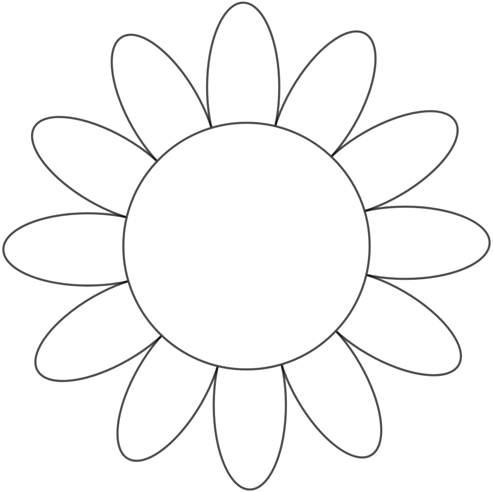
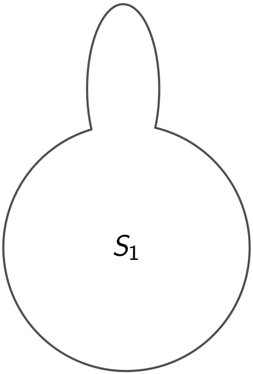
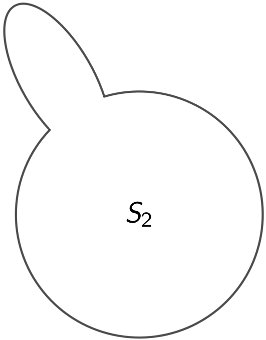
Sunflowers, daisies and local codes

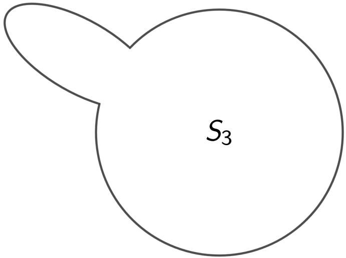
Marcel de Sena
(joint with Tom Gur and Oded Lachish)

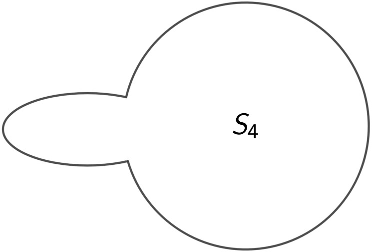


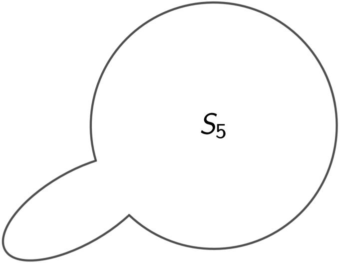
$$\mathcal{S} = \{S_1, S_2, \ldots, S_{12}\}$$

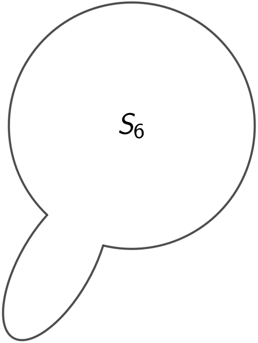

S_1

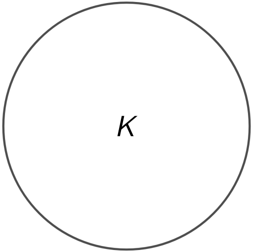

S_2

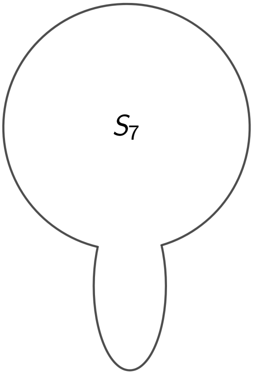

S_3


S_4

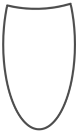

S_5


S_6


K


S_7

$$P_7 = S_7 \setminus K_7$$



Definition

A *sunflower* is a collection $\mathcal{S}$ of q-sets such that $S \cap S' = \cap_{T \in \mathcal{S}} T = K$ for any distinct $S, S' \in \mathcal{S}$. The set K is called the *kernel*, and each $P = S \setminus K$ is a *petal*.

