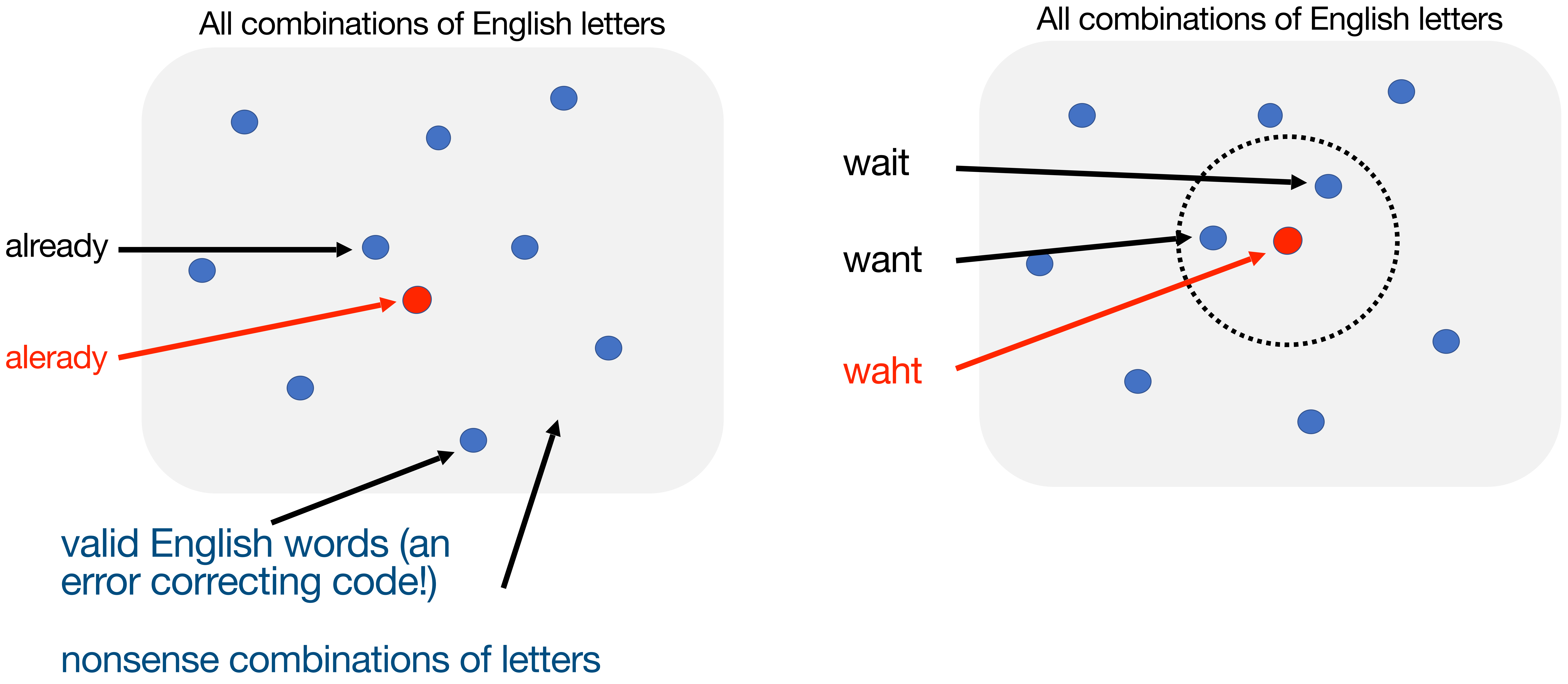
Threshold rates for error-correcting codes

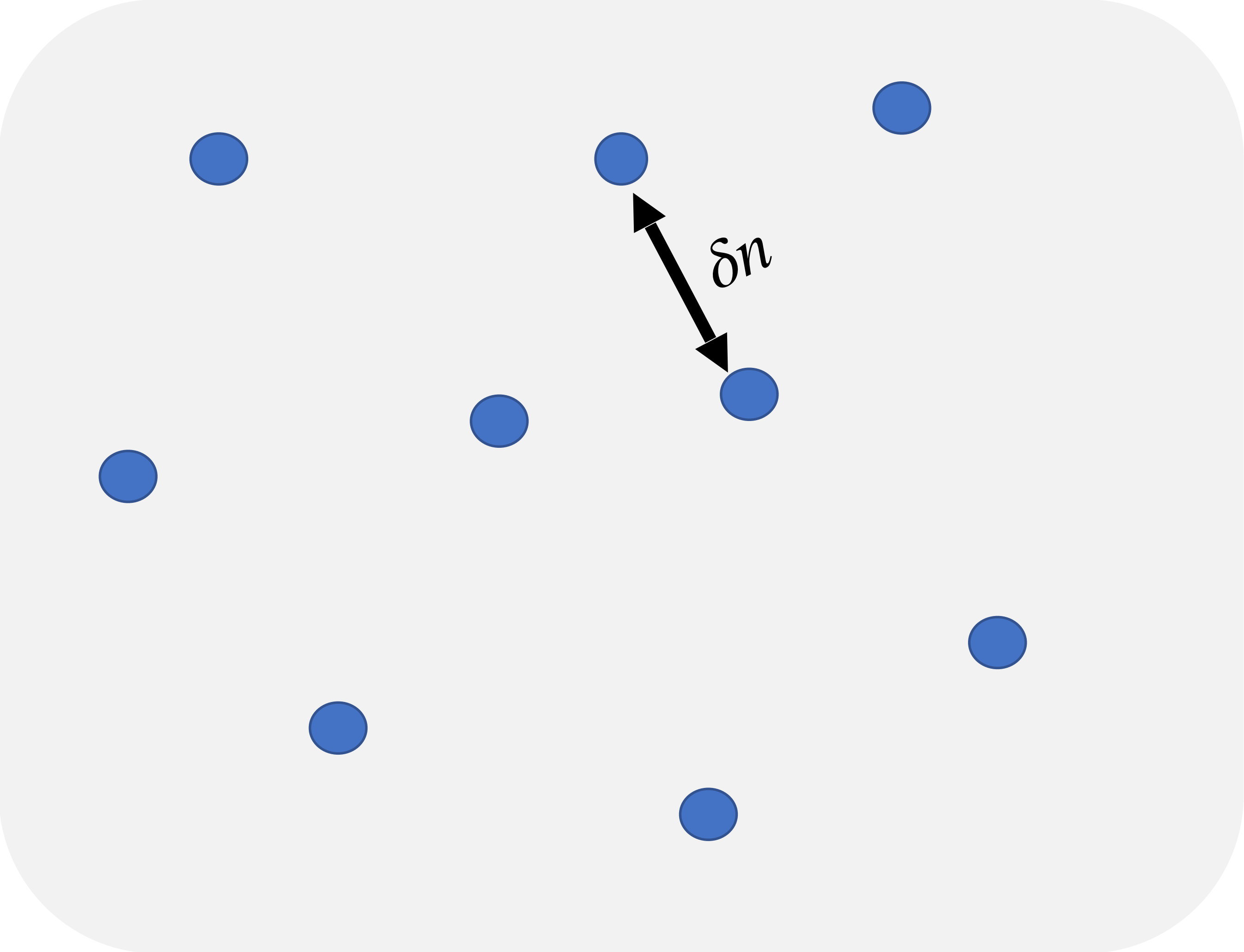
Shashwat Silas. PhD thesis defense. 02/26/2021

yuo alerady knwo waht an erorr-corecting c*de is!

How did you read that?

