New communication strategies for broadcast and interference networks

S. Sandeep Pradhan

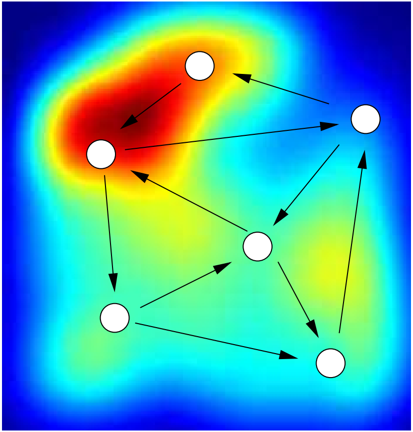
(Joint work with Arun Padakandla and Aria Sahebi)

University of Michigan, Ann Arbor

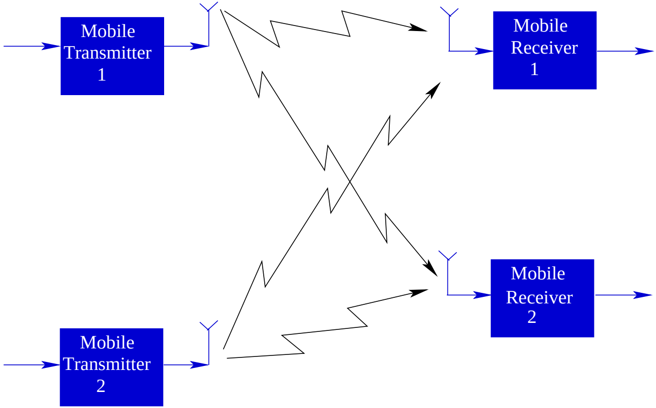
Distributed Information Coding

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

1: Distributed Field Gathering



2: Broadcast and Interference Networks



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

Information theory: Orders of magnitude

- Subatomic scale: $10^{-23} - 10^{-15}$ Physicists
- Atomic scale: $10^{-15} - 10^{-6}$ Chemists

Information theory: Orders of magnitude

- Subatomic scale: $10^{-23} - 10^{-15}$ Physicists
- Atomic scale: $10^{-15} - 10^{-6}$ Chemists
- Human scale: $10^{-6} - 10^{6}$ Biologists
- Astronomical scale: $10^{6} - 10^{27}$ Astromoners

Information theory: Orders of magnitude

- Subatomic scale: $10^{-23} - 10^{-15}$ Physicists
- Atomic scale: $10^{-15} - 10^{-6}$ Chemists
- Human scale: $10^{-6} - 10^{6}$ Biologists
- Astronomical scale: $10^{6} - 10^{27}$ Astromoners
- Information-theory scale: 10^{n}, n sufficiently large.

Probability versus Algebra

Information Theory Tools: based on probability

- Finding the optimal communication system directly is difficult

Probability versus Algebra

Information Theory Tools: based on probability

- Finding the optimal communication system directly is difficult
- Random Coding:
  - Build a collection of communication systems (ensemble)
  - Put a probability distribution on them
  - Show good average performance
  - Craft ensembles using probability

Probability versus Algebra

Information Theory Tools: based on probability

- Finding the optimal communication system directly is difficult
- Random Coding:
  - Build a collection of communication systems (ensemble)
  - Put a probability distribution on them
  - Show good average performance
  - Craft ensembles using probability

Coding Theory Tools: Abstract algebra (groups, fields)

- Exploit algebraic structure to develop algorithms of polynomial complexity for encoding/decoding
- Study a very small ensemble at a time.

Random Coding in networks

- Prob. distribution on a collection of codebooks (ensemble)
- Extensions of Shannon ensembles

Random Coding in networks

- Prob. distribution on a collection of codebooks (ensemble)
- Extensions of Shannon ensembles
- Lot of bad codebooks in the ensemble
- Average performance significantly affected by these bad codes
- Do not achieve optimality in general
- Many problems have remained open for decades.

Coding theory to the rescue ?

- It turns out that algebraic structure can be used to weed out bad codes

Coding theory to the rescue ?

- It turns out that algebraic structure can be used to weed out bad codes
- Gain barely noticeable in point-to-point communication
  - Improvement in second order performance (error exponents)

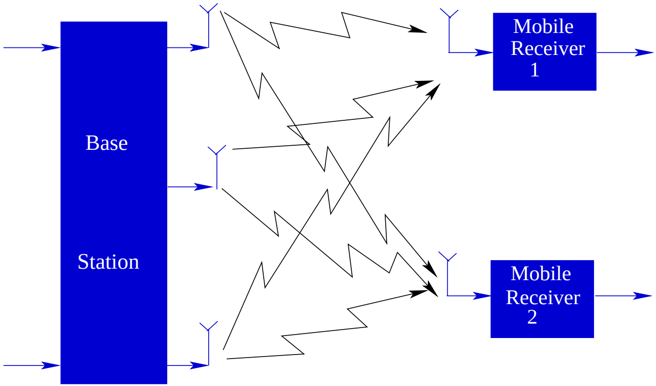
Coding theory to the rescue ?

- It turns out that algebraic structure can be used to weed out bad codes
- Gain barely noticeable in point-to-point communication
  - Improvement in second order performance (error exponents)
- Gains significant in multi-terminal communication

Coding theory to the rescue ?

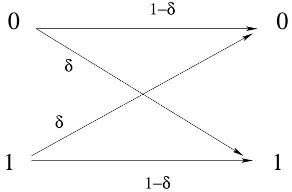
- It turns out that algebraic structure can be used to weed out bad codes
- Gain barely noticeable in point-to-point communication
  - Improvement in second order performance (error exponents)
- Gains significant in multi-terminal communication
- Time for Question?