Sunflower lemma [ER60]

If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower lemma [ER60]

If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower conjecture [ER60]

For some $c : \mathbb{N} \to \mathbb{N}$, if $|\mathcal{S}| \geq c(s)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower lemma [ER60]

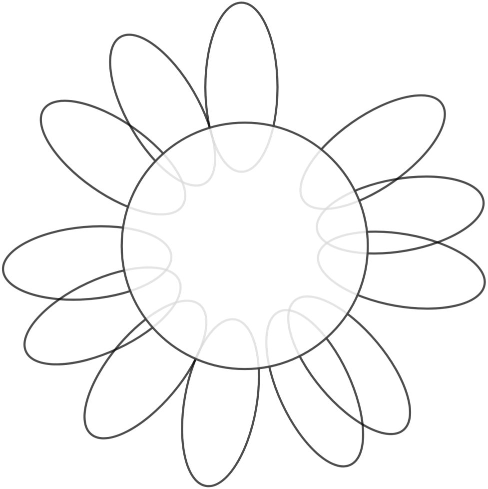
If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower conjecture [ER60]

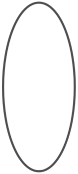
For some $c : \mathbb{N} \to \mathbb{N}$, if $|\mathcal{S}| \geq c(s)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Theorem [ALWZ19]

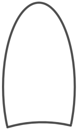
If $|\mathcal{S}| = \Omega(s \log q)^q$, then $\mathcal{S}$ contains a sunflower of size s.



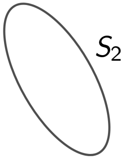
$$\mathcal{D} = \{S_1, S_2, \ldots, S_{12}\}$$

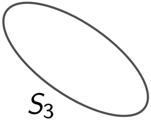


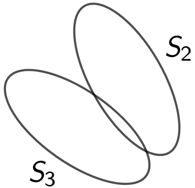
S_1



$$P_1 = S_1 \setminus K$$


S_2

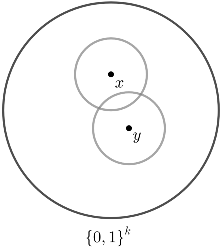

S_3


S_2
S_3

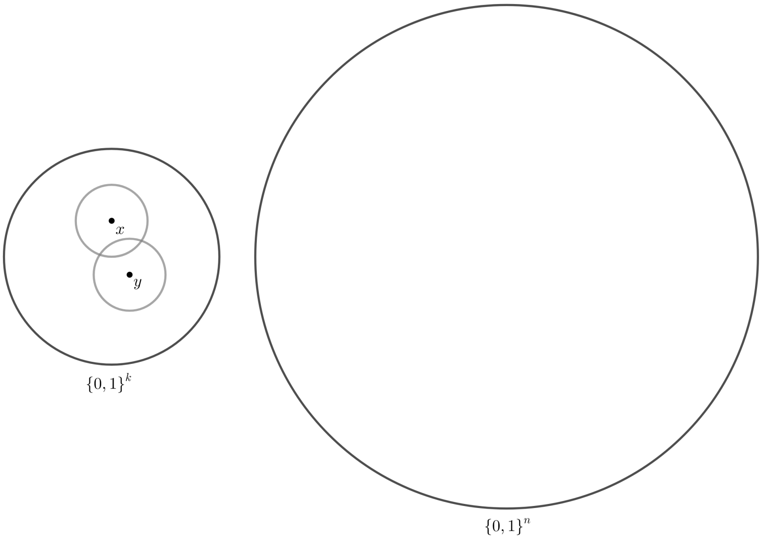
Definition

A *t-daisy* with *kernel* K is a collection $\mathcal{D}$ of q-sets such that each $i \notin K$ is contained in at most t members of $\mathcal{D}$.