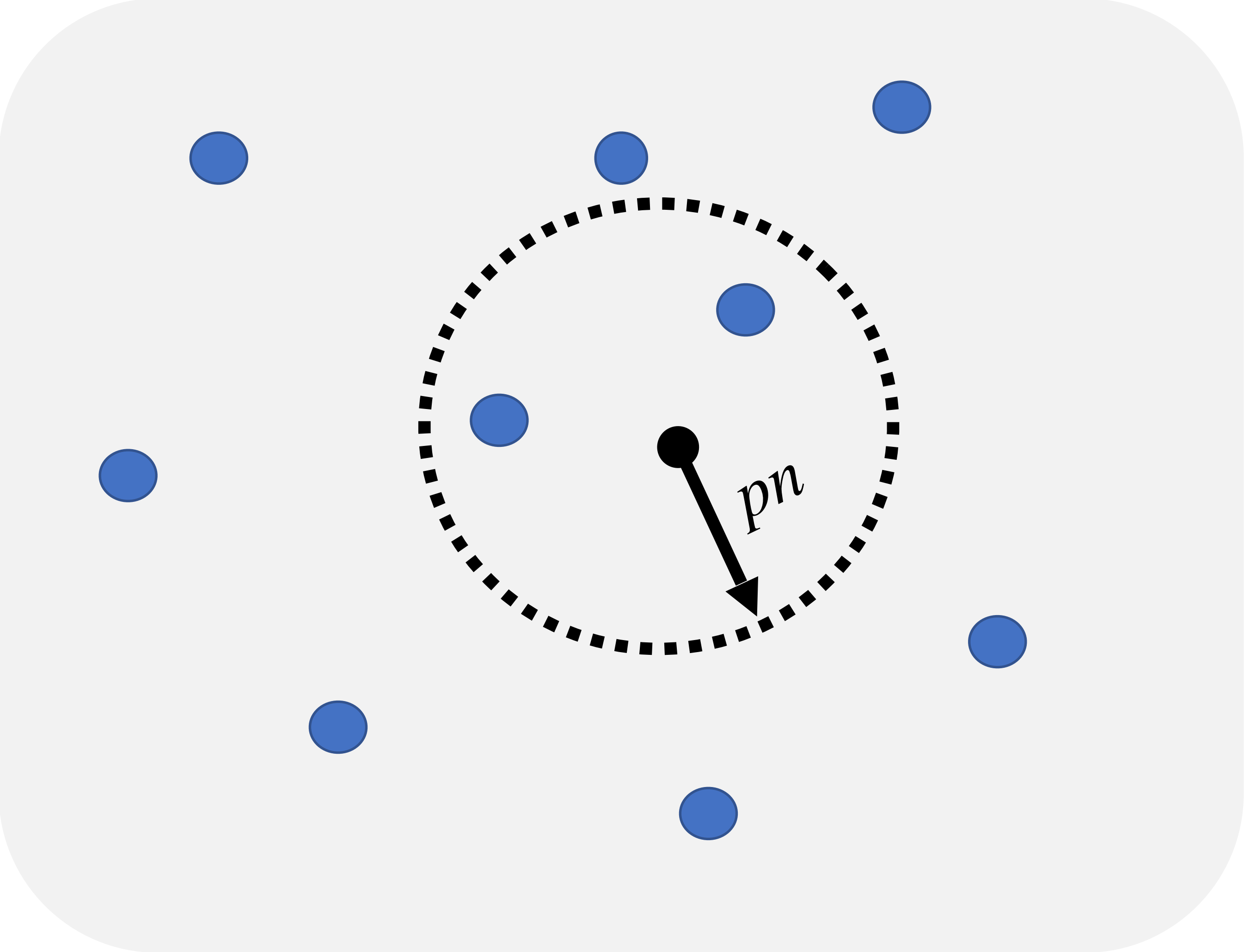


Distance and list-decodability



The closest any two legal words can be is the *distance* of a code.

High distance makes it easier to decode.



If there are $< L$ real words within distance p of any (real or not) word, then the code is *(p,L)-list decodable.*

Small L makes it easier to decode.

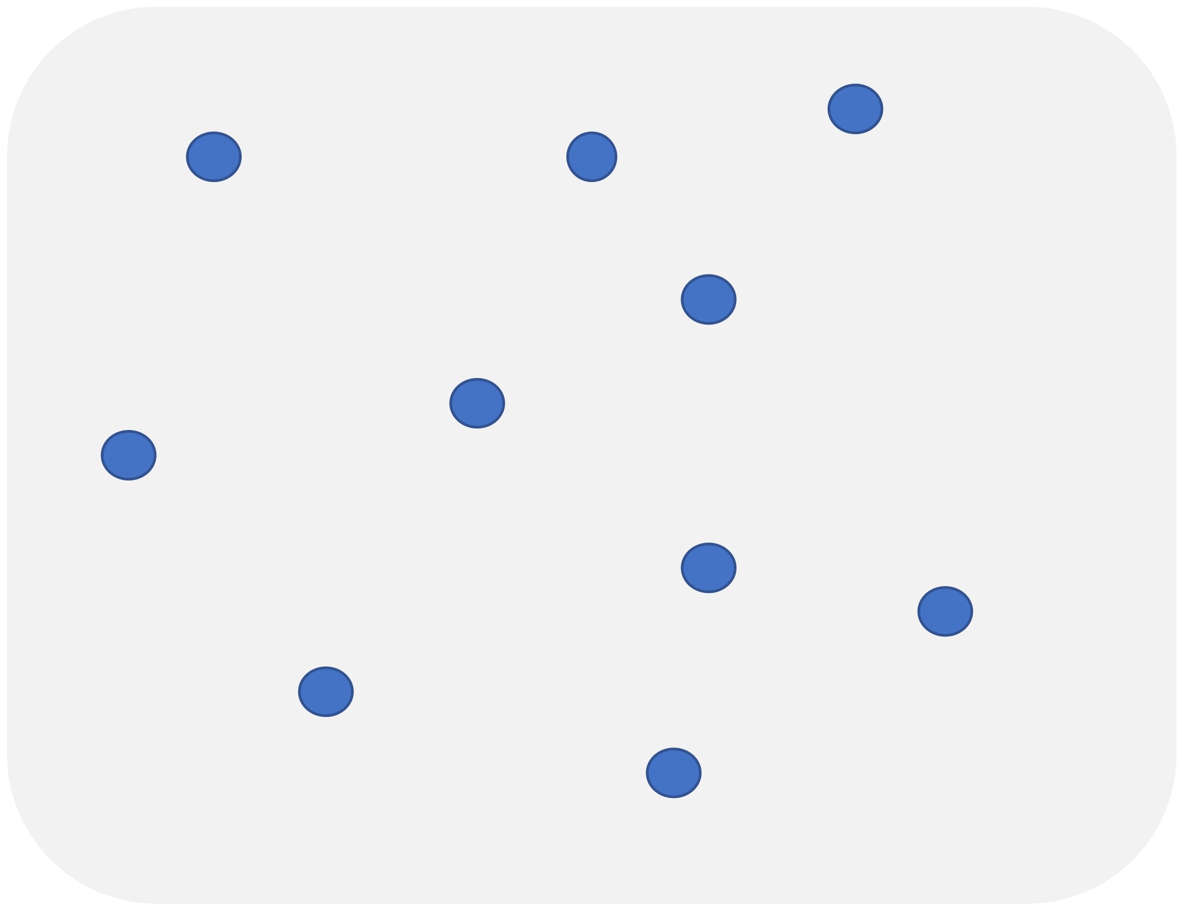
English is not a very good error-correcting code. Many real words are quite similar to each other, so we can't give *mathematical guarantees* about error correction.

Error-correcting codes

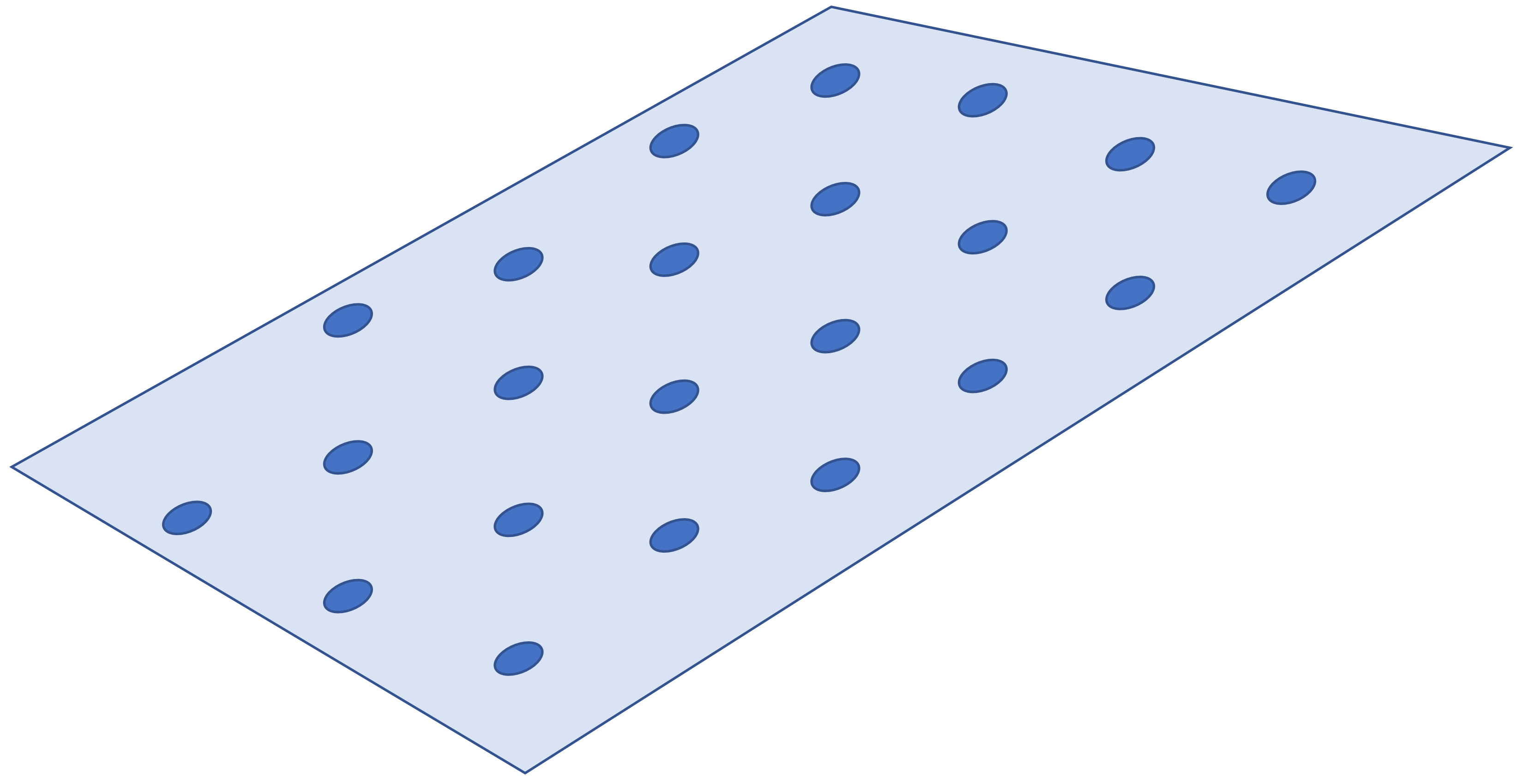
- A code C of blocklength n over an alphabet Σ is just $C \subseteq \Sigma^n$
- The rate $R = \dfrac{\log_{|\Sigma|} |C|}{n} = \dfrac{\text{symbols you want to send}}{\text{symbols you actually send}}$
- There is a trade-off between error-tolerance and rate
- We will think of $\Sigma = \mathbb{F}_q$ for q constant and $n \to \infty$
- The error is adversarial

