Decoding Binary Codes

Fernando Granha Jeronimo
joint work with
Dylan Quintana, Shashank Srivastava and
Madhur Tulsiani

@TTIC Workshop

Goal of the Talk

Goal

Present a new unique decoding result for a family of binary codes

Goal of the Talk

Outline

- Discuss basic properties of codes to give context ($\approx 75\%$)
- State the new unique decoding result ($\approx 10\%$)
- Mention the techniques involved ($\approx 15\%$)

Coding Theory Concepts

Alphabet

$\Sigma = \{0, \ldots, q-1\}$ a set of symbols

Coding Theory Concepts

Alphabet

$\Sigma = \{0, \ldots, q-1\}$ a set of symbols

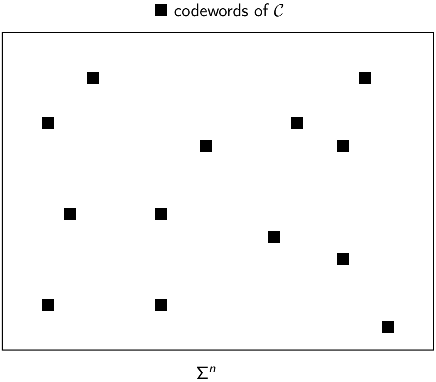
Code

A code is a subset $\mathcal{C} \subseteq \Sigma^n$

Coding Theory Concepts

Code

A code is a subset $\mathcal{C} \subseteq \Sigma^n$



Coding Theory Concepts

Message Set

Suppose we have a set of messages $\mathcal{M}$ of size $|\mathcal{C}|$.

Coding Theory Concepts

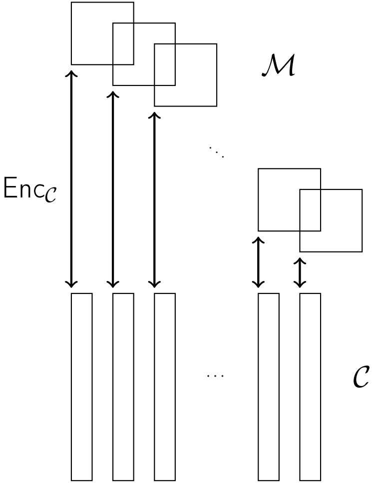
Message Set

Suppose we have a set of messages $\mathcal{M}$ of size $|\mathcal{C}|$.

Encoding Map

$\text{Enc}_{\mathcal{C}} : \mathcal{M} \to \mathcal{C}$ bijection.

Coding Theory Concepts



Coding Theory Concepts

It will be convenient to take $\mathcal{M} = \Sigma^m$ (for some $m \leq n$).

Encoding Map

$\mathrm{Enc}_{\mathcal{C}} : \Sigma^m \to \mathcal{C} \subseteq \Sigma^n$ bijection.

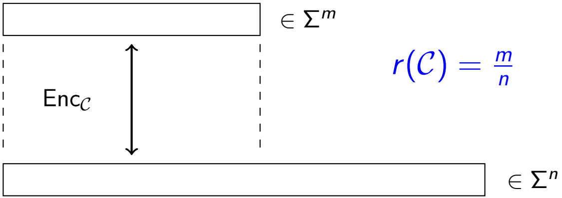
Rate

Fraction of information symbols $\frac{m}{n}$ aka the rate $r(\mathcal{C})$ of $\mathcal{C}$.

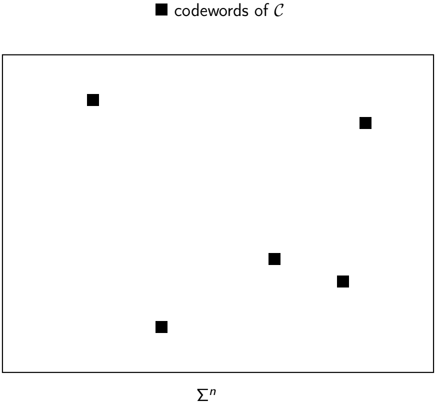
Coding Theory Concepts

Rate

Fraction of information symbols $\frac{m}{n}$ aka the rate $r(\mathcal{C})$ of $\mathcal{C}$.

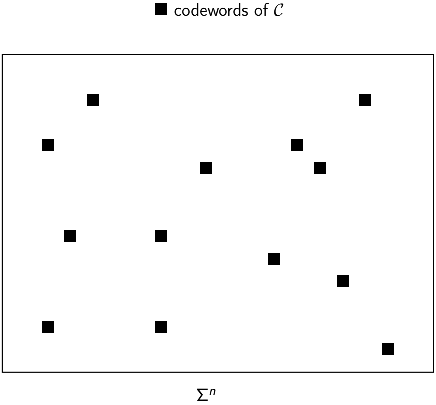


Coding Theory Concepts



Lower rate $r(\mathcal{C})$

Coding Theory Concepts



Higher rate $r(\mathcal{C})$

Coding Theory Concepts

Easy to construct high-rate codes

Take $m = n$ and $\mathrm{Enc}_{\mathcal{C}} \colon \Sigma^m \to \Sigma^n$ to be the identity.
Rate $r(\mathcal{C})$ of $\mathcal{C}$ is $m/n = 1$ (as large as possible).