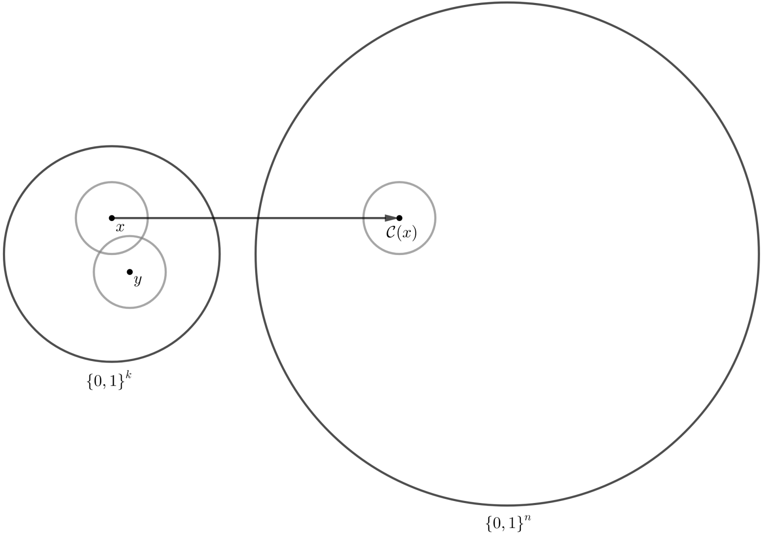
Error-correcting codes



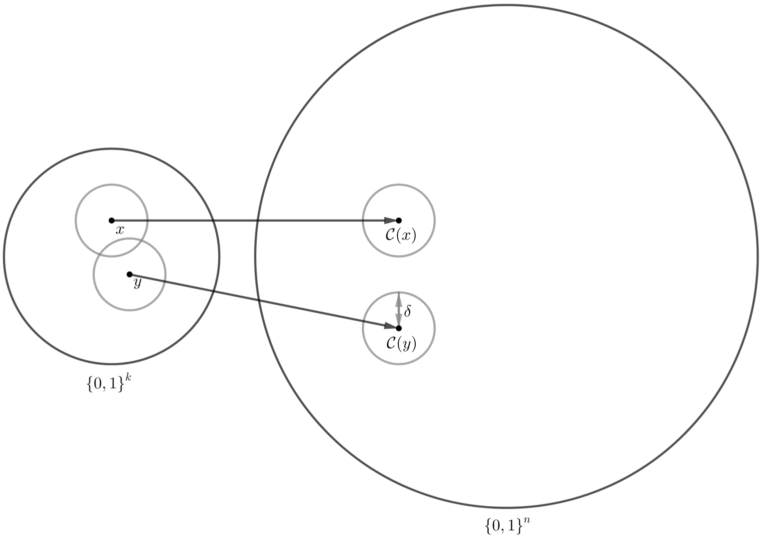
Error-correcting codes



Error-correcting codes



Error-correcting codes



Error-correcting codes

Definition

An *error-correcting code* is an injective function $\mathcal{C} : \Gamma^k \to \Sigma^n$ where the preimage (message) is recoverable after significant corruption of the image (codeword).

If a message is recoverable from at most Δn corrupted coordinates, Δ is the (relative) *distance* of the code. k is its *message length* and n is its *blocklength*.

Error-correcting codes

Definition

A binary *error-correcting code* is an injective function $\mathcal{C} : \{0,1\}^k \to \{0,1\}^n$ where the preimage (message) is recoverable after significant corruption of the image (codeword).

If a message is recoverable from at most Δn corrupted coordinates, Δ is the (relative) *distance* of the code. k is its *message length* and n is its *blocklength*.

Error-correcting codes

Codes with large distance are resilient to corruption; codes with large *rate* k/n have little redundancy. **Goal: find codes with high rate and distance.**

Error-correcting codes

Codes with large distance are resilient to corruption; codes with large *rate* k/n have little redundancy. **Goal: find codes with high rate and distance.**

Singleton bound

For any code $\mathcal{C} : \Gamma^k \to \Sigma^n$,

$$|\Gamma|^{k/n} \leq |\Sigma|^{1-\Delta+1/n}.$$

Error-correcting codes

Codes with large distance are resilient to corruption; codes with large *rate* k/n have little redundancy. **Goal: find codes with high rate and distance.**

Singleton bound

For any binary code $\mathcal{C}$,

$$k/n + \Delta \leq 1 - 1/n < 1.$$

Locally decodable codes

Definition

$\mathcal{C}$ is a *locally decodable code* (LDC) if one need only look at a small number of coordinates of $w \approx \mathcal{C}(x)$ to decode x_i.