Random codes and random linear codes

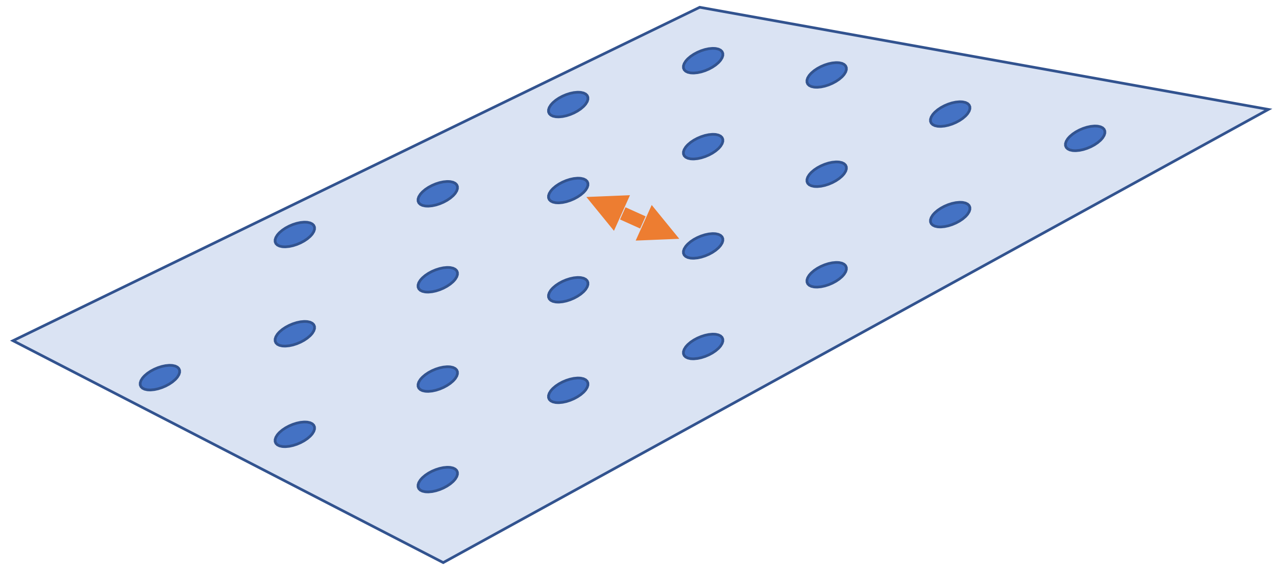
- Σ is the alphabet.
- $C \subseteq \Sigma^n$ is a subset.
- A random code (RC) of 'expected' rate *R* is chosen so that each $x \in \Sigma^n$ is included in C with probability $|\Sigma|^{-n(1-R)}$.

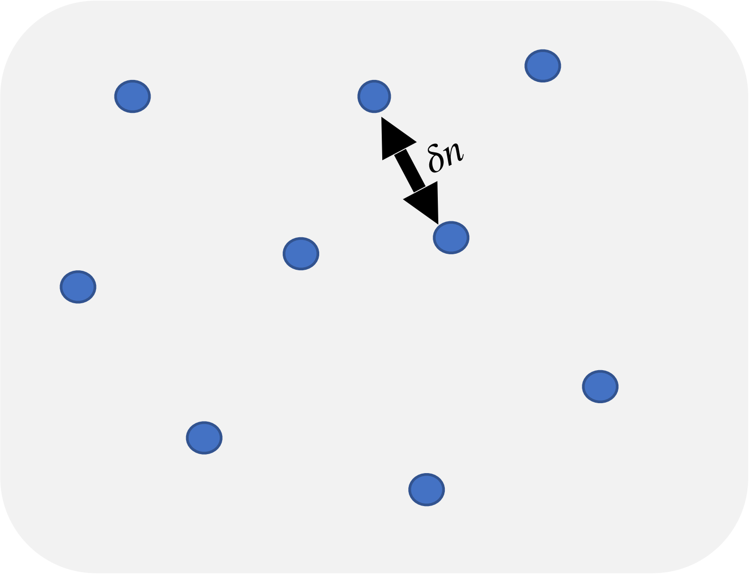


- $\mathbb{F}$ is a finite field.
  - E.g., $\mathbb{F} = \mathbb{F}_2 = \{0,1\}$ with arithmetic mod 2.
- $C \leq \mathbb{F}^n$ is a subspace.
- A random linear code (RLC) of dimension *k* is a random subspace of dimension *k*.
- Rate = *k*/*n*.

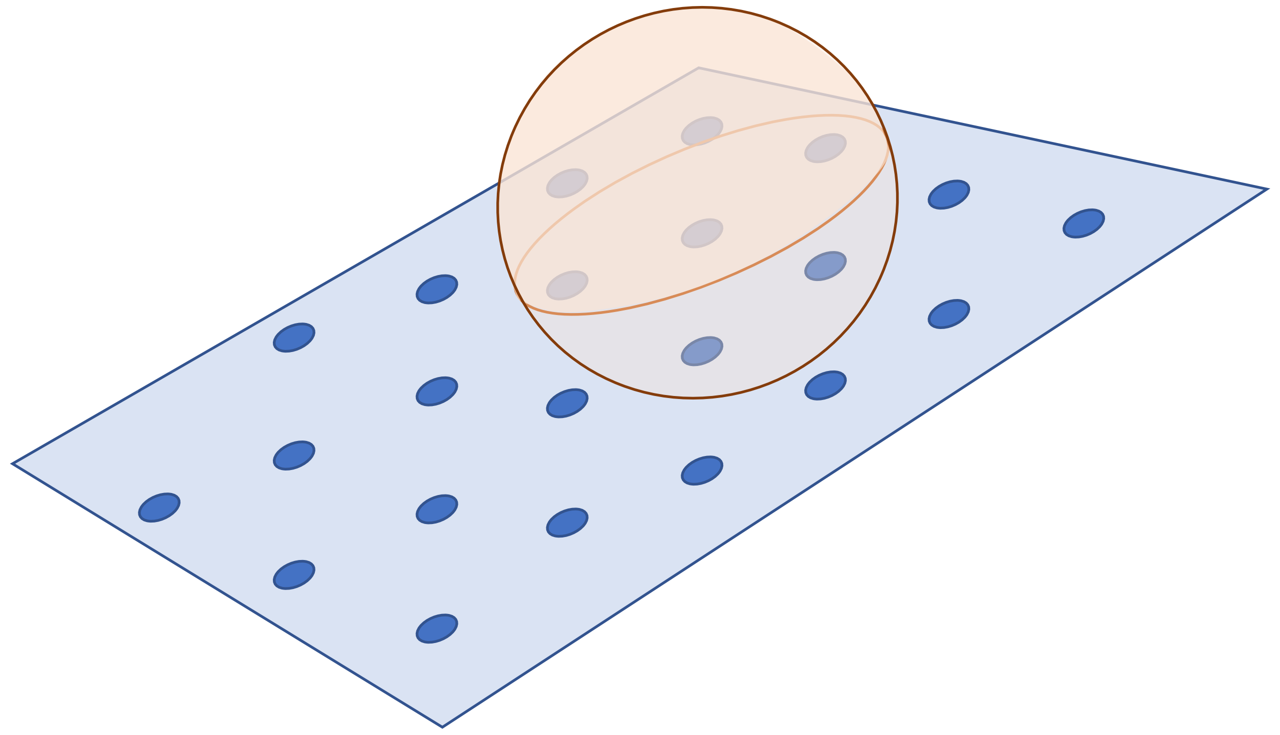


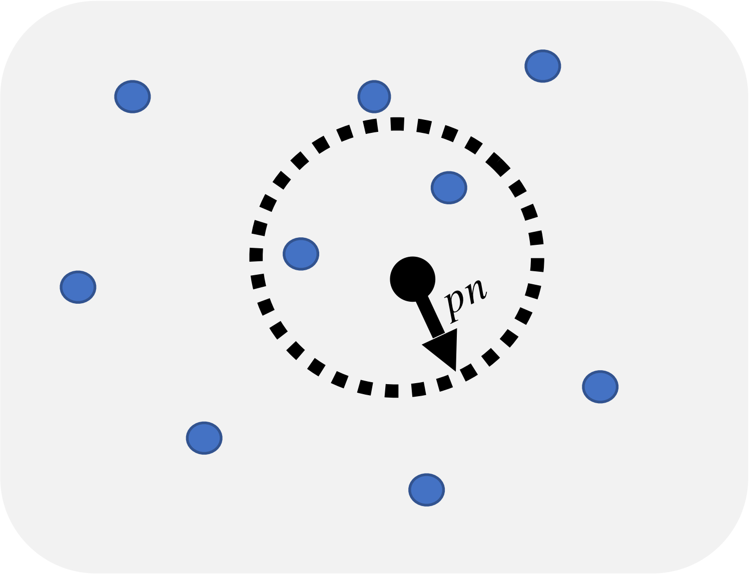
Questions about the combinatorics of codes





- What is the **distance** of a code?