Coding Theory Concepts

Easy to construct high-rate codes

Take $m = n$ and $\mathrm{Enc}_{\mathcal{C}} \colon \Sigma^m \to \Sigma^n$ to be the identity.
Rate $r(\mathcal{C})$ of $\mathcal{C}$ is $m/n = 1$ (as large as possible).

Issue

There are messages $m_1 \neq m_2$ s.t. $\mathrm{Enc}_{\mathcal{C}}(m_1)$ and $\mathrm{Enc}_{\mathcal{C}}(m_2)$ differ in exactly one symbol. If $\mathrm{Enc}_{\mathcal{C}}(m_1)$ is corrupted to $\tilde{x}$ in one symbol, then $\tilde{x}$ may be the same as $\mathrm{Enc}_{\mathcal{C}}(m_2)$.

0	0	$\cdots$	0	0

$\mathrm{Enc}_{\mathcal{C}}(m_1)$

0	0	$\cdots$	0	1

$\mathrm{Enc}_{\mathcal{C}}(m_2)$

Coding Theory Concepts

Issue

There are messages $m_1 \neq m_2$ s.t. $\mathrm{Enc}_{\mathcal{C}}(m_1)$ and $\mathrm{Enc}_{\mathcal{C}}(m_2)$ differ in exactly one symbol. If $\mathrm{Enc}_{\mathcal{C}}(m_1)$ is corrupted to $\tilde{x}$ in one symbol, then $\tilde{x}$ may be the same as $\mathrm{Enc}_{\mathcal{C}}(m_2)$.

0	0	$\cdots$	0	0

$\mathrm{Enc}_{\mathcal{C}}(m_1)$

0	0	$\cdots$	0	1

$\mathrm{Enc}_{\mathcal{C}}(m_2)$

Solution

Require $\mathrm{Enc}_{\mathcal{C}}(m_1)$ and $\mathrm{Enc}_{\mathcal{C}}(m_2)$ to **differ in many positions** for every $m_1 \neq m_2$.

Coding Theory Concepts

Hamming Distance

The Hamming distance between $z, z' \in \Sigma^n$ is

$$\Delta(z, z') := |\{i \mid z_i \neq z'_i\}|.$$

Coding Theory Concepts

Hamming Distance

The Hamming distance between $z, z' \in \Sigma^n$ is

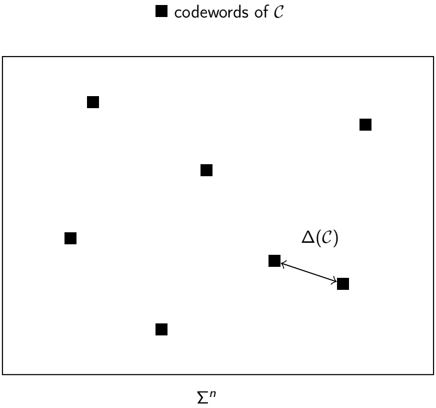
$$\Delta(z, z') := |\{i \mid z_i \neq z'_i\}|.$$

Minimum Distance of a Code

The distance $\Delta(\mathcal{C})$ of $\mathcal{C}$ is

$$\Delta(\mathcal{C}) := \min_{z, z' \in \mathcal{C}:\ z \neq z'} \Delta(z, z').$$

Coding Theory Concepts



Coding Theory Concepts

Easy to construct high-distance codes

Take $m = 1$ and $\mathsf{Enc}_{\mathcal{C}} : \Sigma \to \Sigma^n$ to be the replication map, namely,

$$\mathsf{Enc}_{\mathcal{C}}(\sigma) = \underbrace{\sigma \cdots \sigma}_{n \text{ times}},$$

for $\sigma \in \Sigma$.

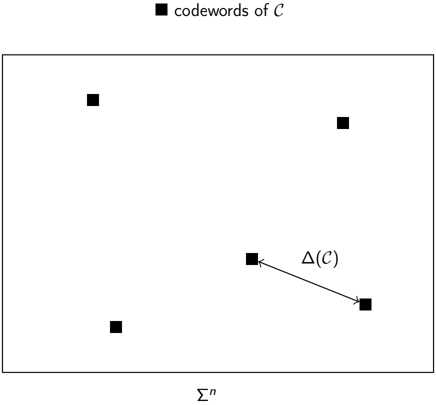
- $\Delta(\mathcal{C}) = n$ (as large as possible).
- Rate of $\mathcal{C}$ is $1/n \to 0$ as $n \to \infty$ (vanishing rate).

Coding Theory Concepts

Tension

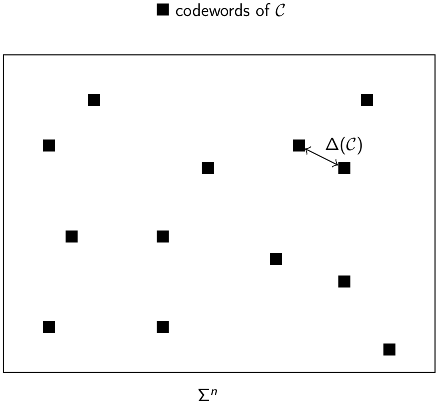
- Increasing the rate $r(\mathcal{C})$ may reduce the distance $\Delta(\mathcal{C})$
- Increasing the distance $\Delta(\mathcal{C})$ may reduce the rate $r(\mathcal{C})$

Coding Theory Concepts



Lower rate $r(\mathcal{C})$

Coding Theory Concepts



Higher rate $r(\mathcal{C})$

Coding Theory Concepts

Question

What is the best trade-off between rate $r(\mathcal{C})$ and distance $\Delta(\mathcal{C})$?

