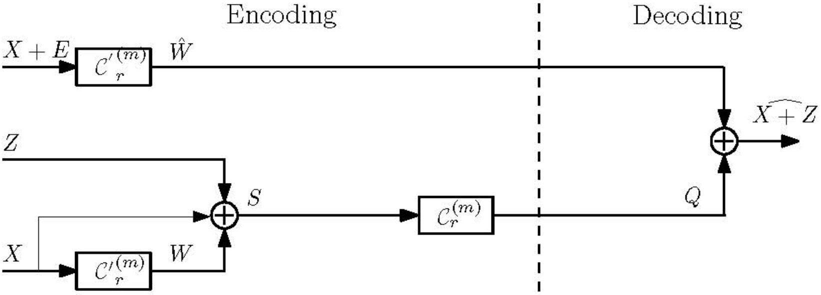
  $\implies \quad X + Z = W + N_\delta + Z, \quad R_1 = 1 - h_b(\delta)$

- Uncertainty at decoder: $N_\delta + Z$,

- Encoder 2: $N_\delta + Z = Q + N_{\delta_1} \implies \quad R_2 = h_b(\delta * p) - h_b(\delta_1)$,

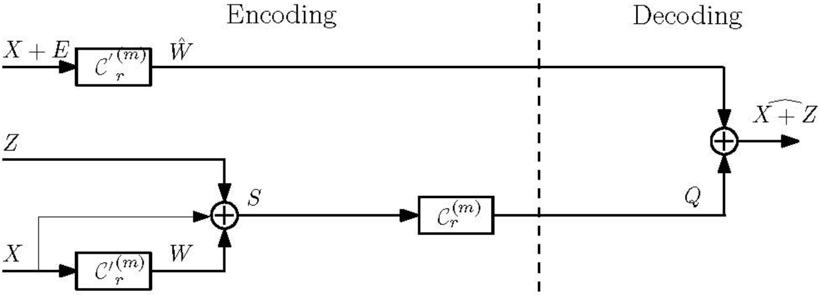
  $D = \delta_1$.

CI is absent: $\epsilon \neq 0$



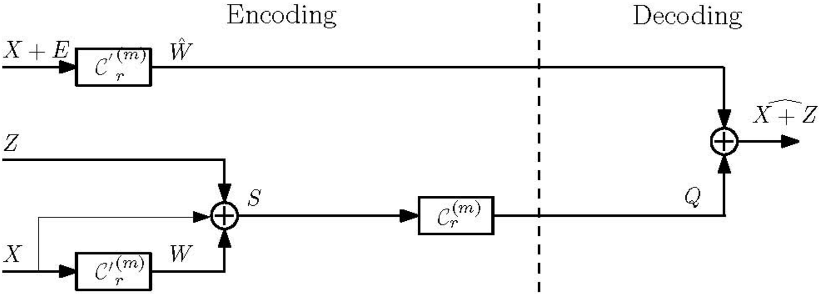
- If $\epsilon \neq 0$: Encoder 2 cannot reconstruct $\hat{W}$

CI is absent: $\epsilon \neq 0$



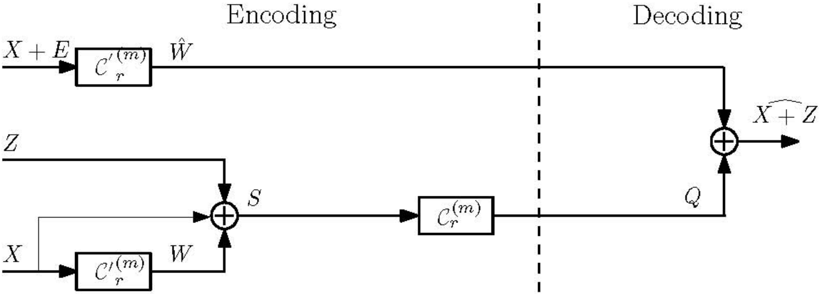
- If $\epsilon \neq 0$: Encoder 2 cannot reconstruct $\hat{W}$
- Instead can only reconstruct W: $\hat{W} - (X + E) - X - W$

CI is absent: $\epsilon \neq 0$



- If $\epsilon \neq 0$: Encoder 2 cannot reconstruct $\hat{W}$
- Instead can only reconstruct W: $\hat{W} - (X + E) - X - W$
- $\implies$ discontinuity in (R_1, R_2, D) as a function of ϵ.