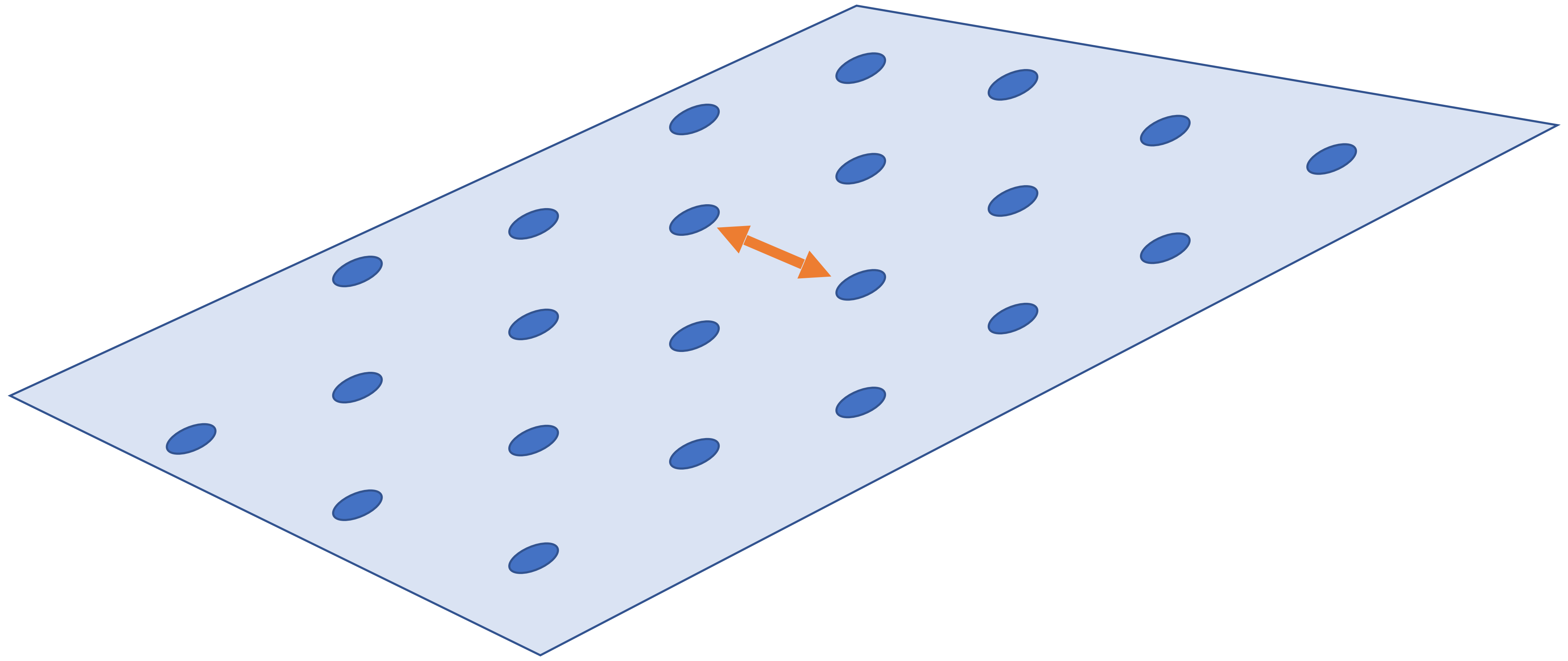




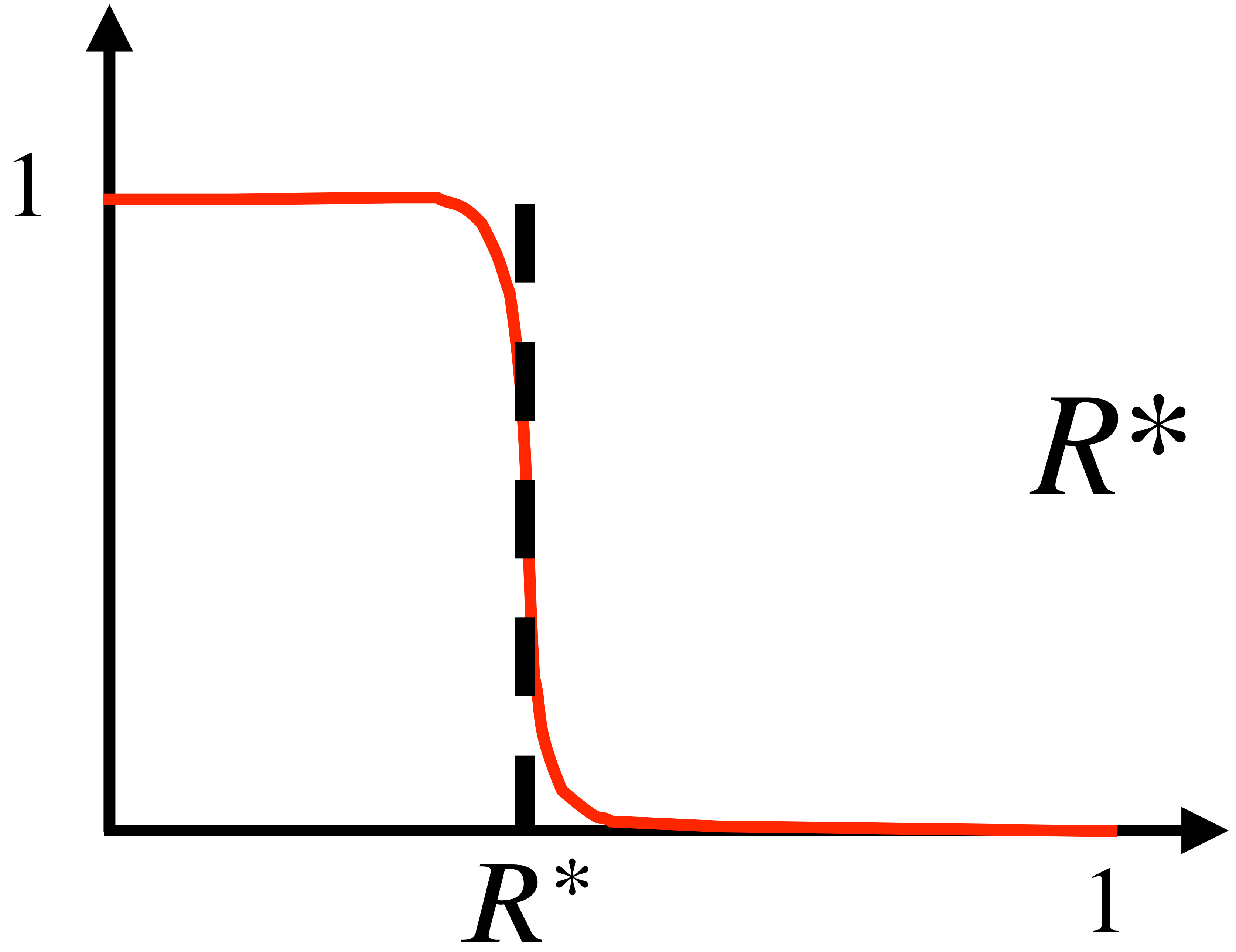
- What is the **list-decodability** of a code?