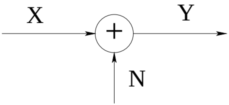
Broadcast Networks



Point-to-point communication

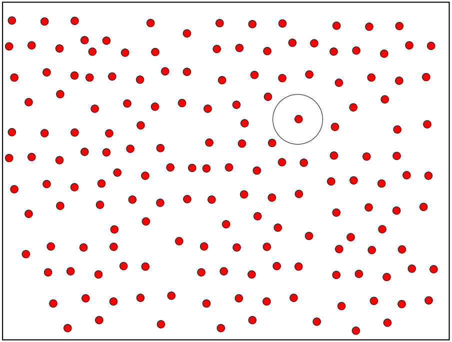
Start with Binary symmetric channel





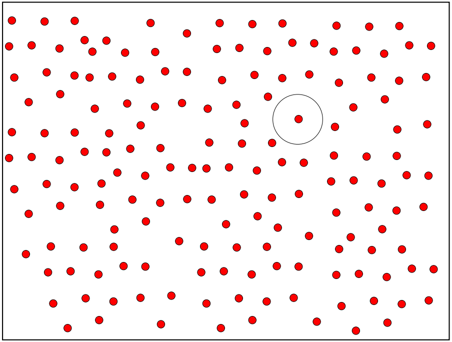
- $N \sim Be(\delta)$, and $+$ is addition modulo 2
- Capacity $= \max_{P(x)} I(X;Y) = 1 - h(\delta)$.

Picture of an optimal code



- Output is within a ball around a transmitted codeword
- Maximum likelyhood decoding

Picture of an optimal code



- Output is within a ball around a transmitted codeword
- Maximum likelyhood decoding
- Time for Question?

Twitter and Eddington Number

- Suppose you to want tweet on a BSC:
- 140 characters

Twitter and Eddington Number

- Suppose you to want tweet on a BSC:
- 140 characters
- Entropy of tweets $= 1.9$ bits/character, $\Rightarrow$ 266 bits.
- Suppose $\delta = 0.11$, then $C = 0.5$ bits/channel use

Twitter and Eddington Number

- Suppose you to want tweet on a BSC:
- 140 characters
- Entropy of tweets $= 1.9$ bits/character, $\Rightarrow$ 266 bits.
- Suppose $\delta = 0.11$, then $C = 0.5$ bits/channel use
- A tweet can be sent by using BSC 532 times.
- Number of possible tweets $= 2^{266}$

Twitter and Eddington Number

- Suppose you to want tweet on a BSC:
- 140 characters
- Entropy of tweets = 1.9 bits/character, $\Rightarrow$ 266 bits.
- Suppose $\delta = 0.11$, then $C = 0.5$ bits/channel use
- A tweet can be sent by using BSC 532 times.
- Number of possible tweets $= 2^{266}$
- Equals the number of protons in the observable universe
- Named after Arthur Eddington.

BSC with cost constaint

- $\frac{1}{n}\mathbb{E}w_H(X^n) \leq q$
- i.e., a codeword has at most q fractions of 1's

BSC with cost constaint

- $\frac{1}{n}\mathbb{E}w_H(X^n) \leq q$
- i.e., a codeword has at most q fractions of 1's
- Capacity-cost function

$$C(q) = \max_{Ew_H(X) \leq q} I(X;Y) = H(Y) - H(Y|X) = h(q * \delta) - h(\delta)$$

- $q * \delta = (1-q)\delta + q(1-\delta)$

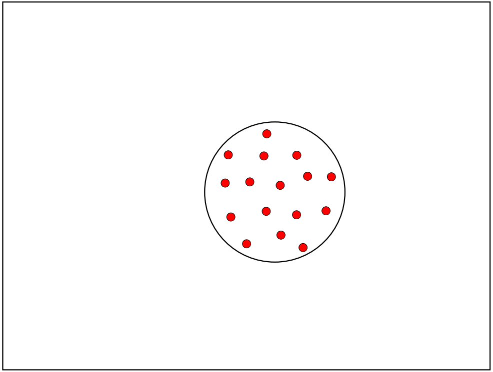
BSC with cost constaint

- $\frac{1}{n}\mathbb{E}w_H(X^n) \leq q$
- i.e., a codeword has at most q fractions of 1's
- Capacity-cost function

$$C(q) = \max_{Ew_H(X) \leq q} I(X;Y) = H(Y) - H(Y|X) = h(q * \delta) - h(\delta)$$

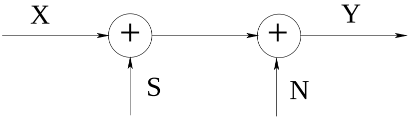
- $q * \delta = (1-q)\delta + q(1-\delta)$
- $X \sim Be(q)$

Picture of an optimal code



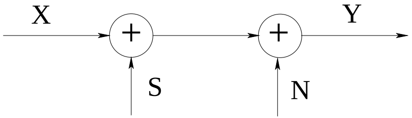
- Big circle: the set of all words with q fraction of 1's

BSC with cost constraint and interference



- $S \sim Be(0.5)$ and $N \sim Be(\delta)$
- S is non-causally observable only at encoder

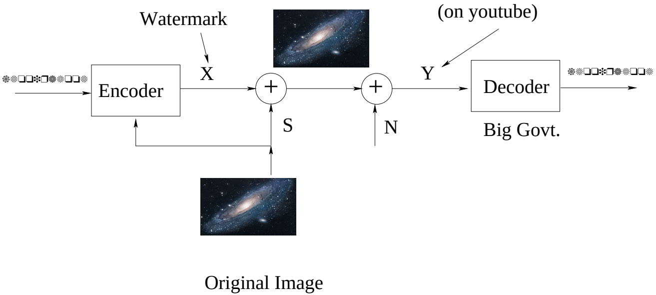
BSC with cost constraint and interference



- $S \sim Be(0.5)$ and $N \sim Be(\delta)$
- S is non-causally observable only at encoder
- $\frac{1}{n}\mathbb{E}w_H(X^n) \leq q$

Applications

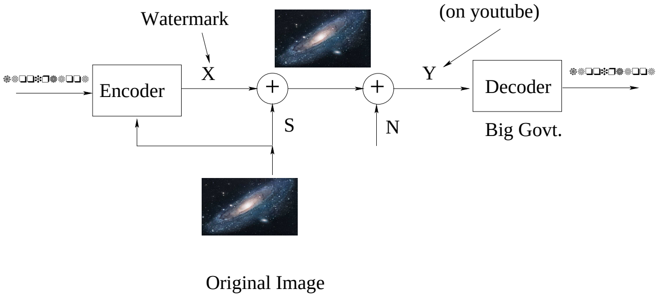
Digital watermarking, data hiding, covert communication



- Blind watermarking

Applications

Digital watermarking, data hiding, covert communication



- Blind watermarking
- You want big govt. but you dont trust it too much

BSC with cost constraint and interference

- Q1: What is the communication strategy?

BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it

BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it
  - You cannot, you do not have enough number of ones.

BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it
  - You cannot, you do not have enough number of ones.
- A2. Ride on the interference

BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it
  - You cannot, you do not have enough number of ones.
- A2. Ride on the interference
  - Nudge the interference with channel input toward a codeword
  - But, you have got just q fraction of ones.

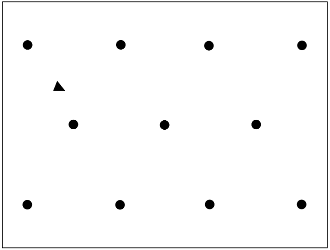
BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it
  - You cannot, you do not have enough number of ones.
- A2. Ride on the interference
  - Nudge the interference with channel input toward a codeword
  - But, you have got just q fraction of ones.
  - Gelfand-Pinsker: Nudge toward a codeword from a set

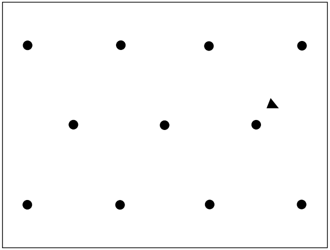
BSC with cost constraint and interference

- Q1: What is the communication strategy?
- A1. Try cancelling it
  - You cannot, you do not have enough number of ones.
- A2. Ride on the interference
  - Nudge the interference with channel input toward a codeword
  - But, you have got just q fraction of ones.
  - Gelfand-Pinsker: Nudge toward a codeword from a set
  - Q2. How large should the set be?
  - Rate of the set: $1 - h(q)$.

Picture of an optimal set of codewords

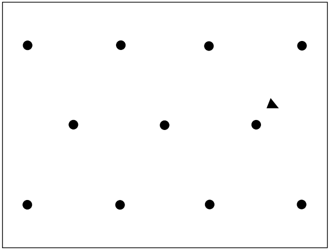


Picture of an optimal set of codewords



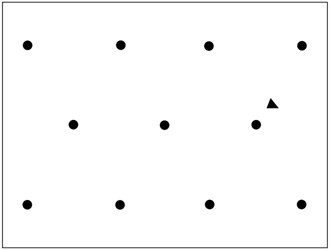
- All these codewords are assigned for a message

Picture of an optimal set of codewords



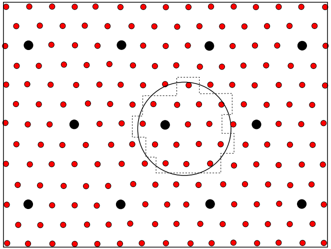
- All these codewords are assigned for a message
- Select a codeword to which you can nudge the interference..
- ..by spending just q fraction of ones $\Rightarrow U = X + S$

Picture of an optimal set of codewords



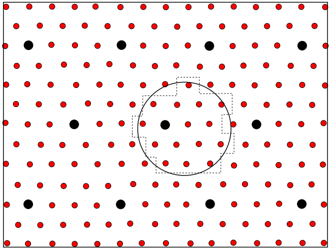
- All these codewords are assigned for a message
- Select a codeword to which you can nudge the interference..
- ..by spending just q fraction of ones $\Rightarrow U = X + S$
- New effective channel: $Y = U + N$ with capacity $1 - h(\delta)$

Precoding for Interference



- Rate of the composite codebook: $1 - h(\delta)$
- Rate of a sub-code-book: $1 - h(q)$
- Transmission rate: difference $= h(q) - h(\delta)$
- Capacity in general case [Gelfand-Pinsker '80]

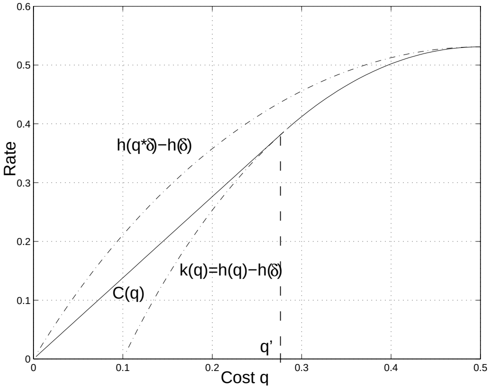
Precoding for Interference



- Rate of the composite codebook: $1 - h(\delta)$
- Rate of a sub-code-book: $1 - h(q)$
- Transmission rate: difference $= h(q) - h(\delta)$
- Capacity in general case [Gelfand-Pinsker '80]

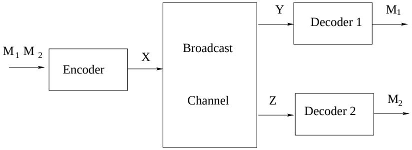
$$C(q) = \max_{P(U,X|S):Ew_H(X) \leq q} I(U;Y) - I(U;S)$$

Picture of Capacity cost function



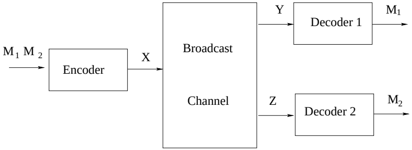
Bottomline: Rate loss as compared to no inferference

Broadcast Channel: Cover '72



- Channel with one input and multiple outputs
- Same signal should contain info. meant for both receivers
- Capacity region still not known in general

Broadcast Channel: Cover '72



- Channel with one input and multiple outputs
- Same signal should contain info. meant for both receivers
- Capacity region still not known in general
- Time for questions?