CI is absent: $\epsilon \neq 0$



- If $\epsilon \neq 0$: Encoder 2 cannot reconstruct $\hat{W}$
- Instead can only reconstruct W: $\hat{W} - (X + E) - X - W$
- $\implies$ discontinuity in (R_1, R_2, D) as a function of ϵ.
- Actual rate distortion region (performance limit) is continuous in ϵ

New Result: Theorem

For the binary one-help-one problem, the following rate-distortion region is achievable for any positive integer n.

$$R_1 \geq 1 - h_b(\delta) + \theta_n \tag{1}$$

$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \tag{2}$$

$$D_2 \leq \delta_1 * \left((1-(1-\epsilon)^n)\left(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta\right)\right) \tag{3}$$

New Result: Theorem

For the binary one-help-one problem, the following rate-distortion region is achievable for any positive integer n.

$$R_1 \geq 1 - h_b(\delta) + \theta_n \tag{1}$$

$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \tag{2}$$

$$D_2 \leq \delta_1 * \left((1 - (1 - \epsilon)^n) \left(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n)} * \delta \right) \right) \tag{3}$$

- This is a single-letter characterization
- Here $\theta_n = \frac{1}{2} \frac{\log n}{n} + O(\frac{1}{n})$.

New Result: Theorem

For the binary one-help-one problem, the following rate-distortion region is achievable for any positive integer n.

$$R_1 \geq 1 - h_b(\delta) + \theta_n \quad (1)$$

$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \quad (2)$$

$$D_2 \leq \delta_1 * \left((1 - (1 - \epsilon)^n) \left(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n)} * \delta \right) \right) \quad (3)$$

- This is a single-letter characterization
- Here $\theta_n = \frac{1}{2} \frac{\log n}{n} + O(\frac{1}{n})$.
- If $n\epsilon \ll 1$ then the distortion is close to $\delta_1 * \left(n\epsilon(\delta + \frac{1}{n} * \delta) \right)$.

New Result: Theorem

For the binary one-help-one problem, the following rate-distortion region is achievable for any positive integer n.

$$R_1 \geq 1 - h_b(\delta) + \theta_n \tag{1}$$

$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \tag{2}$$

$$D_2 \leq \delta_1 * \left((1-(1-\epsilon)^n)\left(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta\right) \right) \tag{3}$$

- This is a single-letter characterization
- Here $\theta_n = \frac{1}{2}\frac{\log n}{n} + O(\frac{1}{n})$.
- If $n\epsilon \ll 1$ then the distortion is close to $\delta_1 * \left(n\epsilon(\delta + \frac{1}{n} * \delta) \right)$.
- The region is continuous in ϵ and contains the CI scheme when $\epsilon = 0$.

New Result: Theorem

For the binary one-help-one problem, the following rate-distortion region is achievable for any positive integer n.

$$R_1 \geq 1 - h_b(\delta) + \theta_n \tag{1}$$

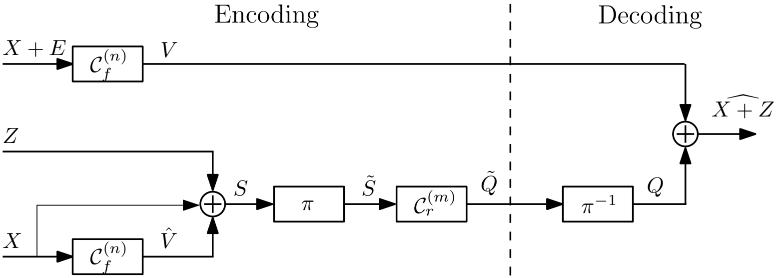
$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \tag{2}$$

$$D_2 \leq \delta_1 * \left((1-(1-\epsilon)^n)\left(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta\right)\right) \tag{3}$$