Distance of random linear codes

$$\text{Distance} = \frac{\min_{x \neq y \in C} \text{Hamming}(x, y)}{n}$$



Probability of a random k dimensional subspace having distance at least δ

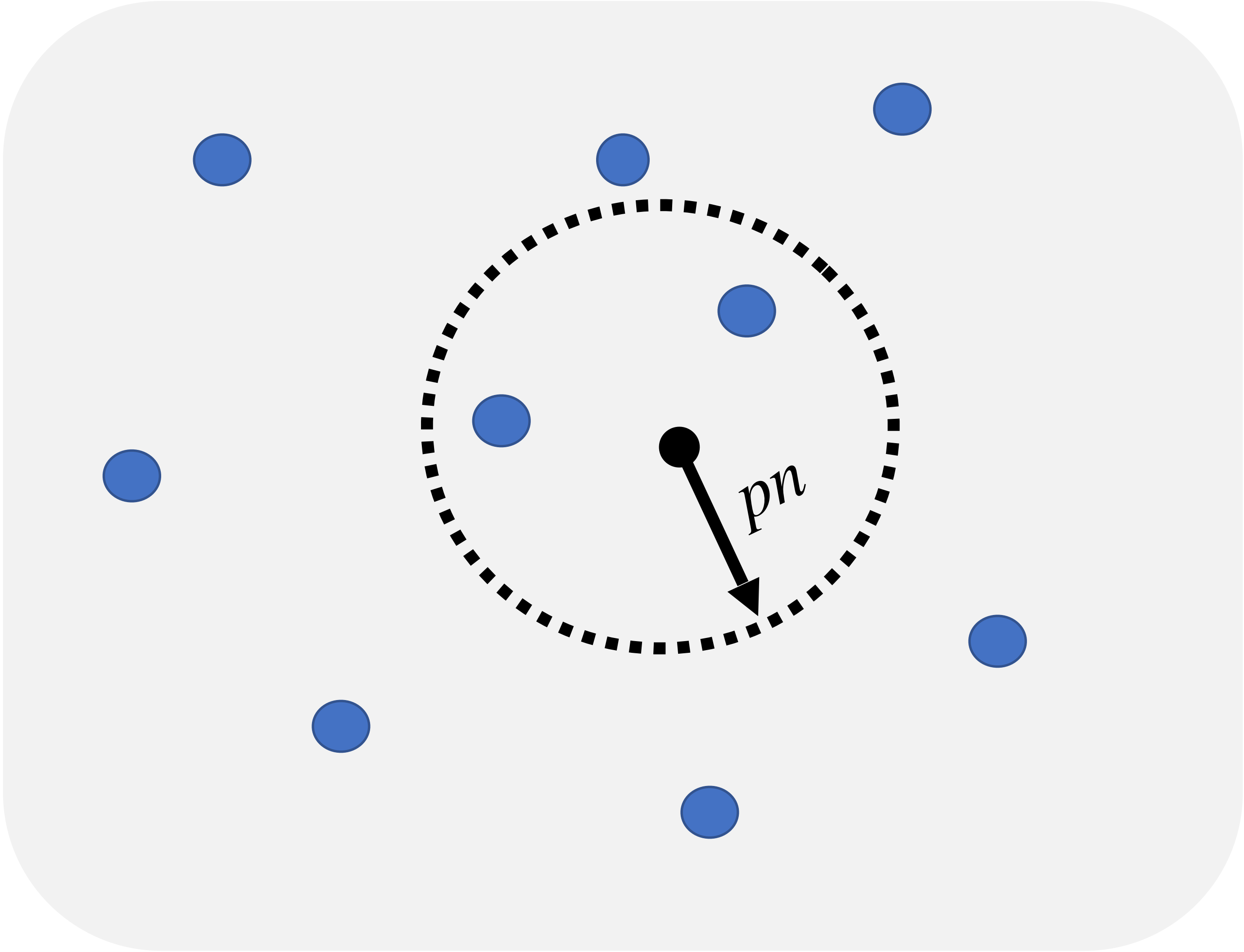


$$R^* = 1 - h_q(\delta)$$

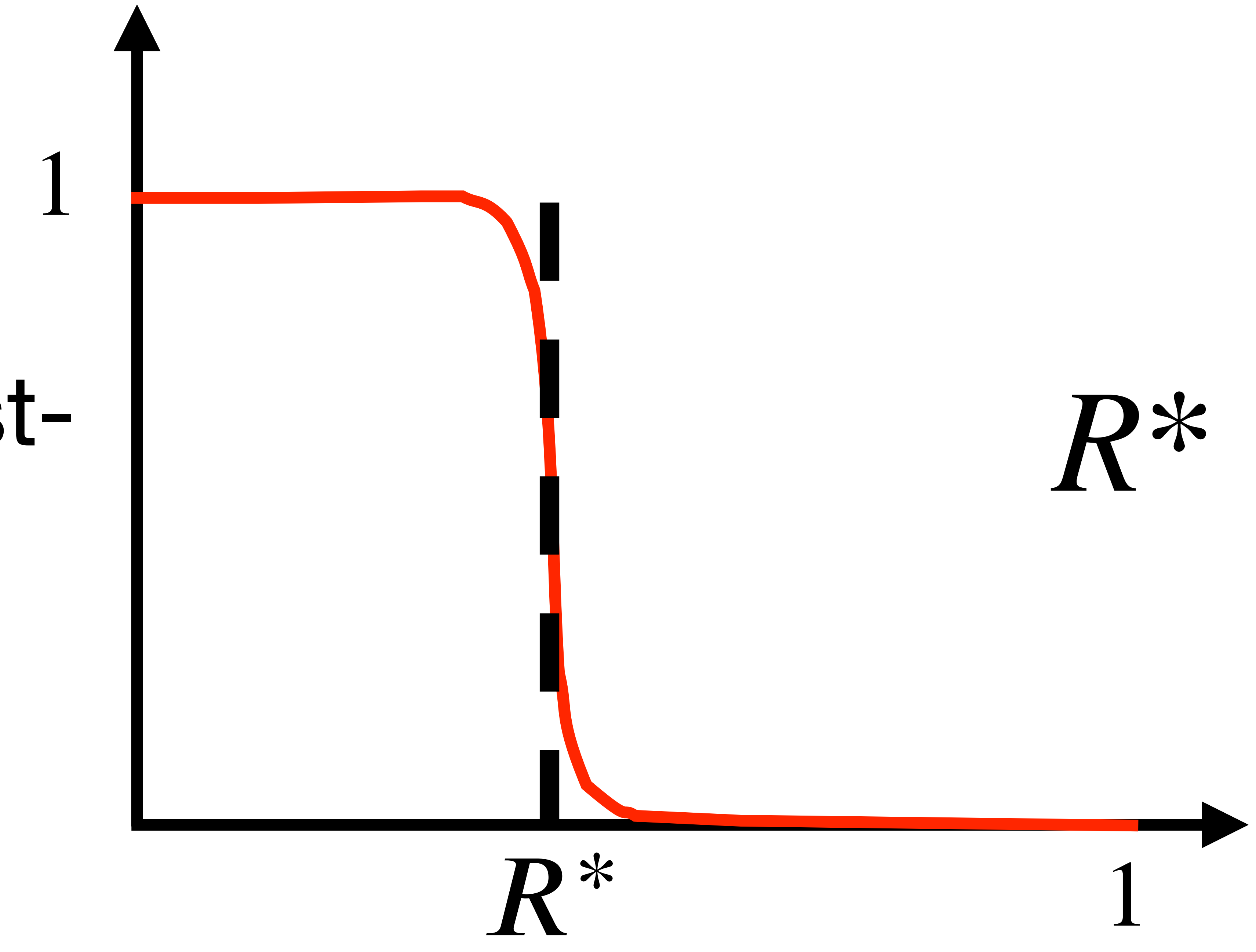
$$h_q(x) = x\log_q(q-1) - x\log_q(x) - (1-x)\log_q(1-x)$$

List-decodability of completely random codes

A code $C \subseteq \mathbb{F}_q^n$ is (p, L)-list decodable if for all $x \in \mathbb{F}_q^n$, $|B_{pn}(x) \cap C| < L$.

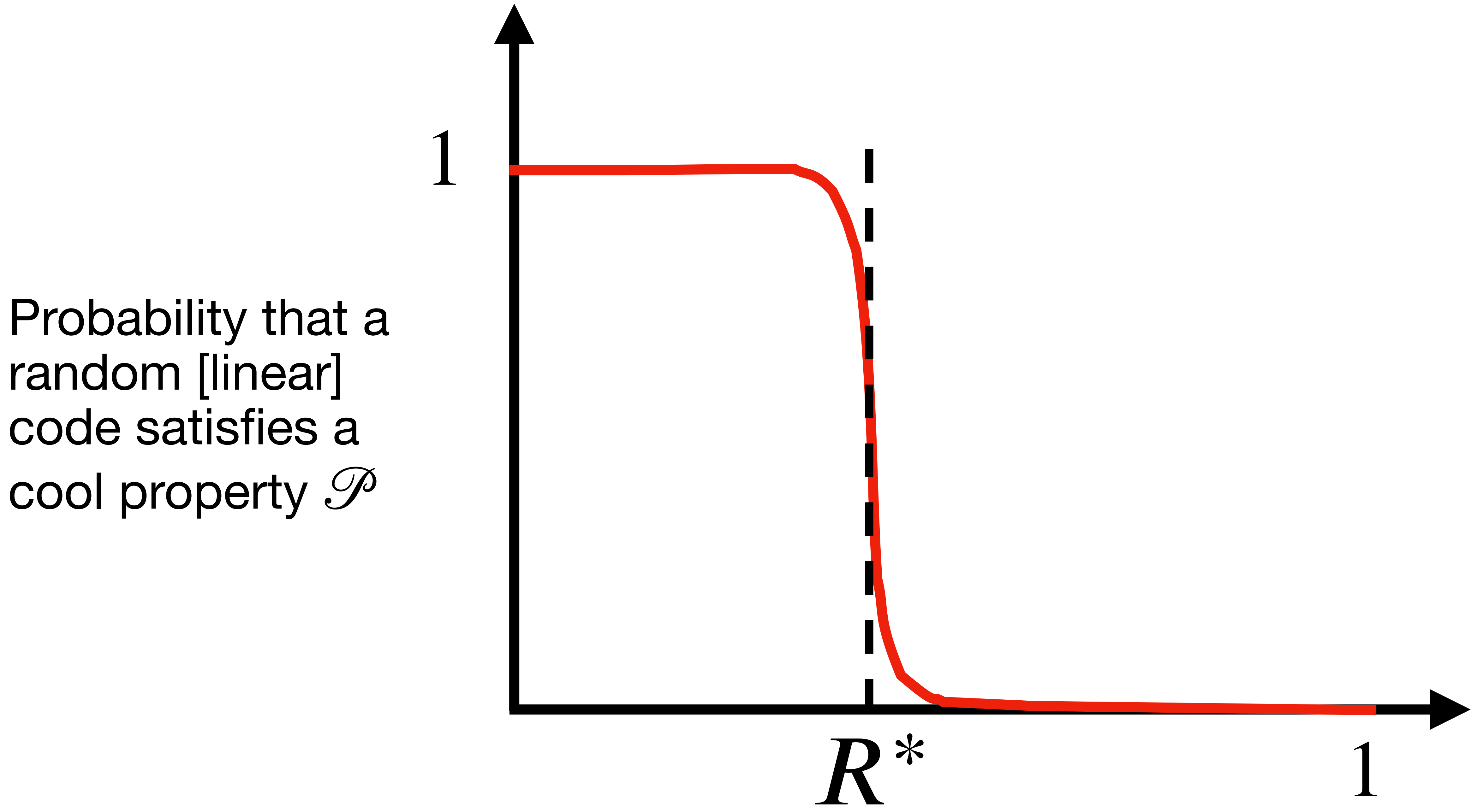


Probability of a random code being $(p, O(1))$-list-decodable



$$R^* = 1 - h_q(p)$$

Threshold rates



- If $R \leq R^* - \varepsilon$, then random [linear] code satisfies property w.h.p.
- If $R \geq R^* + \varepsilon$, then random [linear] code does not satisfy property w.h.p.