Marton's Coding Strategy: Two receivers

- Create a signal that carry information for the second receiver

Marton's Coding Strategy: Two receivers

- Create a signal that carry information for the second receiver
- This signal acts as interference for the signal of the first
- How to tackle (self) interference?

Marton's Coding Strategy: Two receivers

- Create a signal that carry information for the second receiver
- This signal acts as interference for the signal of the first
- How to tackle (self) interference?
  - Make the first receiver decode a large portion of interference
  - This portion is given by a (univariate) function

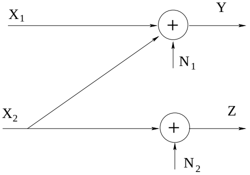
Marton's Coding Strategy: Two receivers

- Create a signal that carry information for the second receiver
- This signal acts as interference for the signal of the first
- How to tackle (self) interference?
  - Make the first receiver decode a large portion of interference
  - This portion is given by a (univariate) function
  - The rest is precoded for using Gelfand-Pinsker strategy

Marton's Coding Strategy: Two receivers

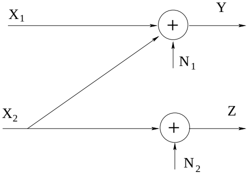
- Create a signal that carry information for the second receiver
- This signal acts as interference for the signal of the first
- How to tackle (self) interference?
  - Make the first receiver decode a large portion of interference
  - This portion is given by a (univariate) function
  - The rest is precoded for using Gelfand-Pinsker strategy
- This strategy is optimal for many special cases
- We do not know whether it is optimal in general

Example: so-called non-degraded channel



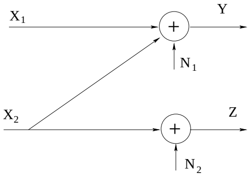
- $N_1 \sim Be(\delta)$, and $N_2 \sim Be(\epsilon)$, and no constraint on X_2
- Hamming weight constraint on X_1: $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$

Example: so-called non-degraded channel



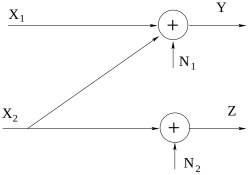
- $N_1 \sim Be(\delta)$, and $N_2 \sim Be(\epsilon)$, and no constraint on X_2
- Hamming weight constraint on X_1: $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$
- Fix $R_2 = 1 - h(\epsilon)$, and assume $\delta < \epsilon$

Example: so-called non-degraded channel



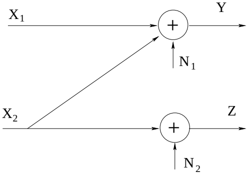
- $N_1 \sim Be(\delta)$, and $N_2 \sim Be(\epsilon)$, and no constraint on X_2
- Hamming weight constraint on X_1: $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$
- Fix $R_2 = 1 - h(\epsilon)$, and assume $\delta < \epsilon$
- When $q * \delta \leq \epsilon$, Rec. 1 can decode interference completely

Example: so-called non-degraded channel



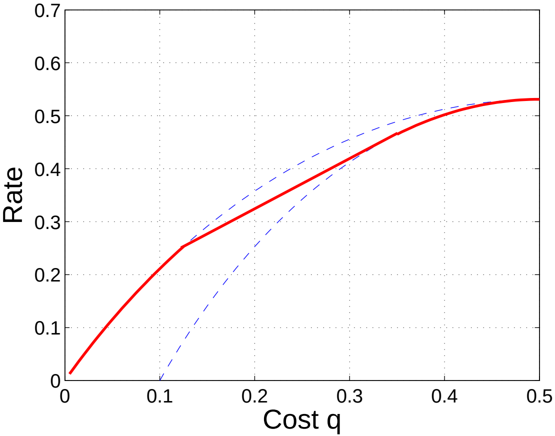
- $N_1 \sim Be(\delta)$, and $N_2 \sim Be(\epsilon)$, and no constraint on X_2
- Hamming weight constraint on X_1: $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$
- Fix $R_2 = 1 - h(\epsilon)$, and assume $\delta < \epsilon$
- When $q * \delta \leq \epsilon$, Rec. 1 can decode interference completely
  - a.k.a no interference $\Rightarrow R_1 = h(q * \delta) - h(\delta)$

Example: so-called non-degraded channel



- $N_1 \sim Be(\delta)$, and $N_2 \sim Be(\epsilon)$, and no constraint on X_2
- Hamming weight constraint on X_1: $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$
- Fix $R_2 = 1 - h(\epsilon)$, and assume $\delta < \epsilon$
- When $q * \delta \leq \epsilon$, Rec. 1 can decode interference completely
    - a.k.a no interference $\Rightarrow R_1 = h(q * \delta) - h(\delta)$
- Otherwise, precode for X_2: $\Rightarrow R_1 = h(q) - h(\delta)$

Picture of Rate Region



Decode a *univariate* function of interference & precode for the rest

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:
  - (self) interference of signals of Rec. 2 and Rec. 3

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:
  - (self) interference of signals of Rec. 2 and Rec. 3
  - Decode a univariate function of signal meant for Rec. 2...
  - .. and a univariate function of signal meant for Rec. 3.