- This is a single-letter characterization
- Here $\theta_n = \frac{1}{2}\frac{\log n}{n} + O(\frac{1}{n})$.
- If $n\epsilon \ll 1$ then the distortion is close to $\delta_1 * \left(n\epsilon(\delta + \frac{1}{n} * \delta)\right)$.
- The region is continuous in ϵ and contains the CI scheme when $\epsilon = 0$.
- To get this performance via BT approach, we need multi-letterization

New Coding Approach

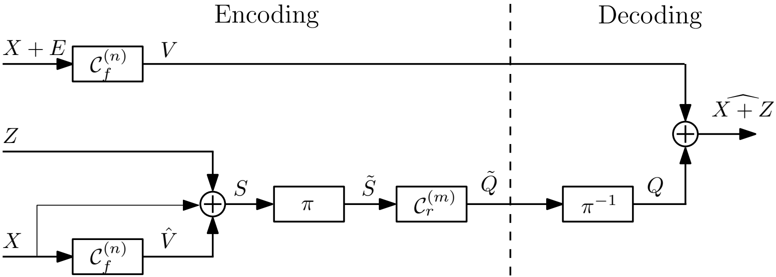
- Here is a block-diagram of the scheme:



- n is finite and m is infinitely large

New Coding Approach

- Here is a block-diagram of the scheme:



- n is finite and m is infinitely large
- 3 components: $C_f^{(n)}$, $C_r^{(m)}$ and π.

Specifics of Coding

- $C_f^{(n)}$ is an n-length code for quantizing a BSS to a distortion δ with rate $R(n, \delta) = 1 - h_b(\delta) + \theta_n$. [Kostina-Verdu 12]

Specifics of Coding

- $C_f^{(n)}$ is an n-length code for quantizing a BSS to a distortion δ with rate $R(n, \delta) = 1 - h_b(\delta) + \theta_n$. [Kostina-Verdu 12]
- $C_r^{(m)}$ is an m-length code for quantizing a BBS $\sim (p * \delta)$ to distortion δ_1, with rates approaching $h_b(p * \delta) - h_b(\delta_1)$. [Shannon 59]

Specifics of Coding

- $C_f^{(n)}$ is an n-length code for quantizing a BSS to a distortion δ with rate $R(n, \delta) = 1 - h_b(\delta) + \theta_n$. [Kostina-Verdu 12]
- $C_r^{(m)}$ is an m-length code for quantizing a BBS $\sim (p * \delta)$ to distortion δ_1, with rates approaching $h_b(p * \delta) - h_b(\delta_1)$. [Shannon 59]
- Interleaver $\pi_i \in S_n, i \in [1 : m]$

Encoders

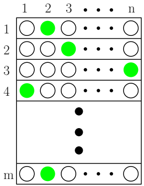
- Encoder 1:
  - Upon receiving a sequence $(X+E)(1:m,1:n)$ takes $V(i,1:n) = argmin_{c^n \in C_n^f} d_h(c^n,(X+E)(i,1:n))$ transmits the index of $V(i,1:n)$ in C_f^n.