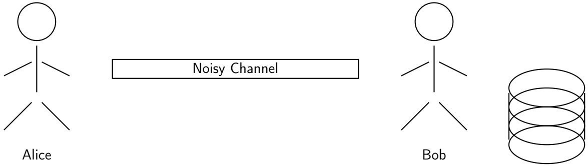
Coding Theory Concepts

Question

What is the best trade-off between rate $r(\mathcal{C})$ and distance $\Delta(\mathcal{C})$?

Applications

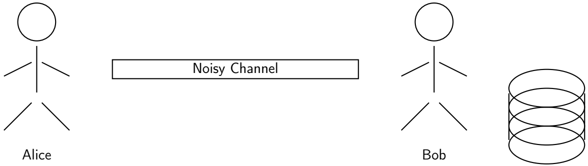
- **Optimally** storing data robustly against corruptions
- **Optimally** communicating via a noisy channel



Coding Theory Concepts

Applications

- **Optimally** storing data robustly against corruptions
- **Optimally** communicating via a noisy channel



Question

What do we mean by "optimally"?

Coding Theory Concepts

Question

What do we mean by "optimally"?

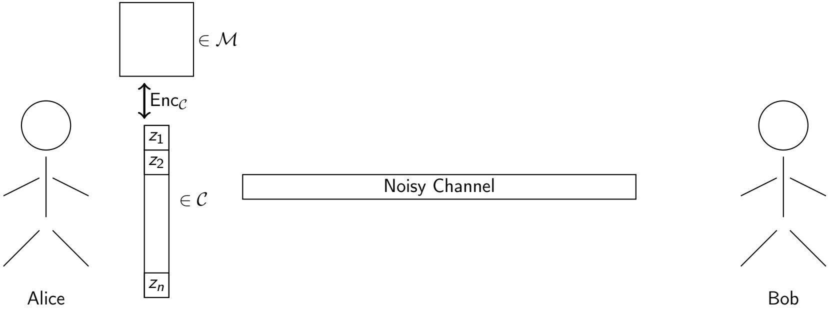
To answer the question above we need to define an error model

Error Model

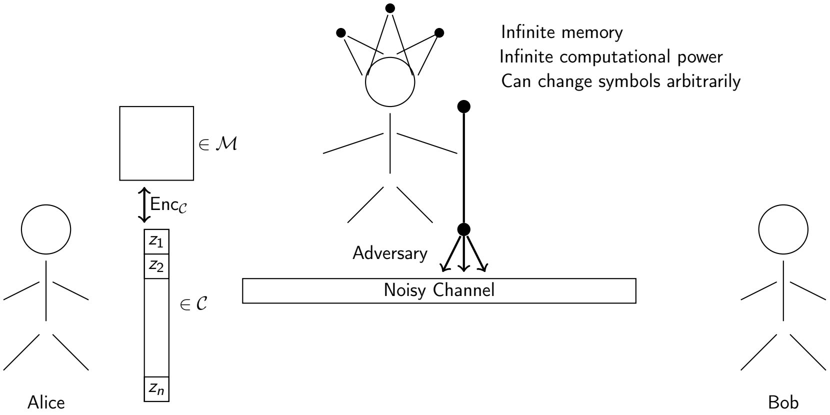
What is the error model?

Coding Theory Concepts

Alice encodes her message and sends $z_1, \ldots, z_n$, one symbol at a time, to Bob



Coding Theory Concepts

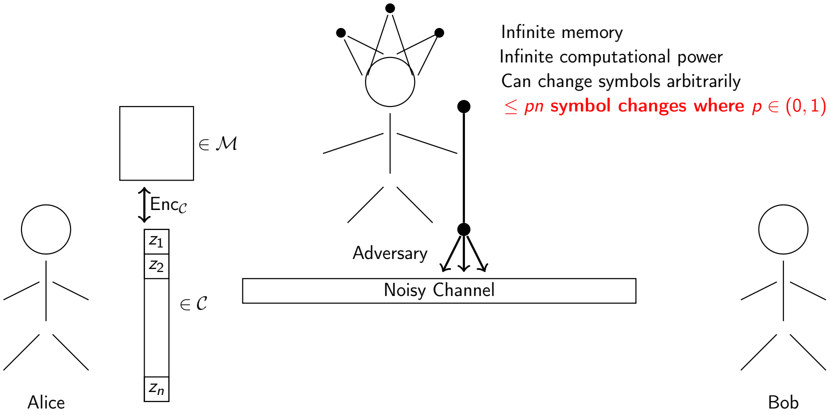


Coding Theory Concepts

Issue

Adversary is too powerful. For instance, adversary can map all code words to $\underbrace{00\ldots0}_{n}$

Coding Theory Concepts



Coding Theory Concepts

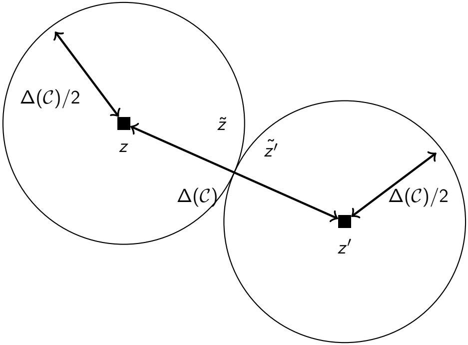
Question

How large can we take $p \in [0, 1)$ to be?