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:
  - (self) interference of signals of Rec. 2 and Rec. 3
  - Decode a univariate function of signal meant for Rec. 2...
  - .. and a univariate function of signal meant for Rec. 3.
  - Precode for the rest

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:
  - (self) interference of signals of Rec. 2 and Rec. 3
  - Decode a univariate function of signal meant for Rec. 2...
  - .. and a univariate function of signal meant for Rec. 3.
  - Precode for the rest
- All these being done using random codes
- No need for linear or algebraic codes till now

Broadcast with more receivers

- Marton's strategy can be easily extended
- Consider 3 receiver case: At receiver 1:
  - (self) interference of signals of Rec. 2 and Rec. 3
  - Decode a univariate function of signal meant for Rec. 2...
  - .. and a univariate function of signal meant for Rec. 3.
  - Precode for the rest
- All these being done using random codes
- No need for linear or algebraic codes till now
- We can show that such a strategy is strictly suboptimal

New Strategy

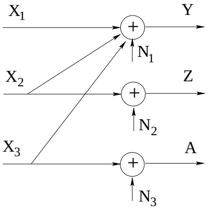
- Decode a bivariate function of the signals meant for other two

New Strategy

- Decode a bivariate function of the signals meant for other two
- It turns out that to exploit this we need linear codes

New Strategy

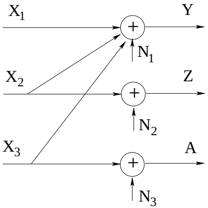
- Decode a bivariate function of the signals meant for other two
- It turns out that to exploit this we need linear codes



- $N_2, N_3 \sim Be(\epsilon)$, and no constraints on X_2 and X_3
- $N_1 \sim Be(\delta)$ and the usual : $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$

New Strategy

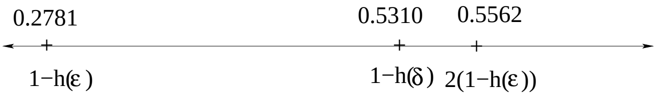
- Decode a bivariate function of the signals meant for other two
- It turns out that to exploit this we need linear codes



- $N_2, N_3 \sim Be(\epsilon)$, and no constraints on X_2 and X_3
- $N_1 \sim Be(\delta)$ and the usual : $\frac{1}{n}\mathbb{E}w_H(X_1^n) \leq q$
- Let $R_2 = R_3 = 1 - h(\epsilon)$, the incorrigble brutes!
- Let $\delta < \epsilon$

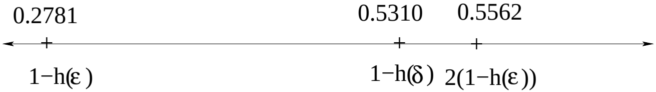
Deficiency of random codes

- $\delta = 0.1$ and $\epsilon = 0.2$



Deficiency of random codes

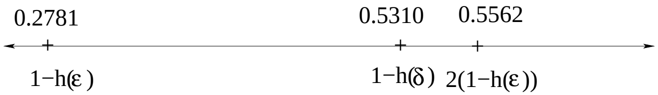
- $\delta = 0.1$ and $\epsilon = 0.2$



- Marton wishes to decode "full" interference: (X_2, X_3):
  - $1 - h(q * \delta) > 2(1 - h(\epsilon))$

Deficiency of random codes

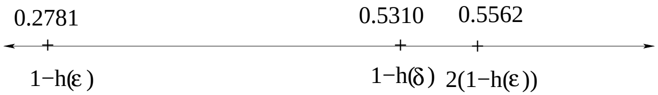
- $\delta = 0.1$ and $\epsilon = 0.2$



- Marton wishes to decode "full" interference: (X_2, X_3):
  - $1 - h(q * \delta) > 2(1 - h(\epsilon))$
  - a.k.a never going to happen
  - Marton ends up doing precoding incurring rate loss

Deficiency of random codes

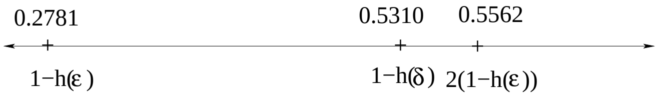
- $\delta = 0.1$ and $\epsilon = 0.2$



- Marton wishes to decode "full" interference: (X_2, X_3):
  - $1 - h(q * \delta) > 2(1 - h(\epsilon))$
  - a.k.a never going to happen
  - Marton ends up doing precoding incurring rate loss
- New Approach: Try decoding *actual* interference: $X_2 + X_3$

Deficiency of random codes

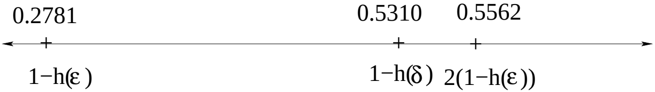
- $\delta = 0.1$ and $\epsilon = 0.2$



- Marton wishes to decode "full" interference: (X_2, X_3):
  - $1 - h(q * \delta) > 2(1 - h(\epsilon))$
  - a.k.a never going to happen
  - Marton ends up doing precoding incurring rate loss
- New Approach: Try decoding *actual* interference: $X_2 + X_3$
  - Benefit if the range of $X_2 + X_3$ is $\lll$ range of (X_2, X_3)