Encoders

- Encoder 1:
  - Upon receiving a sequence $(X+E)(1:m,1:n)$ takes $V(i,1:n) = argmin_{c^n \in C_n^f} d_h(c^n, (X+E)(i,1:n))$ transmits the index of $V(i,1:n)$ in C_f^n.
- Encoder 2:
  - Upon receiving a sequence $X(1:m,1:n)$ calculates $\hat{V}(i,1:n) = argmin_{c^n \in C_n^f} d_h(c^n, X(i,1:n))$.
  - Calculates $S(i,1:n) = (\hat{V}+X+Z)(i,1:n)$.
  - Let $\tilde{S}(i,1:n) = \pi_i(S(i,1:n))$. Quantizes each $\tilde{S}(1:m,j)$ using C_r^m to get $\tilde{Q}(1:m,j)$.
  - Transmits the index of $\tilde{Q}(1:m,j)$ in C_r^m.

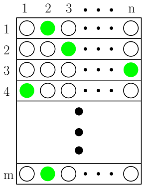
Observation

- Note that $\tilde{S}(1:m, j)$ is a DMS:



Observation

- Note that $\tilde{S}(1:m, j)$ is a DMS:



- The distribution of $\tilde{S}(1:m, j)$ is $Be(p * \delta)$:

$$
\begin{aligned}
P(\tilde{S}(i,j)=1) &= P(X(i,\pi_i(j)) + \hat{V}(i,\pi_i(j)) + Z(i,\pi_i(j)) = 1) \\
&= p * P(X(i,\pi_i(j)) + \hat{V}(i,\pi_i(j)) = 1) \\
&= p * \frac{1}{n} \sum_{j'=1}^{n} E(w_H(X(i,j') + \hat{V}(i,j'))) \\
&= p * \delta
\end{aligned}
$$

Decoder

- Decoder:
  - Calculates $Q(i, 1:n) = \pi_i^{-1}\left(\tilde{Q}(i, 1:n)\right)$.
  - Declares $(Q+V)(1:m, 1:n)$ as the reconstruction.

Decoder

- Decoder:
  - Calculates $Q(i, 1:n) = \pi_i^{-1}\left(\tilde{Q}(i, 1:n)\right)$.
  - Declares $(Q+V)(1:m, 1:n)$ as the reconstruction.
- Calculating distortion:

$$
\begin{aligned}
D &= \frac{1}{mn} E\{d_H((X+Z)(1{:}m, 1{:}n), (Q+V)(1{:}m, 1{:}n))\} \\
&= \frac{1}{mn} E\{w_H((\hat{V}+V+T)(1{:}m, 1{:}n))\} \\
&= (\delta_1 * \frac{1}{mn} E\{w_H((\hat{V}+V)(1{:}m, 1{:}n))\})
\end{aligned}
$$

Decoder

- Decoder:
  - Calculates $Q(i, 1:n) = \pi_i^{-1}\left(\tilde{Q}(i, 1:n)\right)$.
  - Declares $(Q+V)(1:m, 1:n)$ as the reconstruction.
- Calculating distortion:

$$
\begin{aligned}
D &= \frac{1}{mn}E\{d_H((X+Z)(1{:}m, 1{:}n), (Q+V)(1{:}m, 1{:}n))\} \\
&= \frac{1}{mn}E\{w_H((\hat{V}+V+T)(1{:}m, 1{:}n))\} \\
&= (\delta_1 * \frac{1}{mn}E\{w_H((\hat{V}+V)(1{:}m, 1{:}n))\})
\end{aligned}
$$

- Note that $V(i, 1:n) = \hat{V}(i, 1:n)$ if $E(1:n) = 0$.

$$
D = \left(\delta_1 * \frac{1}{n}E\{w_H((\hat{V}+V)(1, 1{:}n)|E(1, 1{:}n) \neq 0)\}P(E(1, 1{:}n) \neq 0)\right)
$$

Decoder

- Simplifying the previous equations we can get:

$$D \leq \delta_1 * (1 - (1 - \epsilon)^n)\left(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n} * \delta)\right)$$

Decoder

- Simplifying the previous equations we can get:

$$D \leq \delta_1 * (1 - (1 - \epsilon)^n)\left(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n} * \delta)\right)$$