Coding Theory Concepts

Question

How large can we take $p \in [0, 1)$ to be? In theory, any $p \in [0, \Delta(\mathcal{C})/2)$ is valid for unique decoding.



Coding Theory Concepts

Error Model

We will consider this adversarial error model also known as **Hamming model**.

Figure: Richard W. Hamming (source: mathshistory.st-andrews.ac.uk).

Coding Theory Concepts

What do we mean by optimal storage/communication?

If we want to be robust against a p fraction of adversarial errors, what is the best possible rate (equivalently the least amount of redundancy needed)?

Coding Theory Concepts

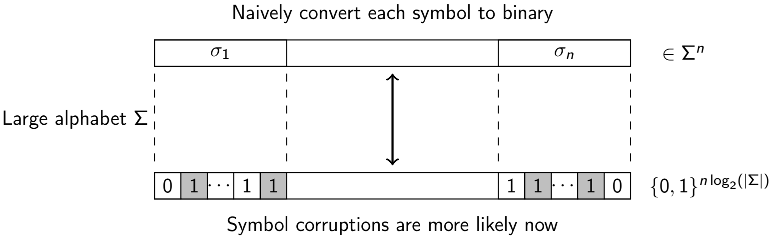
Code Parameters

- We have seen the role of distance and rate
- What about the role of the alphabet size?

Coding Theory Concepts

Large Alphabet Issue

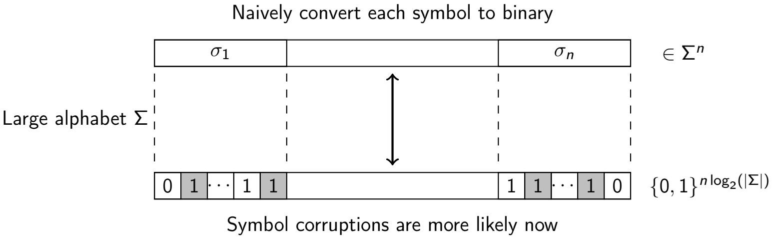
- Many information systems are inherent binary
- Naively "binarifying" a code may ruin its distance guarantee



Coding Theory Concepts

Large Alphabet Issue

- Many information systems are inherent binary
- Naively "binarifying" a code may ruin its distance guarantee



Coding Theory Concepts

Solution

Use a binary code (i.e., $\Sigma = \mathbb{F}_2 = \{0, 1\}$) from the start

Coding Theory Concepts

Solution

Use a binary code (i.e., $\Sigma = \mathbb{F}_2 = \{0, 1\}$) from the start

Issue

Binary codes are not that well understood (more on it shortly)

Coding Theory Concepts

Use the probabilistic method as an yardstick for binary codes

Random Construction

We will construct a random linear binary code and observe its rate vs distance trade-off

Coding Theory Concepts

Digression: Linear Binary Code

A linear binary code $\mathcal{C}$ has $\mathrm{Enc}_{\mathcal{C}}$ as a linear operator $\mathrm{G} : \mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$

Coding Theory Concepts

Digression: Linear Binary Code

A linear binary code $\mathcal{C}$ has $\text{Enc}_{\mathcal{C}}$ as a linear operator $\text{G} : \mathbb{F}_2^m \to \mathbb{F}_2^n$

Fact

If $\mathcal{C} \subseteq \mathbb{F}_2^n$ is linear, then $\Delta(\mathcal{C}) = \min_{z \in \mathcal{C} \setminus \{0\}} |\{i : z_i = 1\}| / n$

Coding Theory Concepts

Take $\mathsf{G} \in \mathbb{F}_2^{n \times m}$ uniformly at random, that is,

$$\mathsf{G} = \begin{pmatrix} g_{1,1} & \cdots & g_{1,m} \\ \vdots & \ddots & \vdots \\ g_{n,1} & \cdots & g_{n,m} \end{pmatrix}$$

where each $g_{i,j}$ is uniform in $\mathbb{F}_2$.

Coding Theory Concepts

Let $x \in \mathbb{F}_2^m$ be a non-zero vector and $j^* = \max\{j : x_i = 1\}$. Then

$$Gx = \sum_{j:\, x_j=1} \begin{pmatrix} g_{1,j} \\ \vdots \\ g_{n,j} \end{pmatrix} = \sum_{j:\, x_j=1, j<j^*} \begin{pmatrix} g_{1,j} \\ \vdots \\ g_{n,j} \end{pmatrix} + \begin{pmatrix} g_{1,j^*} \\ \vdots \\ g_{n,j^*} \end{pmatrix}$$

Hence, $(Gx)_i$'s are uniformly and independently distributed in $\mathbb{F}_2$.