Locally decodable codes

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that, if w is δ-close to $\mathcal{C}(x)$, then, $\forall i$,

$$\mathbb{P}[D^w(i) = x_i] \geq 2/3$$

and D makes $q = o(n)$ queries to w.

Locally decodable codes

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that, if w is δ-close to $\mathcal{C}(x)$, then, $\forall i$,

$$\mathbb{P}[D^w(i) = x_i] \geq 2/3$$

and D makes $q = O(1)$ queries to w.

Relaxed locally decodable codes

Definition

$\mathcal{C}$ is a *relaxed locally decodable code* (RLDC) if $\mathcal{C}$ is (almost) locally decodable but D can sometimes fail and return $\perp$.

Relaxed locally decodable codes

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that

- if $w = \mathcal{C}(x)$, then
$$\mathbb{P}[D^w(i) = x_i] \geq 2/3;$$
- if w is δ-close to $\mathcal{C}(x)$, then
$$\mathbb{P}[D^w(i) \in \{x_i, \perp\}] \geq 2/3;$$

and D makes $q = O(1)$ queries to w.

Theorem [GL19]

Any *one-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{2^q}\right)}.$$

Theorem [GL19]

Any *one-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{2^q}\right)}.$$

Theorem

Any *two-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{q^2}\right)}.$$

One-sided RLDCs

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that

- if $w = \mathcal{C}(x)$, then
$$\mathbb{P}[D^w(i) = x_i] = 1;$$
- if w is δ-close to $\mathcal{C}(x)$, then
$$\mathbb{P}[D^w(i) = x_i] \geq 2/3;$$

and D makes $q = o(n)$ queries to w.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
    - Randomness reduction
    - Independence from decision trees
    - Soundness amplification
    - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

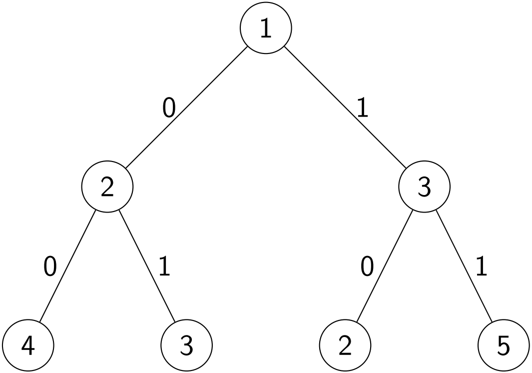
1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

Overview

1. Local decoder D' as decision trees and predicates
2. Preprocessing: from D', obtain D after
   - Randomness reduction
   - Independence from decision trees
   - Soundness amplification
   - Combinatorialisation
3. From D, obtain *global* decoder G – using daisies!
4. G decodes k bits *of a valid codeword* with high probability and $o(n)$ queries: information theoretically, $n = \omega(k)$.