PART I: Informal results

A. Characterization theorems

B. Some applications

PART II: Proof outline for RLCs

A. Local properties

B. Threshold for containing a type

PART III: Formal results for RC and RLC

A. Characterization theorem for RLCs

B. Characterization theorem for RCs

PART IV: LDPC Codes

A. Definitions

B. Reduction

PART I: Informal results

A. Characterization theorems

B. Some applications

PART III: Formal results for RC and RLC

A. Characterization theorem for RLCs

B. Characterization theorem for RCs

PART II: Proof outline for RLCs

A. Local properties

B. Threshold for containing a type

PART IV: LDPC Codes

A. Definitions

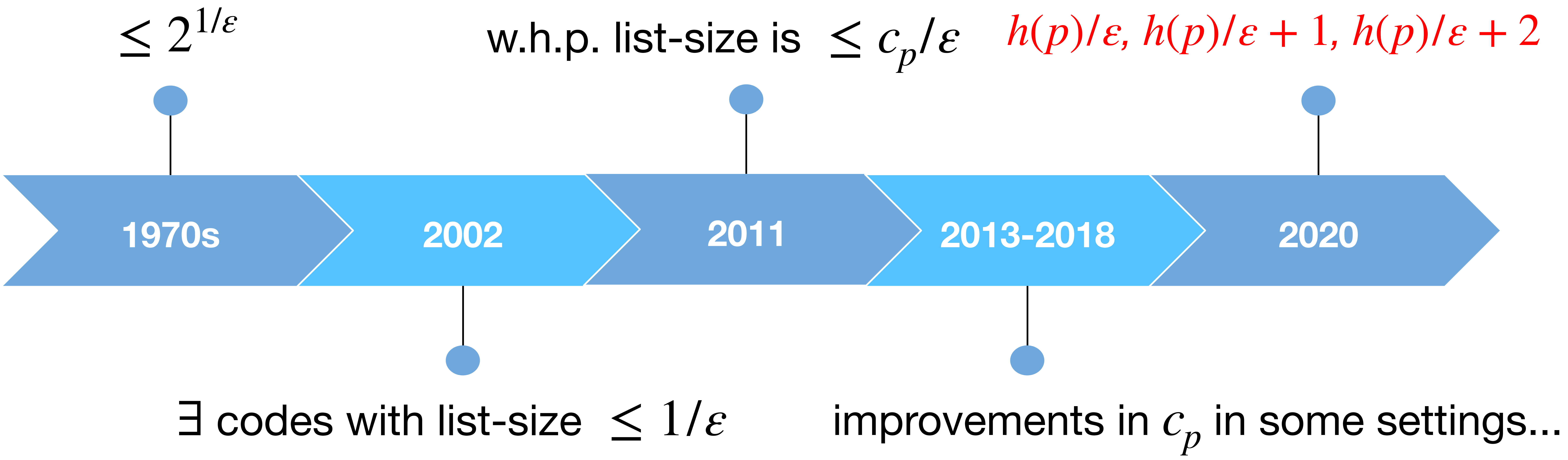
B. Reduction

A. Characterization theorems

1. All local properties of RLCs have a threshold rate and we characterize it.
2. All symmetric properties of RCs have a threshold rate and we characterize it.
3. Both local and symmetric are broad classes of properties, and include distance, list-decodability and many natural properties.
4. We show that LDPC codes achieve every local property a random linear code achieves.

B. Some applications

What is the list-size of a binary RLC of rate *R* = *1* - *h(p)* - *ε*?



[ZP81], [GHSZ02], [GHK11], [CGV13], [Woo13], [RW18], [LW18], [GLMRSW20] and others

B. Some applications

- List-size for list-recovery of a random linear code of rate $R^* - \varepsilon$ is $\ell^{\Omega(1/\varepsilon)}$
- The threshold rate for a random code to be a perfect hashing code is

$$R^* = \frac{1}{q} \log_q \left(\frac{1}{1 - q!/q^q} \right)$$

- Threshold rates for $(p,3)-$list decodability.
- Further results about list-recovery of random codes

PART I: Informal results

A. Characterization theorems

B. Some applications

PART III: Formal results for RC and RLC

A. Characterization theorem for RLCs

B. Characterization theorem for RCs

PART II: Proof outline for RLCs

A. Local properties

B. Threshold for containing a type

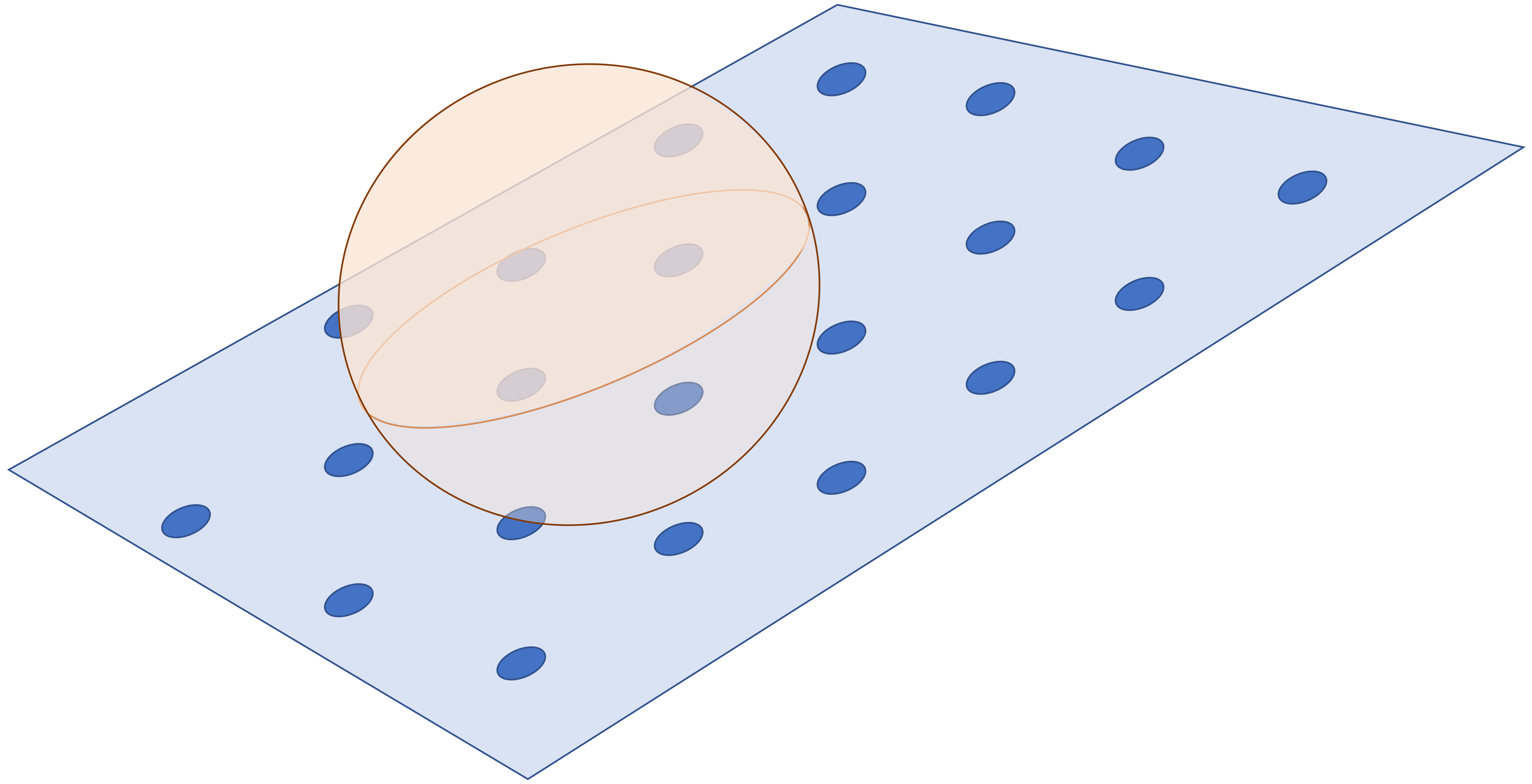
PART IV: LDPC Codes

A. Definitions

B. Reduction

A. Local properties

- Many code properties are satisfied $\Longleftrightarrow$ no bad set of vectors lies in the code.
- E.g. code (p, L)-list dec $\Longleftrightarrow$ contains no bad set of L vectors in a radius p ball.



A. Local properties

- Group these bad sets of codewords which define property into collections of bad types (special distributions).