Coding Theory Concepts

Set $\mathbf{X}_i = \mathbf{1}\left[(Gx)_i = 1\right]$ and $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$.

Coding Theory Concepts

Set $\mathbf{X}_i = \mathbf{1}[(\mathrm{G}x)_i = 1]$ and $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$.
Note that $\mathbf{E}\mathbf{X} = n/2$. By the Chernoff bound,

$$\Pr[|\mathbf{X} - \mathbf{E}\mathbf{X}| > \beta n] \leq \exp(-O(\beta^2 n)).$$

Coding Theory Concepts

Set $\mathbf{X}_i = \mathbf{1}[(Gx)_i = 1]$ and $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$.
Note that $\mathbf{E}\mathbf{X} = n/2$. By the Chernoff bound,

$$\Pr[|\mathbf{X} - \mathbf{E}\mathbf{X}| > \beta n] \leq \exp(-O(\beta^2 n)).$$

By union bound,

$$\Pr_G[\Delta(\mathcal{C}) < 1/2 - \beta] = \Pr_G[\exists x \in \mathbb{F}_2^m \setminus \{0\} : ||Gx|| < 1/2 - \beta]$$

Coding Theory Concepts

Set $\mathbf{X}_i = \mathbf{1}[(Gx)_i = 1]$ and $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$.
Note that $\mathbf{E}\mathbf{X} = n/2$. By the Chernoff bound,

$$\Pr[|\mathbf{X} - \mathbf{E}\mathbf{X}| > \beta n] \le \exp(-O(\beta^2 n)).$$

By union bound,

$$\Pr_G[\Delta(\mathcal{C}) < 1/2 - \beta] = \Pr_G[\exists x \in \mathbb{F}_2^m \setminus \{0\} : ||Gx|| < 1/2 - \beta]$$
$$\le 2^m \cdot \exp(-O(\beta^2 n)),$$

Coding Theory Concepts

Set $\mathbf{X}_i = \mathbf{1}[(\mathrm{G}x)_i = 1]$ and $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$.
Note that $\mathbf{E}\mathbf{X} = n/2$. By the Chernoff bound,

$$\Pr[|\mathbf{X} - \mathbf{E}\mathbf{X}| > \beta n] \leq \exp(-O(\beta^2 n)).$$

By union bound,

$$\Pr_{\mathrm{G}}[\Delta(\mathcal{C}) < 1/2 - \beta] = \Pr_{\mathrm{G}}[\exists x \in \mathbb{F}_2^m \setminus \{0\} : ||Gx|| < 1/2 - \beta] \leq 2^m \cdot \exp(-O(\beta^2 n)),$$

which vanishes for $n = \Theta(m/\beta^2)$, i.e., $r(\mathcal{C}) = \Theta(\beta^2)$.

Coding Theory Concepts

Theorem (Gilbert–Varshamov Bound 1950's (asymptotic version))

There are binary codes of distance $1/2 - \beta$ *and rate* $\Theta(\beta^2)$.

Coding Theory Concepts

Theorem (Gilbert–Varshamov Bound 1950's (asymptotic version))

There are binary codes of distance $1/2 - \beta$ *and rate* $\Theta(\beta^2)$.

Question

Can we do better?

Coding Theory Concepts

Theorem (Gilbert–Varshamov Bound 1950's (asymptotic version))

There are binary codes of distance $1/2 - \beta$ *and rate* $\Theta(\beta^2)$.

Question

Can we do better?

We do not know, but certainly not by much!

Theorem (LP Bound MRRW 1977)

Binary codes of distance $1/2 - \beta$ *can have rate at most* $O(\beta^2 \log(1/\beta))$ *(if any).*

Coding Theory Concepts

Theorem (Gilbert–Varshamov Bound 1950's (asymptotic version))

There are binary codes of distance $1/2 - \beta$ *and rate* $\Theta(\beta^2)$.

Guessing...

The Gilbert–Varshamov is quite old and optimal (rate vs distance) binary codes are quite fundamental so this part of coding theory should be well established by now.

Coding Theory Concepts

Guessing...

The Gilbert–Varshamov is quite old and optimal (rate vs distance) binary codes are quite fundamental so this part of coding theory should be well established by now.

Guess is not correct

Binary codes are not that well understood (specially compared to larger alphabet codes). We lack:

- explicit constructions,
- decoding algorithmic tools, and
- tighter impossibility results.

Coding Theory Concepts

Wait, aren't we essentially done?

Random linear codes achieve the Gilbert–Varshamov bound thereby having a nearly optimal rate vs distance trade-off.

Coding Theory Concepts

Wait, aren't we essentially done?

Random linear codes achieve the Gilbert–Varshamov bound thereby having a nearly optimal rate vs distance trade-off.