Theorem

The new rate-distortion region strictly contains the BT rate region

Hamming Codes

- Using Hamming code of length $2^r - 1$:

$$R_1 = 1 - \frac{r}{2^r - 1}$$

$$R_2 = h_b(\frac{1}{2^r} * p) - h_b(\delta_1)$$

$$D \leq \delta_1 * (1 - (1 - \epsilon)^n)\left(\frac{\epsilon}{(1 - (1 - \epsilon)^{2^r - 1})} * \frac{1}{2^r - 1} + \frac{1}{2^r - 1}\right)$$

Hamming Codes

- Using Hamming code of length $2^r - 1$:

$$R_1 = 1 - \frac{r}{2^r - 1}$$

$$R_2 = h_b(\frac{1}{2^r} * p) - h_b(\delta_1)$$

$$D \leq \delta_1 * (1 - (1-\epsilon)^n)\left(\frac{\epsilon}{(1-(1-\epsilon)^{2^r-1})} * \frac{1}{2^r - 1} + \frac{1}{2^r - 1}\right)$$

- A good scheme for BT seems to be to time-share between the following points to avoid double quantization.

$$r_1 = 1, r_2 = h_b(p) - h_b(\delta_1)$$

$$r'_1 = 0, r'_2 = 1 - h_b(\delta_1)$$

Hamming Codes

- Using Hamming code of length $2^r - 1$:

$$R_1 = 1 - \frac{r}{2^r - 1}$$
$$R_2 = h_b(\frac{1}{2^r} * p) - h_b(\delta_1)$$
$$D \leq \delta_1 * (1 - (1 - \epsilon)^n)\left(\frac{\epsilon}{(1 - (1 - \epsilon)^{2^r - 1})} * \frac{1}{2^r - 1} + \frac{1}{2^r - 1}\right)$$

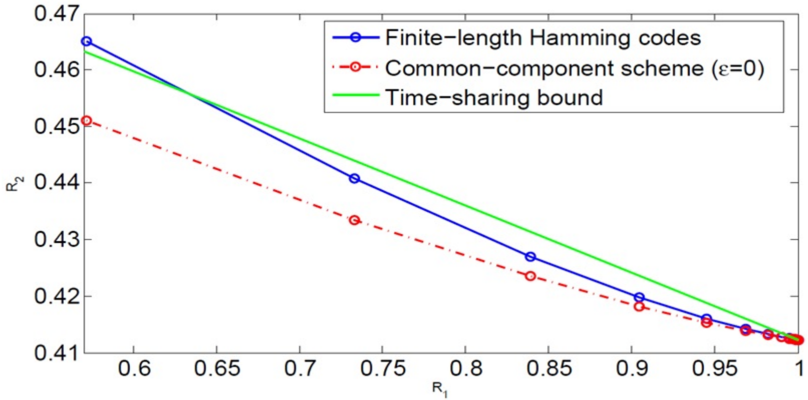
- A good scheme for BT seems to be to time-share between the following points to avoid double quantization.

$$r_1 = 1, r_2 = h_b(p) - h_b(\delta_1)$$
$$r'_1 = 0, r'_2 = 1 - h_b(\delta_1)$$

- Also the CI scheme for $\epsilon = 0$ gives an outer bound.

Numerical Results

- Comparison between the three bounds:($\delta_1 = 0.1, p = 0.3, \epsilon = 10^{-10}$)



Conclusions

- As block-length of quantization is increased
  - Correlation Transfer Efficiency decreases
  - Source Representation Efficiency increases

Conclusions

- As block-length of quantization is increased
  - Correlation Transfer Efficiency decreases
  - Source Representation Efficiency increases
- There is a sweet spot for block-length n where overall efficiency is maximum

Conclusions

- As block-length of quantization is increased
  - Correlation Transfer Efficiency decreases
  - Source Representation Efficiency increases
- There is a sweet spot for block-length n where overall efficiency is maximum
- To get this performance in BT framework, we need multi-letterization (n-letter)