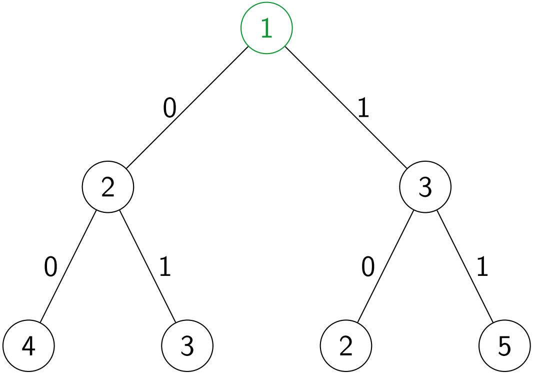
Local decoders and decision trees

$$w = (1, 0, 0, \ldots)$$



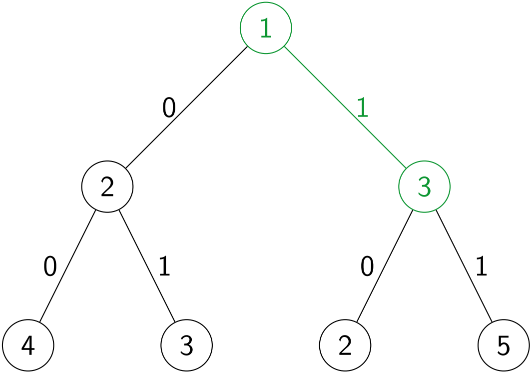
Local decoders and decision trees

$$w = (1, 0, 0, \dots)$$



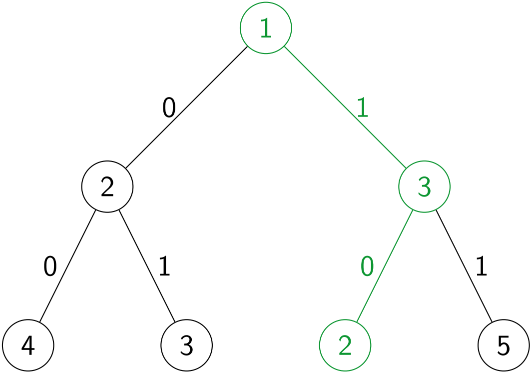
Local decoders and decision trees

$$w = (1, 0, 0, \ldots)$$



Local decoders and decision trees

$$w = (1, 0, 0, \ldots)$$



Local decoders and decision trees

Input w and tree T determine set $S = \{1, 2, 3\}$.

$$D^w(i) = f_{i,T}(1, 0, 0)$$

Local decoders and decision trees

$$D(i) = (\mu_i, \{f_{i,T} : T \in \mathcal{T}_i\})$$

Distribution μ_i over decision trees $\mathcal{T}_i$ capture randomness of D.
Predicates $f_{i,T} : \{0,1\}^q \to \{0,1,\perp\}$ determine its output.

Preprocessing

Lemma (randomness reduction)

$\exists$ relaxed decoder D with query complexity $O(q')$ and randomness complexity $\log(n) + O(1)$.

Decoder D':

- message length k;
- blocklength n;
- randomness complexity r;
- decoding radius δ;
- query complexity q';
- soundness ε'.

Preprocessing

Lemma (independence from decision trees)

$\exists$ local decoder D with soundness $O(\varepsilon')$ whose predicates only depends on sets.

Decoder D':

- message length k;
- blocklength n;
- randomness complexity r;
- decoding radius δ;
- query complexity q';
- soundness ε'.

Preprocessing

Lemma (soundness amplification)

For any $\varepsilon > 0$, $\exists$ relaxed decoder D with query complexity $O(q' \cdot \log(\varepsilon'/\varepsilon))$ and soundness ε.

Decoder D':

- message length k;
- blocklength n;
- randomness complexity r;
- decoding radius δ;
- query complexity q';
- soundness ε'.

Preprocessing

Lemma (combinatorialisation)

$\exists$ *combinatorial* relaxed decoder D with soundness $O(\varepsilon')$.

Decoder D':

- message length k;
- blocklength n;
- randomness complexity r;
- decoding radius δ;
- query complexity q';
- soundness ε'.

Preprocessing

Corollary

There exists a *combinatorial* relaxed decoder D with:

- message length k;
- blocklength n;
- randomness complexity $\log(n) + \rho$;
- decoding radius δ;
- query complexity $q = O(q')$;
- soundness $\varepsilon = O\left(\min\{q^{-1}, 2^{-\rho}\}\right)$.

Daisy partition lemma

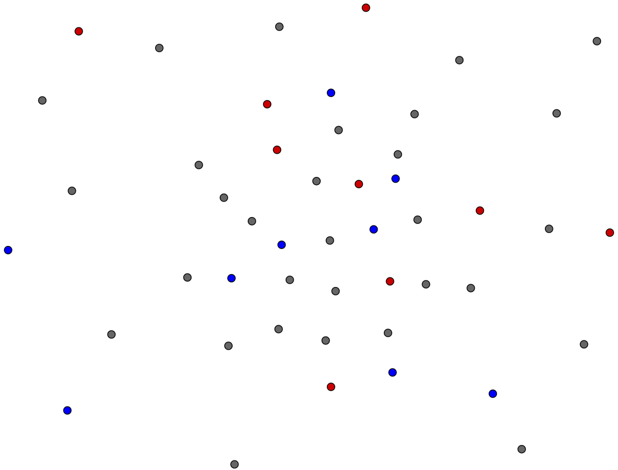
Let μ be a distribution over $2^{[n]}$ whose support is $\mathcal{S} \subseteq \binom{[n]}{q}$, with $|\mathcal{S}| = \alpha n$.
Define $m = \max\{1, j-1\}$. Then $\mathcal{S}$ can be partitioned into

$$\{\mathcal{D}_j : j \in [q]\},$$

where $\mathcal{D}_j$ is a $\alpha n^{m/q}$-daisy with petals of size j.
The kernel of $\mathcal{D}_j$ satisfies $|K_j| \leq q n^{1-j/q}$.