Quite Far

- Decoding random linear code is likely to be **hard**. Known algorithms run in time $2^{\Omega(n)}$.
- Given $\mathrm{G} \in \mathbb{F}_2^{n \times m}$ sampled uniformly, how do we certify $\Delta(\mathcal{C}) \geq 1/2 - \beta$? (in principle $\Delta(\mathcal{C})$ can be small)

Coding Theory Concepts

Quest for Explicit Construction

A code $\mathcal{C} \subseteq \Sigma^n$ is explicit if the encoding $\mathrm{Enc}_{\mathcal{C}}(\cdot)$ can be computed in time $\mathrm{poly}(n, |\Sigma|)$.

- Advantage: avoid the issue of not knowing $\Delta(\mathcal{C})$

Coding Theory Concepts

Adversarial Error Regime (Hamming model)

Holy Grail of Coding Theory

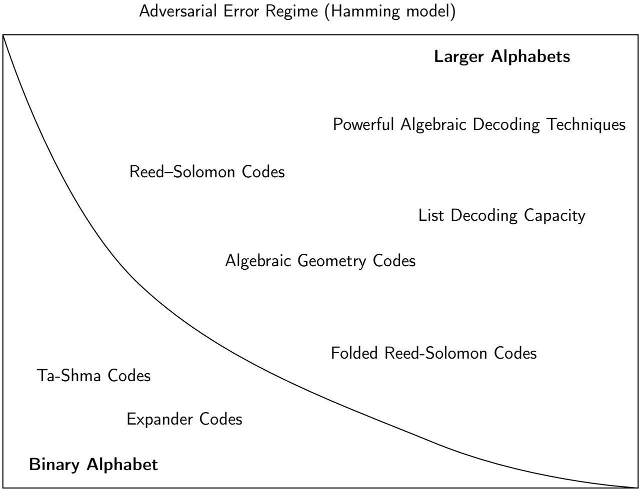
- Code $\mathcal{C}$ is over small alphabet Σ (ideally binary)
- Code $\mathcal{C}$ is explicit
- Code $\mathcal{C}$ achieves optimal parameters
- Code $\mathcal{C}$ is efficiently decodable

Context

Previous Results

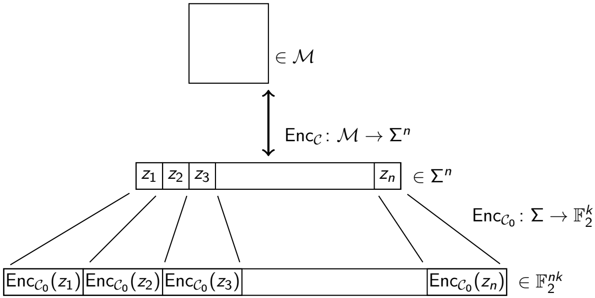
We take a detour through the state-of-the-art techniques.

Context


Adversarial Error Regime (Hamming model)
Larger Alphabets
Powerful Algebraic Decoding Techniques
Reed–Solomon Codes
List Decoding Capacity
Algebraic Geometry Codes
Folded Reed-Solomon Codes
Ta-Shma Codes
Expander Codes
Binary Alphabet

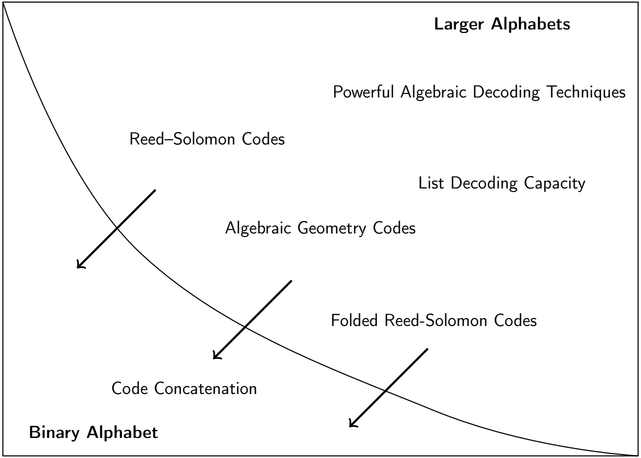
Context

"Binarifying" codes (a second approach) via code concatenation of $\mathcal{C}$ and $\mathcal{C}_0$



Context

Popular approach: obtain results by concatenating with binary codes



Context

In this line of code concatenation, the closest results to our work are:

Theorem (Guruswami–Indyk'04)

There are efficiently decodable non-explicit codes at the Gilbert–Varshamov bound

Context

In this line of code concatenation, the closest results to our work are:

Theorem (Guruswami–Indyk'04)

There are efficiently decodable non-explicit codes at the Gilbert–Varshamov bound

Theorem (Guruswami–Rudra'06)

There are explicit binary codes list decodable from radius $1/2 - \beta$ *and rate* $\Omega(\beta^3)$ *(at the Zyablov bound)*

Context

In this line of code concatenation, the closest results to our work are:

Theorem (Guruswami–Indyk'04)

There are efficiently decodable non-explicit codes at the Gilbert–Varshamov bound

Theorem (Guruswami–Rudra'06)

There are explicit binary codes list decodable from radius $1/2 - \beta$ *and rate* $\Omega(\beta^3)$ *(at the Zyablov bound)*

Context

Possible Issue

Is our lack knowledge a result of the relatively few “genuinely” binary techniques?