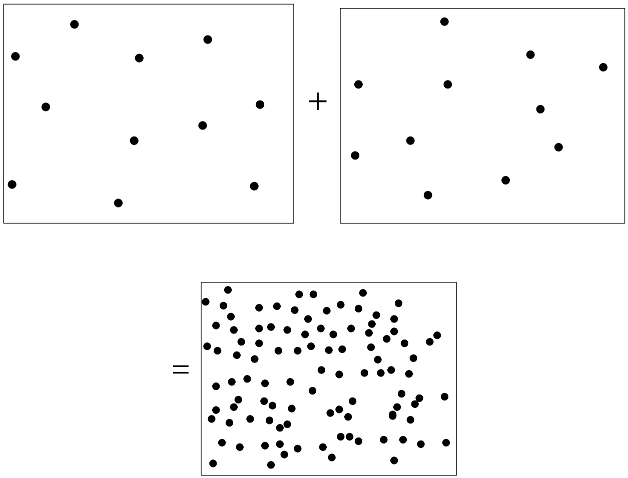
Deficiency of random codes

- $\delta = 0.1$ and $\epsilon = 0.2$

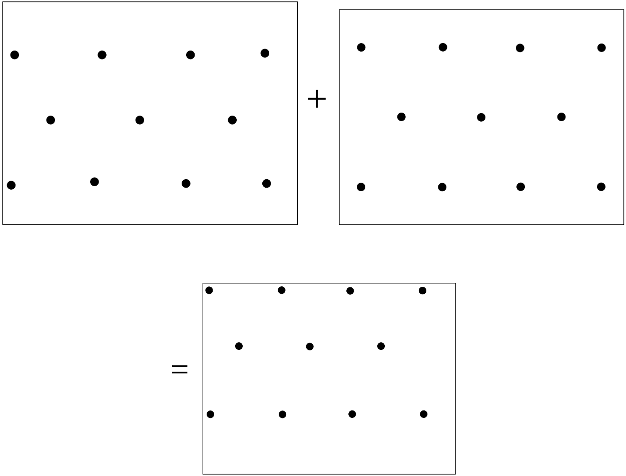


- Marton wishes to decode "full" interference: (X_2, X_3):
  - $1 - h(q * \delta) > 2(1 - h(\epsilon))$
  - a.k.a never going to happen
  - Marton ends up doing precoding incurring rate loss
- New Approach: Try decoding *actual* interference: $X_2 + X_3$
  - Benefit if the range of $X_2 + X_3$ is $\lll$ range of (X_2, X_3)
  - If X_2 and X_3 are "random", this wont happen

Picture of sum of two random sets



Picture of sum of two cosets of a linear code



Exploits of Linear Codes

- The "incorrigible brutes" can have their capacities

Exploits of Linear Codes

- The "incorrigible brutes" can have their capacities
- We just need their codebooks to behave "algebraic"
- We know that linear codes achieve the capacity of BSC

Exploits of Linear Codes

- The "incorrigible brutes" can have their capacities
- We just need their codebooks to behave "algebraic"
- We know that linear codes achieve the capacity of BSC
- rate of X_2 = rate of X_3 = rate of $X_2 + X_3 = 1 - h(\epsilon)$
- Since $\delta < \epsilon$, we have for small q: $q * \delta < \epsilon$

Exploits of Linear Codes

- The "incorrigible brutes" can have their capacities
- We just need their codebooks to behave "algebraic"
- We know that linear codes achieve the capacity of BSC
- rate of X_2 = rate of X_3 = rate of $X_2 + X_3 = 1 - h(\epsilon)$
- Since $\delta < \epsilon$, we have for small q: $q * \delta < \epsilon$
- Hence $1 - h(q * \delta) > 1 - h(\epsilon)$
- Rec. 1 can decode the actual interference and subtract it off

Exploits of Linear Codes

- The "incorrigible brutes" can have their capacities
- We just need their codebooks to behave "algebraic"
- We know that linear codes achieve the capacity of BSC
- rate of X_2 = rate of X_3 = rate of $X_2 + X_3 = 1 - h(\epsilon)$
- Since $\delta < \epsilon$, we have for small q: $q * \delta < \epsilon$
- Hence $1 - h(q * \delta) > 1 - h(\epsilon)$
- Rec. 1 can decode the actual interference and subtract it off
- Then decodes her message at rate $h(q * \delta) - h(\delta)$
- $R_1 = h(q * \delta) - h(\delta)$, $R_2 = R_3 = 1 - h(\epsilon)$

Symmetry and addition saved the world

We have banked on

Symmetry and addition saved the world

We have banked on

- Channels of Rec. 2 and 3 are symmetric
  - so uniform input distribution achieves capacity

Symmetry and addition saved the world

We have banked on

- Channels of Rec. 2 and 3 are symmetric
  - so uniform input distribution achieves capacity
- Interference in the broadcast channel is additive

Symmetry and addition saved the world

We have banked on

- Channels of Rec. 2 and 3 are symmetric
  - so uniform input distribution achieves capacity
- Interference in the broadcast channel is additive

But Shannon theory is all about not getting bogged down in an example

- Objective is to develop a theory for general case

However?

- Caution: Even in point-to-point communication
  - In general, linear codes do not achieve Shannon capacity of an arbitrary discrete memoryless channel

However?

- Caution: Even in point-to-point communication
    - In general, linear codes do not achieve Shannon capacity of an arbitrary discrete memoryless channel
- What hope do we have in using them for network communication for the arbitrary discrete memoryless case?

Thesis

- Algebraic structure in codes may be necessary in a fundamental way

Thesis

- Algebraic structure in codes may be necessary in a fundamental way
- Algebraic structure alone is not sufficient
- A right mix of algebraic structure along with non-linearity
- Nested algebraic code appears to be a universal structure

Noisy Channel Coding in Point-to-point case

- Given: Channel I/P$= X$, O/P$=Y$, with $p_{Y|X}$, and cost function $w(x)$
- Find: maximum transmission rate R for a target cost W.

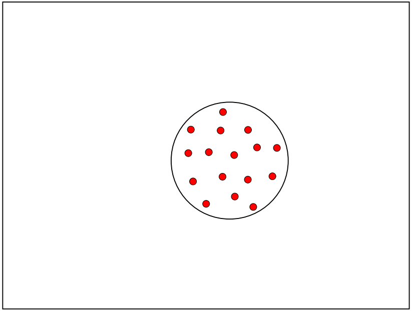
Noisy Channel Coding in Point-to-point case

- Given: Channel I/P= X, O/P=Y, with $p_{Y|X}$, and cost function $w(x)$
- Find: maximum transmission rate R for a target cost W.
- Answer: Shannon Capacity-Cost function (Shannon '49)

$$C(W) = \max_{p_X : Ew \leq W} I(X;Y)$$

Picture of a near-optimal channel code

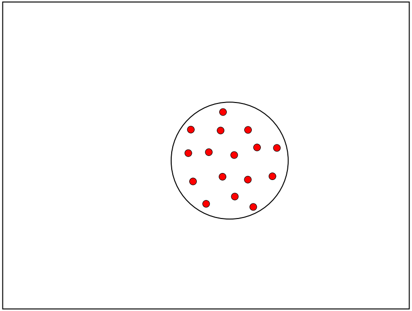
Obtained from Shannon ensemble



- Box $= \mathcal{X}^n$
- Red dot $=$ codeword
- $\mathcal{C}$ $=$ code book

Picture of a near-optimal channel code

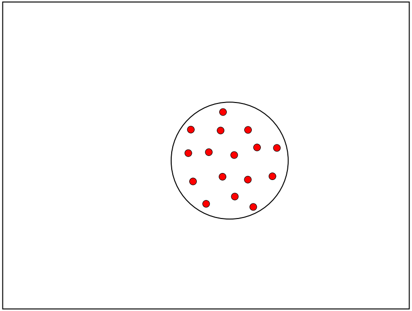
Obtained from Shannon ensemble



- Box $= \mathcal{X}^n$
- Red dot $=$ codeword
- $\mathcal{C}$ $=$ code book
- $\mathcal{C}$ has Packing Property
- $\mathcal{C}$ has Shaping Property