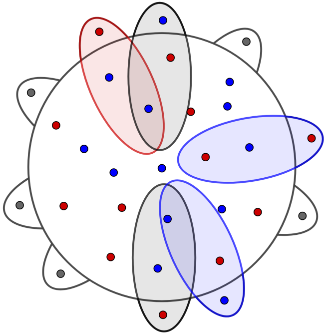
Global decoder G: construction

Binomial sampling with $p = n^{-\frac{1}{2q^2}}$



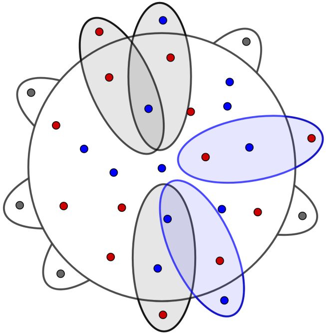
Global decoder G: construction

$\mathcal{D}_1$, assignment 1



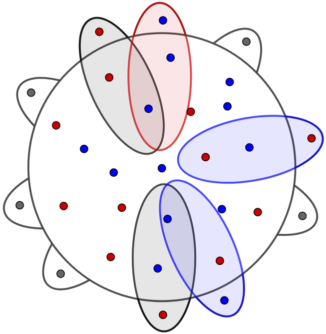
Global decoder G: construction

$\mathcal{D}_1$, assignment 2



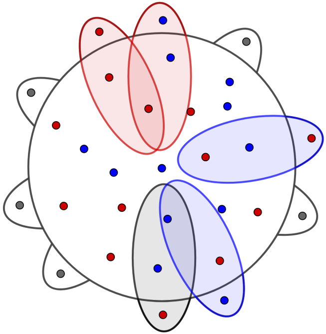
Global decoder G: construction

$\mathcal{D}_1$, assignment 3



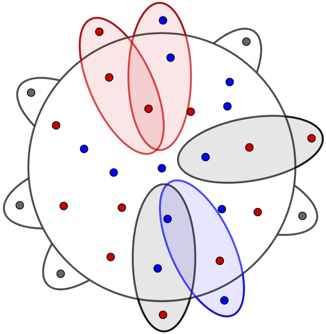
Global decoder G: construction

$\mathcal{D}_1$, assignment 4



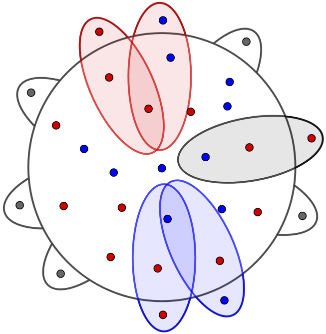
Global decoder G: construction

$\mathcal{D}_1$, assignment 5



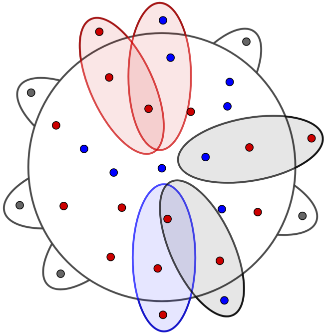
Global decoder G: construction

$\mathcal{D}_1$, assignment 6



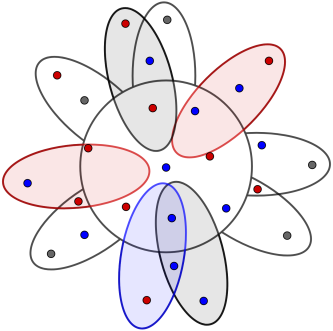
Global decoder G: construction

$\mathcal{D}_1$, assignment $2^{|K_1|}$



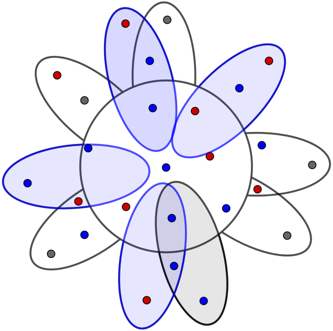
Global decoder G: construction

$\mathcal{D}_2$, assignment 1



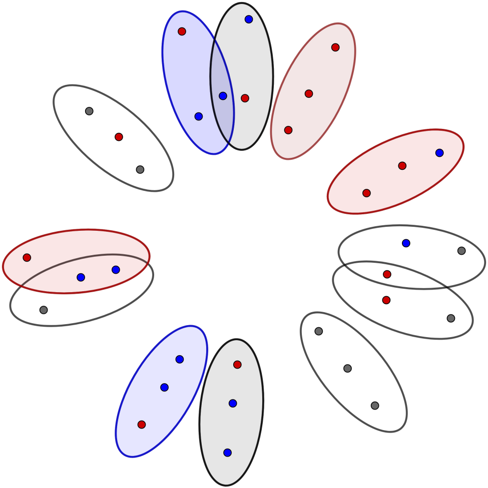
Global decoder G: construction

$\mathcal{D}_2$, assignment $\leq 2^{|K_2|}$: output $\bullet$



Global decoder G: construction

$\mathcal{D}_3$, assignment $\leq 2^{|K_3|}$



Global decoder G: analysis

Volume lemma, upper bound

For every daisy and kernel assignment κ, the *bad* q-sets $\mathcal{B}$ (that decode to the wrong value) cover a small fraction of the codeword. Thus, $|\mathcal{B}| = O(n)$.

Lemma (soundness)

For every daisy $\mathcal{D}_j$ and kernel assignment κ, the collection of bad queried q-sets satisfies $|\mathcal{B} \cap \mathcal{Q}_j| < \tau_j$ with high probability.

Global decoder G: analysis

Volume lemma, lower bound