Context

A "genuinely" binary technique was discovered leading to the following breakthrough result

Theorem (Ta-Shma 2017)

*For every $\beta > 0$, there are **explict** codes near the Gilbert–Varshamov bound, namely, codes $\mathcal{C}$ with*

- *distance $\Delta(\mathcal{C}) \geq 1/2 - \beta$, and*
- *rate $r(\mathcal{C}) = \Omega(\beta^{2+\epsilon})$,*

where $\epsilon \to 0$ as $\beta \to 0$.

Context

Ta-Shma's codes score highly on the holy grail scale

Holy Grail of Coding Theory

- Code $\mathcal{C}$ is over binary alphabet Σ
- Code $\mathcal{C}$ is explicit (only explicit construction in this regime)
- Code $\mathcal{C}$ achieves near optimal parameters
- Code $\mathcal{C}$ is efficiently decodable (not known)

Context

Missing Piece

It was left open whether Ta-Shma's codes can be efficiently decoded leaving the possibility of this being a computationally hard task

Context

Missing Piece

It was left open whether Ta-Shma's codes can be efficiently decoded leaving the possibility of this being a computationally hard task

Striking Reality

In this adversarial regime, we are not storing/transmitting data as efficiently as it is theoretically possible because we do not know explicit efficiently decodable (near) optimal binary codes.

- What is the energy cost of this inefficiency?
- What is the storage cost of this inefficiency?

Main Result

Disclaimer

The next result was not thoroughly peer reviewed

Main Result

Theorem (Main Result Informal)

Ta-Shma's codes can be efficiently decoded

Main Result

More precisely, we have:

Theorem (Main Result)

For every $\beta > 0$, there are ***explict Ta-Shma*** *codes near the Gilbert–Varshamov bound, namely, codes $\mathcal{C}$ with*

- *distance $\Delta(\mathcal{C}) \geq 1/2 - \beta$, and*
- *rate $r(\mathcal{C}) = \Omega(\beta^{2+\epsilon})$,*
- ***$\mathcal{C}$ is uniquely decodable in time*** $n^{(1/\beta)^{O(1)}}$,

where $\epsilon \to 0$ as $\beta \to 0$.

Main Result

More precisely, we have:

Theorem (Main Result)

For every $\beta > 0$, there are ***explict Ta-Shma*** *codes near the Gilbert–Varshamov bound, namely, codes $\mathcal{C}$ with*

- *distance $\Delta(\mathcal{C}) \geq 1/2 - \beta$, and*
- *rate $r(\mathcal{C}) = \Omega(\beta^{2+\epsilon})$,*
- *$\mathcal{C}$* ***is uniquely decodable in time*** *$n^{(1/\beta)^{O(1)}}$,*

where $\epsilon \to 0$ as $\beta \to 0$. ***Furthermore, if*** *$\epsilon > 0$* ***is a constant, then unique decoding takes time*** $\mathrm{poly}(n/\beta)$.

Main Result

Adversarial Error Regime (Hamming model)

Holy Grail of Coding Theory

- Code $\mathcal{C}$ is over binary alphabet Σ
- Code $\mathcal{C}$ is explicit
- Code $\mathcal{C}$ achieves near optimal parameters
- Code $\mathcal{C}$ is efficiently decodable

Main Result

Question

Are we “nearly” done now?

Main Result

Question

Are we “nearly” done now?

Not really

Albeit polynomial time, the decoding algorithm might be too slow for practical use

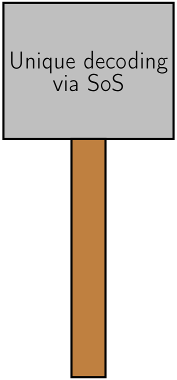
Figure: (source: wikipedia.org).

Bird's-eye view of Techniques

What are the techniques?

We will just mention the techniques at a very high-level

Techniques



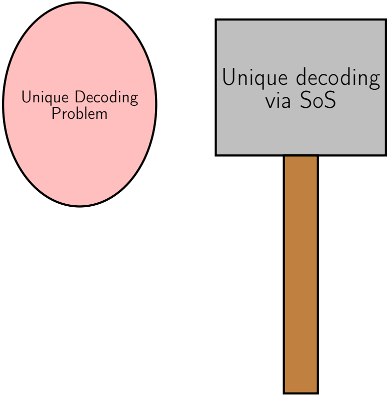
Techniques

Sum-of-Squares (SoS)

Sum-of-Squares is a semi-definite programming hierarchy

- It generalizes linear programming
- Captures the state-of-the-art results for many problems (MAX-CUT and other CSPs)
- Level d of SoS runs in time $n^{O(d)}$ where n is the number of variables