Picture of a near-optimal channel code

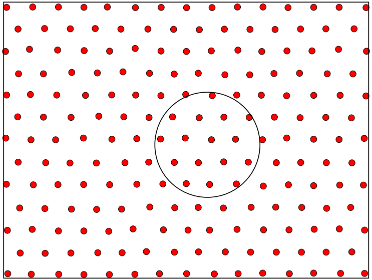
Obtained from Shannon ensemble



- Box $= \mathcal{X}^n$
- Red dot $=$ codeword
- $\mathcal{C}$ $=$ code book
- $\mathcal{C}$ has Packing Property
- $\mathcal{C}$ has Shaping Property
- Shape Region $=$ Typical set
- Size of code $= I(X;Y)$
- Codeword density $=$

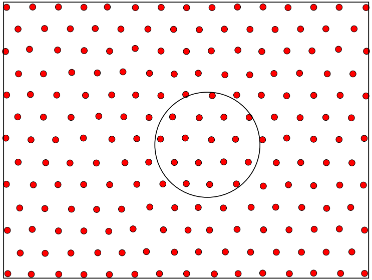
  $I(X;Y) - H(X) = -H(X|Y)$

New Result: An optimal linear code



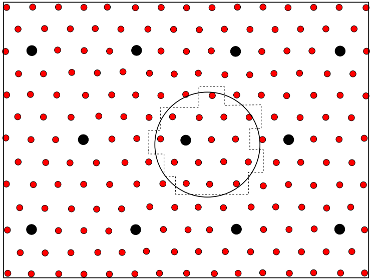
- Let $|\mathcal{X}| = p$, prime no.
- $\mathcal{C}_1$ = code book
- $\mathcal{C}_1$ has Packing Property
- Size of code
  $= \log|\mathcal{X}| - H(X|Y)$

New Result: An optimal linear code



- Let $|\mathcal{X}| = $ p, prime no.
- $\mathcal{C}_1 = $ code book
- $\mathcal{C}_1$ has Packing Property
- Size of code
  $= \log |\mathcal{X}| - H(X|Y)$
- Finite field is $\mathbb{Z}_p$
- Bounding Region $= \mathcal{X}^n$
- Density $= -H(X|Y)$

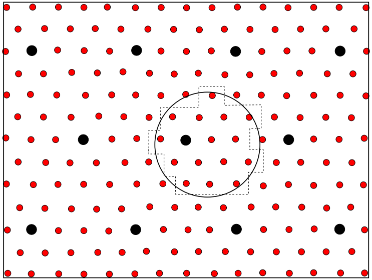
New Theorem: An optimal nested linear code



Going beyond symmetry

- $\mathcal{C}_1$ fine code (red & black)
- $\mathcal{C}_2$ coarse code (black)
- $\mathcal{C}_1$ has Packing property

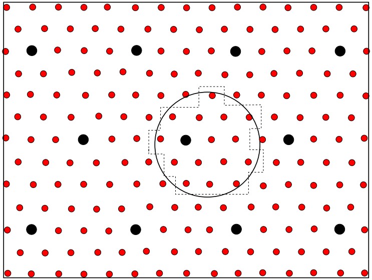
New Theorem: An optimal nested linear code



Going beyond symmetry

- $\mathcal{C}_1$ fine code (red & black)
- $\mathcal{C}_2$ coarse code (black)
- $\mathcal{C}_1$ has Packing property
- $\mathcal{C}_2$ has Shaping property
- Size of $\mathcal{C}_1 = \log|\mathcal{X}| - H(X|Y)$
- Size of $\mathcal{C}_2 = \log|\mathcal{X}| - H(X)$

New Theorem: An optimal nested linear code



Going beyond symmetry

- $\mathcal{C}_1$ fine code (red & black)
- $\mathcal{C}_2$ coarse code (black)
- $\mathcal{C}_1$ has Packing property
- $\mathcal{C}_2$ has Shaping property
- Size of $\mathcal{C}_1 = \log|\mathcal{X}| - H(X|Y)$
- Size of $\mathcal{C}_2 = \log|\mathcal{X}| - H(X)$
- Code book$= \mathcal{C}_1/\mathcal{C}_2$
- Code book size $= I(X;Y)$
- Achieves $C(W)$

Going beyond addition

- $X_2 \vee X_3$ (logical OR function)

Going beyond addition

- $X_2 \vee X_3$ (logical OR function)
- What kind of glasses you wear so this looks like addition?