Under the correct kernel assignment, for some daisy $\mathcal{D}_j$, the queried q-sets $\mathcal{Q}_j$ cover a large fraction of the codeword. Thus, $|\mathcal{Q}_j| = \Omega(n)$.

Lemma (completeness)

For some daisy $\mathcal{D}_j$, under the correct kernel assignment, $|\mathcal{Q}_j| \geq 2\tau_j$ with high probability. Thus, the *good* sets $\mathcal{G} = \mathcal{Q}_j \setminus \mathcal{B}$ satisfy $|\mathcal{G}| \geq \tau_j$.

Global decoder G: analysis

Lemma

For any $x \in \{0,1\}^k$, G makes $O(n^{1-\frac{1}{2q^2}})$ queries to $C(x)$ and satisfies $\mathbb{P}[G^{C(x)} = x] \geq 2/3$.

Theorem

Any RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n^{1-\frac{1}{2q^2}} = \Omega(k).$$

Global decoder G: analysis

Lemma

For any $x \in \{0,1\}^k$, G makes $O(n^{1-\frac{1}{2q^2}})$ queries to $C(x)$ and satisfies $\mathbb{P}[G^{C(x)} = x] \geq 2/3$.

Theorem

Any RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = \Omega\left(k^{1+\frac{1}{2q^2-1}}\right) = k^{1+\Omega\left(\frac{1}{q^2}\right)}.$$

Theorem [BGH+04]

There exist RLDCs with message length k and blocklength n satisfying

$$n = k^{1+O\left(\frac{1}{\sqrt{q}}\right)}.$$

Open problem

What is the largest $\alpha \in [1/2, 2]$ such that there exist RLDCs with

$$n = k^{1+\Omega\left(\frac{1}{q^{\alpha}}\right)}?$$

References

Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang.
Improved bounds for the sunflower lemma, 2019.

Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan.
Robust PCPs of proximity, shorter PCPs and applications to coding.
In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 1–10, 2004.

P. Erdös and R. Rado.
Intersection theorems for systems of sets.
Journal of the London Mathematical Society, s1-35(1):85–90, 1960.

Tom Gur and Oded Lachish.
A lower bound for relaxed locally decodable codes.
CoRR, abs/1904.08112, 2019.