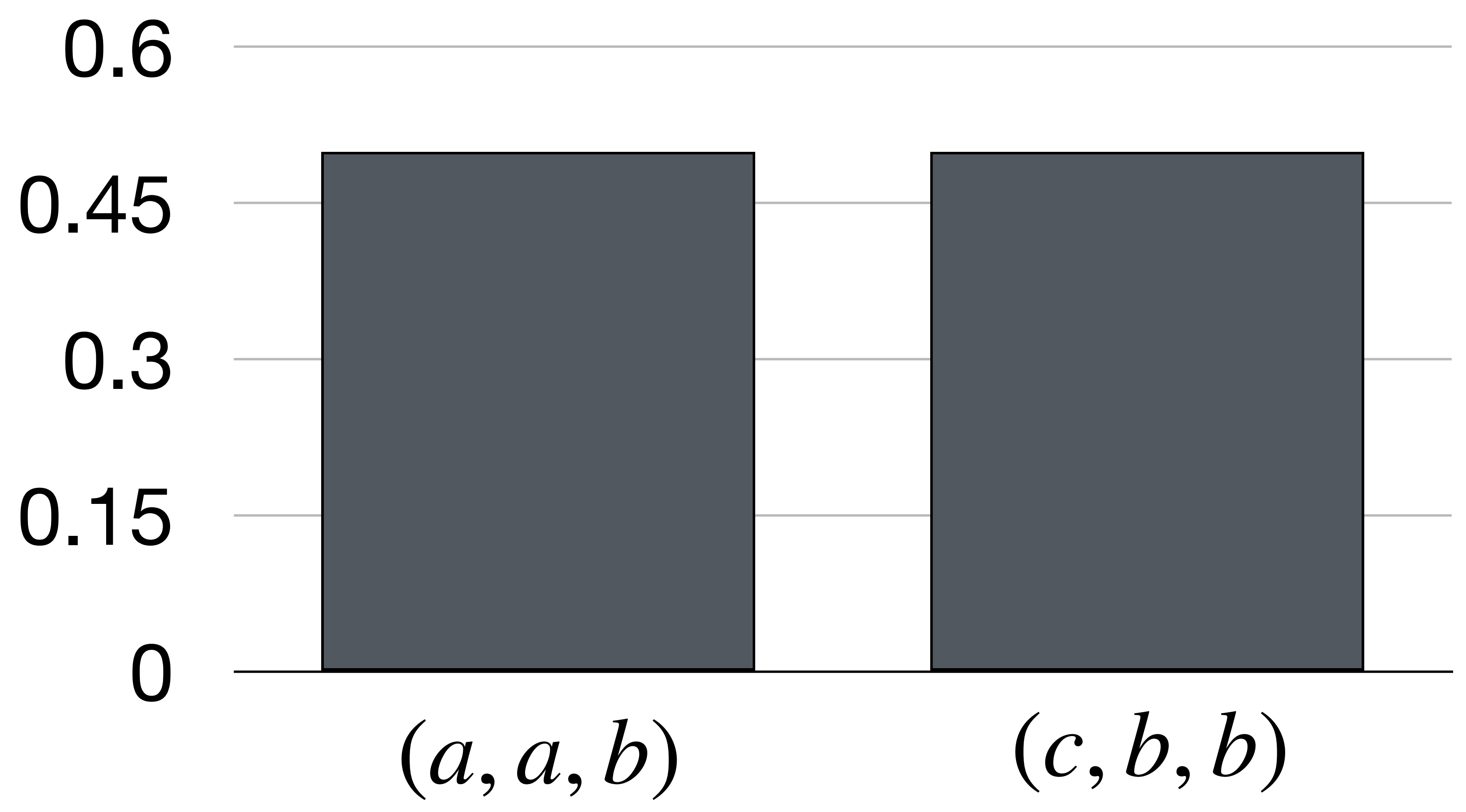
- Then: property is satisfied $\iff$ no set of codewords with such a bad type is in code.

$$b_1 = \begin{bmatrix} a \\ a \\ \vdots \\ a \\ c \\ \vdots \\ c \\ c \end{bmatrix} \quad b_2 = \begin{bmatrix} a \\ a \\ \vdots \\ a \\ b \\ \vdots \\ b \\ b \end{bmatrix} \quad b_3 = \begin{bmatrix} b \\ b \\ \vdots \\ b \\ b \\ \vdots \\ b \\ b \end{bmatrix} \longrightarrow B = \begin{bmatrix} a & a & b \\ a & a & b \\ & \vdots & \\ a & a & b \\ c & b & b \\ & \vdots & \\ c & b & b \\ c & b & b \end{bmatrix}$$

PART II: Proof outline for RLCs

A. Local properties (Types)

$$B = \begin{bmatrix} a & a & b \\ a & a & b \\ & \vdots & \\ a & a & b \\ c & b & b \\ & \vdots & \\ c & b & b \\ c & b & b \end{bmatrix} \quad B' = \begin{bmatrix} a & a & b \\ a & a & b \\ & \vdots & \\ c & b & b \\ a & a & b \\ & \vdots & \\ c & b & b \\ c & b & b \end{bmatrix} \quad B'' = \begin{bmatrix} a & a & b \\ c & b & b \\ & \vdots & \\ a & a & b \\ c & b & b \\ & \vdots & \\ a & a & b \\ c & b & b \end{bmatrix}$$

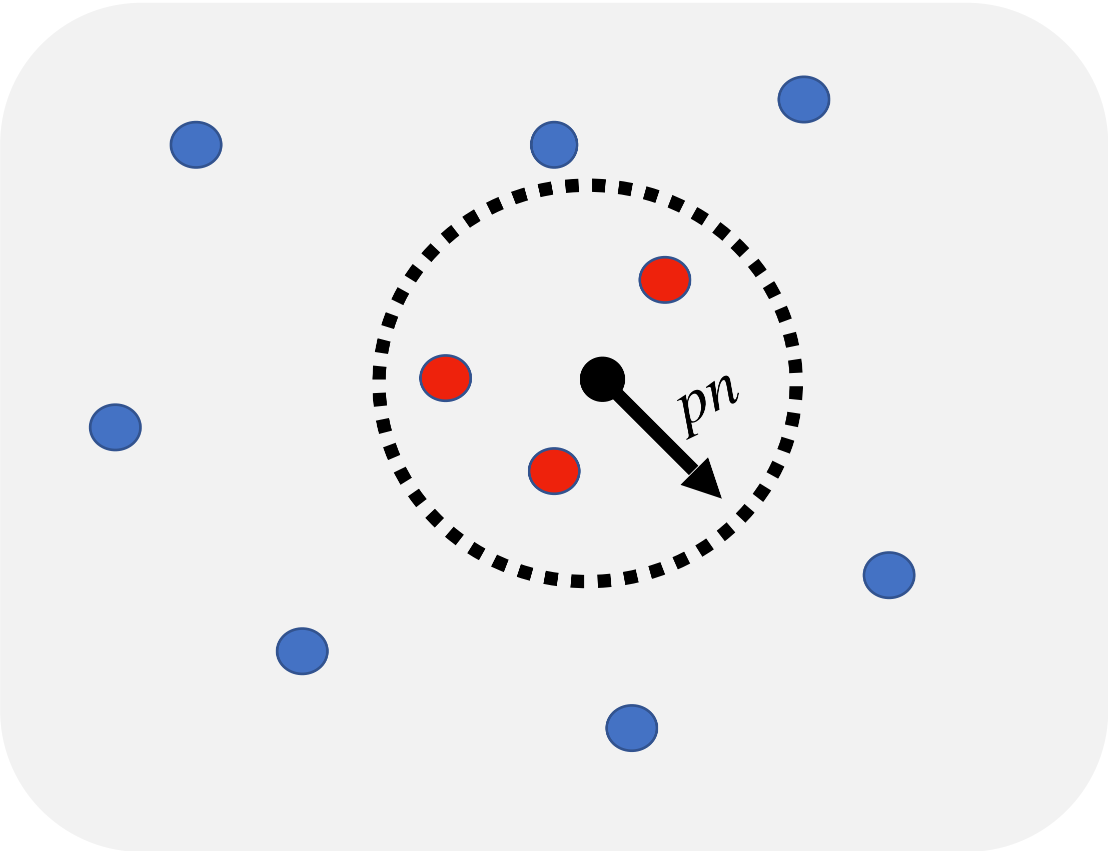


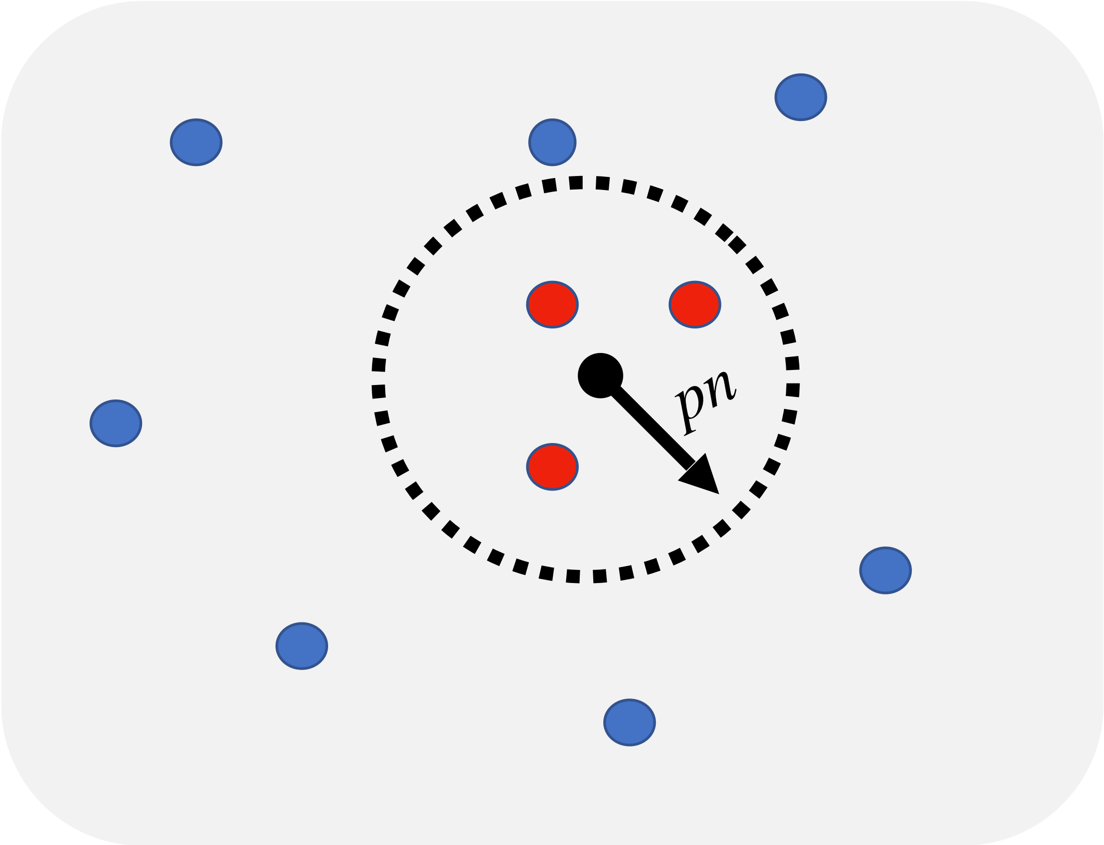
- Two matrices B, B' are the same type if they are row permutations of each other.
- A type is the empirical distribution of the rows of a matrix.
- Here, $\text{type}(B) = \text{type}(B') = \text{type}(B'') = \beta$ is a distribution over Σ^3 such that $\beta(a, a, b) = \beta(c, b, b) = 0.5$ and $\beta(x) = 0$ for all other x in Σ^3.