Going beyond addition

- $X_2 \vee X_3$ (logical OR function)
- What kind of glasses you wear so this looks like addition?
- Can be embedded in the addition table in $\mathbb{F}_3$

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Going beyond addition

- $X_2 \vee X_3$ (logical OR function)
- What kind of glasses you wear so this looks like addition?
- Can be embedded in the addition table in $\mathbb{F}_3$

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

- Map binary sources into $\mathbb{F}_3$, and use linear codes built on $\mathbb{F}_3$
- Can do better than traditional random coding

Going beyond addition

- $X_2 \vee X_3$ (logical OR function)
- What kind of glasses you wear so this looks like addition?
- Can be embedded in the addition table in $\mathbb{F}_3$

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

- Map binary sources into $\mathbb{F}_3$, and use linear codes built on $\mathbb{F}_3$
- Can do better than traditional random coding
- In general we 'embed' bivariate functions in groups

Groups - An Introduction

- G - a finite abelian group of order n
- $G \cong \mathbb{Z}_{p_1^{e_1}} \times \mathbb{Z}_{p_2^{e_2}} \cdots \times \mathbb{Z}_{p_k^{e_k}}$
- G isomorphic to direct product of possibly repeating primary cyclic groups

$$g \in G \Leftrightarrow g = (g_1, \ldots, g_k), \;\; g_i \in \mathbb{Z}_{p_i^{e_i}}$$

- Call g_i as the ith digit of g

Groups - An Introduction

- G - a finite abelian group of order n
- $G \cong \mathbb{Z}_{p_1^{e_1}} \times \mathbb{Z}_{p_2^{e_2}} \cdots \times \mathbb{Z}_{p_k^{e_k}}$
- G isomorphic to direct product of possibly repeating primary cyclic groups

$$g \in G \Leftrightarrow g = (g_1, \ldots, g_k), \;\; g_i \in \mathbb{Z}_{p_i^{e_i}}$$

- Call g_i as the ith digit of g
- Prove coding theorems for primary cyclic groups

Nested Group Codes

- Group code over $\mathbb{Z}_{p^r}^n$: $\mathcal{C} < \mathbb{Z}_{p^r}^n$
- $\mathcal{C} = \text{Image}(\phi)$ for some homomorphism $\phi \colon \mathbb{Z}_{p^r}^k \to \mathbb{Z}_{p^r}^n$

Nested Group Codes

- Group code over $\mathbb{Z}_{p^r}^n$: $\mathcal{C} < \mathbb{Z}_{p^r}^n$
- $\mathcal{C} = \text{Image}(\phi)$ for some homomorphism $\phi\colon \mathbb{Z}_{p^r}^k \to \mathbb{Z}_{p^r}^n$
- $(\mathcal{C}_1, \mathcal{C}_2)$ nested if $\mathcal{C}_2 \subset \mathcal{C}_1$

Nested Group Codes

- Group code over $\mathbb{Z}_{p^r}^n$: $\mathcal{C} < \mathbb{Z}_{p^r}^n$
- $\mathcal{C} = \text{Image}(\phi)$ for some homomorphism $\phi\colon \mathbb{Z}_{p^r}^k \to \mathbb{Z}_{p^r}^n$
- $(\mathcal{C}_1, \mathcal{C}_2)$ nested if $\mathcal{C}_2 \subset \mathcal{C}_1$
- We need:
  - $\mathcal{C}_1 < \mathbb{Z}_{p^r}^n$: "good" packing code
  - $\mathcal{C}_2 < \mathbb{Z}_{p^r}^n$: "good" covering code

Good Group Packing Codes

- Good group channel code $\mathcal{C}_2$ for the triple $(\mathcal{U}, \mathcal{V}, P_{UV})$
- Assume $\mathcal{U} = \mathbb{Z}_{p^r}$ for some prime p and exponent $r > 0$

Good Group Packing Codes

- Good group channel code $\mathcal{C}_2$ for the triple $(\mathcal{U}, \mathcal{V}, P_{UV})$
- Assume $\mathcal{U} = \mathbb{Z}_{p^r}$ for some prime p and exponent $r > 0$

Lemma

Exists for large n if

$$\frac{1}{n} \log |\mathcal{C}_2| \leq \log p^r - \max_{0 \leq i < r} \left(\frac{r}{r-i} \right) (H(U|V) - H([U]_i|V))$$

Good Group Packing Codes

- Good group channel code $\mathcal{C}_2$ for the triple $(\mathcal{U}, \mathcal{V}, P_{UV})$
- Assume $\mathcal{U} = \mathbb{Z}_{p^r}$ for some prime p and exponent $r > 0$

Lemma

Exists for large n if

$$\frac{1}{n} \log |\mathcal{C}_2| \leq \log p^r - \max_{0 \leq i < r} \left(\frac{r}{r-i} \right) (H(U|V) - H([U]_i|V))$$

- $[U]_i$ is a function of U and depends on the group
- Extra penalty for imposing group structure beyond linearity

Good Group Packing Codes

- Good group channel code $\mathcal{C}_2$ for the triple $(\mathcal{U}, \mathcal{V}, P_{UV})$
- Assume $\mathcal{U} = \mathbb{Z}_{p^r}$ for some prime p and exponent $r > 0$

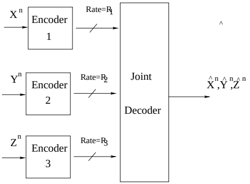
Lemma

Exists for large n if

$$\frac{1}{n} \log |\mathcal{C}_2| \leq \log p^r - \max_{0 \leq i < r} \left(\frac{r}{r-i} \right) (H(U|V) - H([U]_i|V))$$

- $[U]_i$ is a function of U and depends on the group
- Extra penalty for imposing group structure beyond linearity
- Time for questions?

A Distributed Source Coding Problem



- Encoders observe different components of a vector source
- Central decoder receives quantized observations from the encoders
- Given source distribution p_{XYZ}
- Best known rate region - Berger-Tung Rate Region, '77

Conclusions

- Presented a nested group codes based coding scheme
- Can recover known rate regions of broadcast channel
- Offers rate gains over random coding coding scheme

Conclusions

- Presented a nested group codes based coding scheme
- Can recover known rate regions of broadcast channel
- Offers rate gains over random coding coding scheme
- New bridge between probability and algebra, between information theory and coding theory

Conclusions

- Presented a nested group codes based coding scheme
- Can recover known rate regions of broadcast channel
- Offers rate gains over random coding coding scheme
- New bridge between probability and algebra, between information theory and coding theory
- It was thought that probability and algebra are nemesis

Conclusions

- Presented a nested group codes based coding scheme
- Can recover known rate regions of broadcast channel
- Offers rate gains over random coding coding scheme
- New bridge between probability and algebra, between information theory and coding theory
- It was thought that probability and algebra are nemesis
- Instead the match made in heaven