Techniques



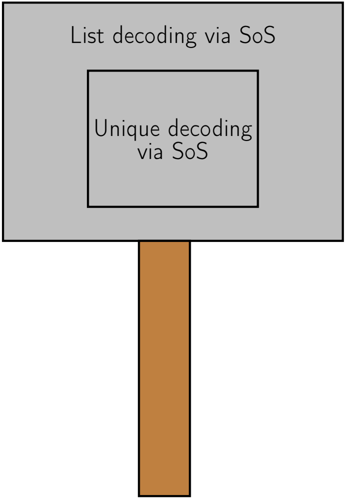
Techniques

First Hammer Effect

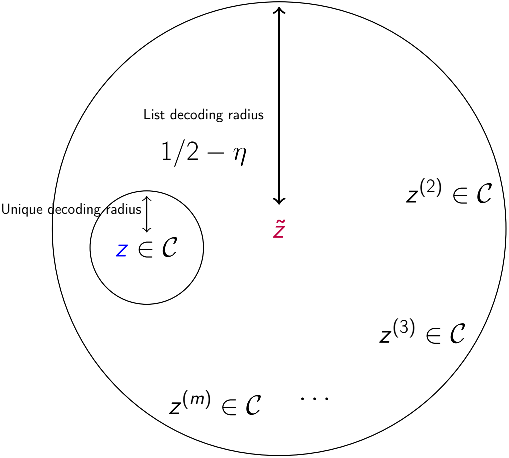
Can decode explicit binary codes $\mathcal{C}$ satisfying

- $\Delta(\mathcal{C}) \geq 1/2 - \beta$, and
- rate $r(\mathcal{C}) = 2^{-\text{polylog}(1/\beta)} \ll \beta^{2+\epsilon}$ (not even polynomial rate)

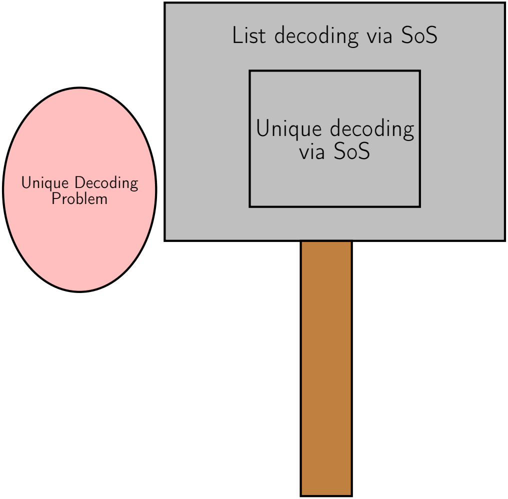
Bird's-eye view of Techniques



Bird's-eye view of Techniques



Bird's-eye view of Techniques

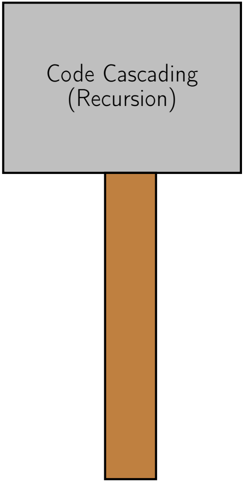


Bird's-eye view of Techniques

Second Hammer Effect

Some parameters are better but $r(\mathcal{C})$ still not even polynomial

Bird's-eye view of Techniques



Bird's-eye view of Techniques

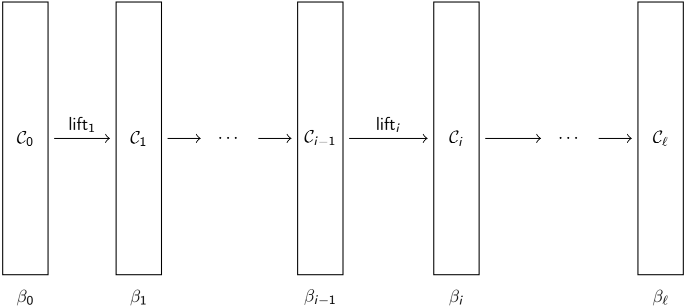
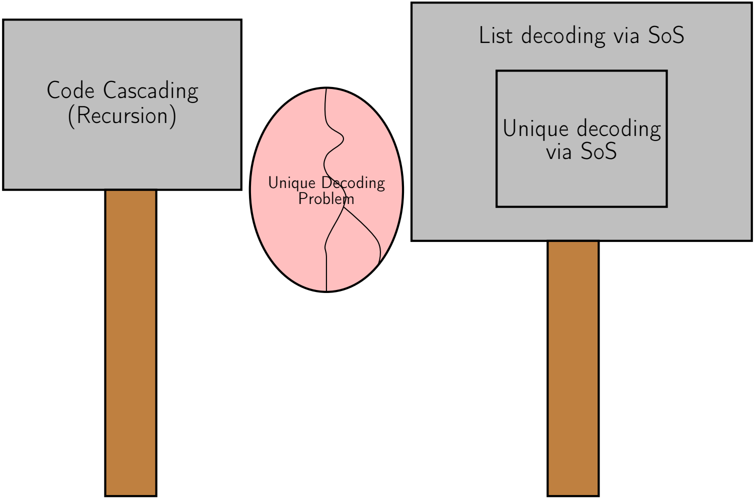


Figure: Code cascading: recursive construction of codes.

Bird's-eye view of Techniques



Bird's-eye view of Techniques

Second and Third Hammers Effect

Decode Ta-Shma's codes with nearly optimal rate

Open Problems

A central open problem in coding theory [Guruswami'10]

Extended Holy Grail of Coding Theory

- Code $\mathcal{C}$ is over binary alphabet Σ
- Code $\mathcal{C}$ is explicit
- Code $\mathcal{C}$ achieves optimal parameters
- Code $\mathcal{C}$ is efficiently **list decodable**

That's all.

Thank you!