PART II: Proof outline for RLCs

A. Local properties

$$B = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ & \vdots & \\ & \vdots & \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \pi B = \pi \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ & \vdots & \\ & \vdots & \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} \end{bmatrix} \qquad \pi x = \pi \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

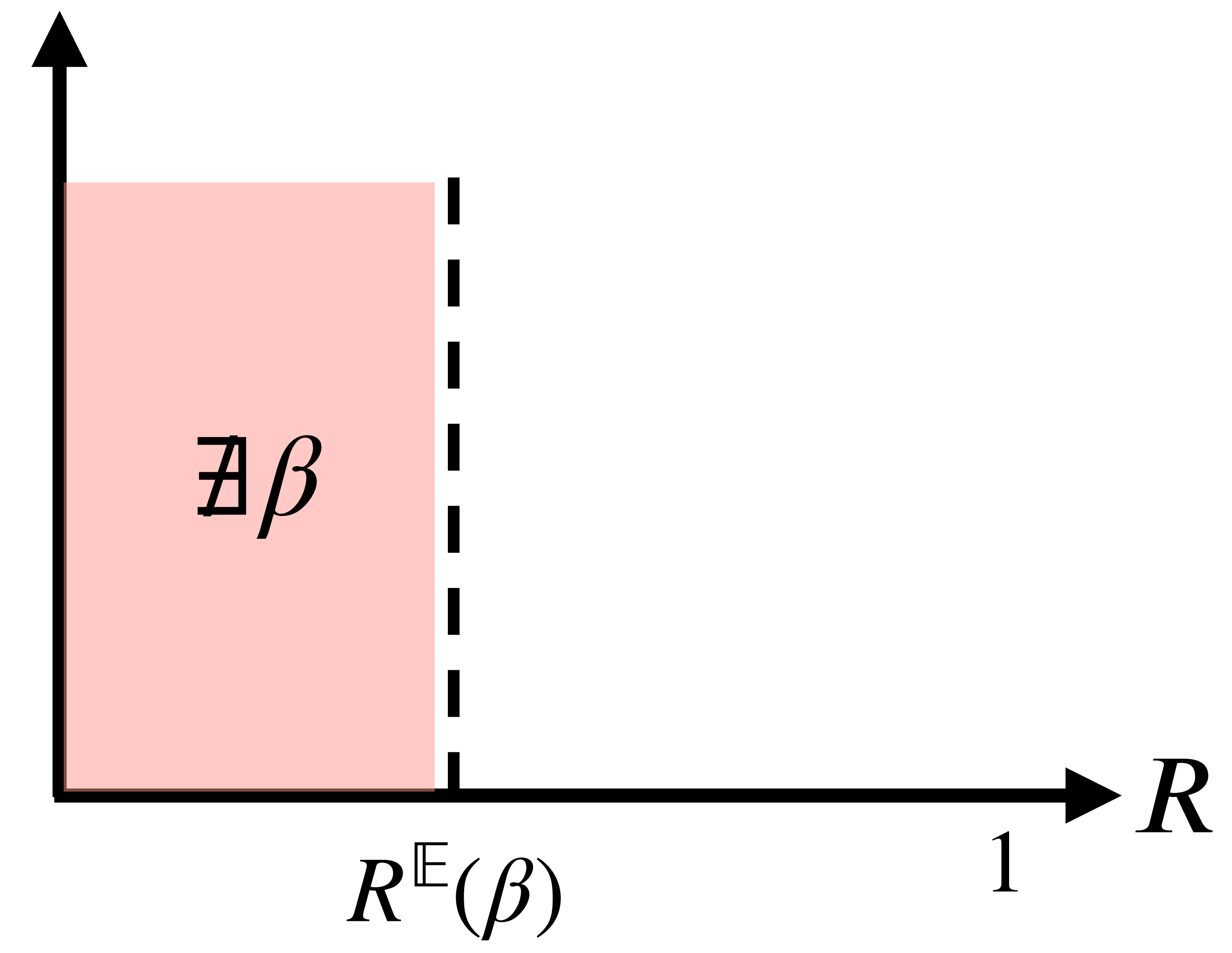




- An *ℓ-local property $\mathscr{P}$* is defined by a set of bad types T over $\mathbb{F}_q^\ell$.
- $\mathscr{P}$ is satisfied $\iff$ no bad type from T is in code.

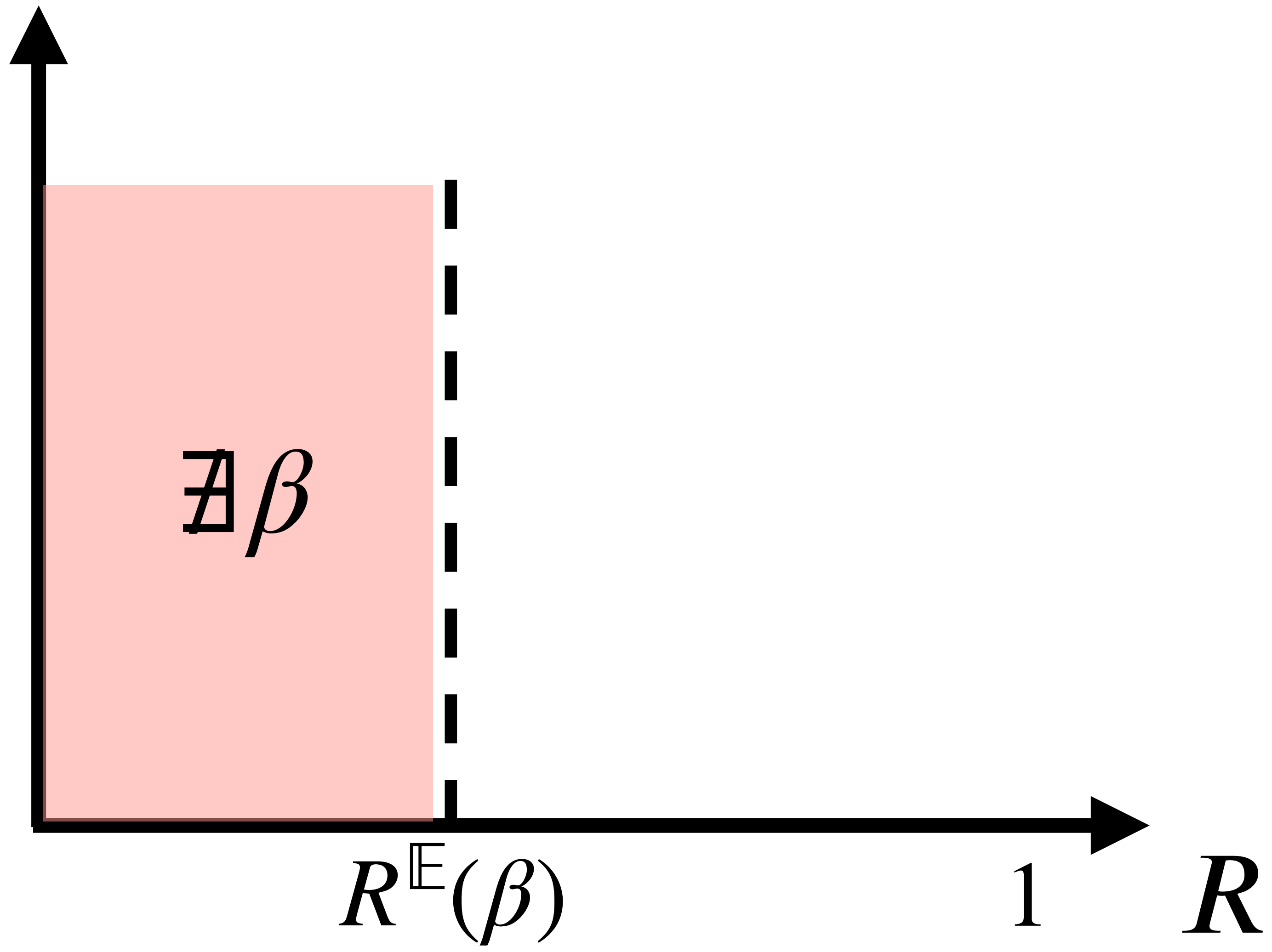
B. Threshold for containing a type

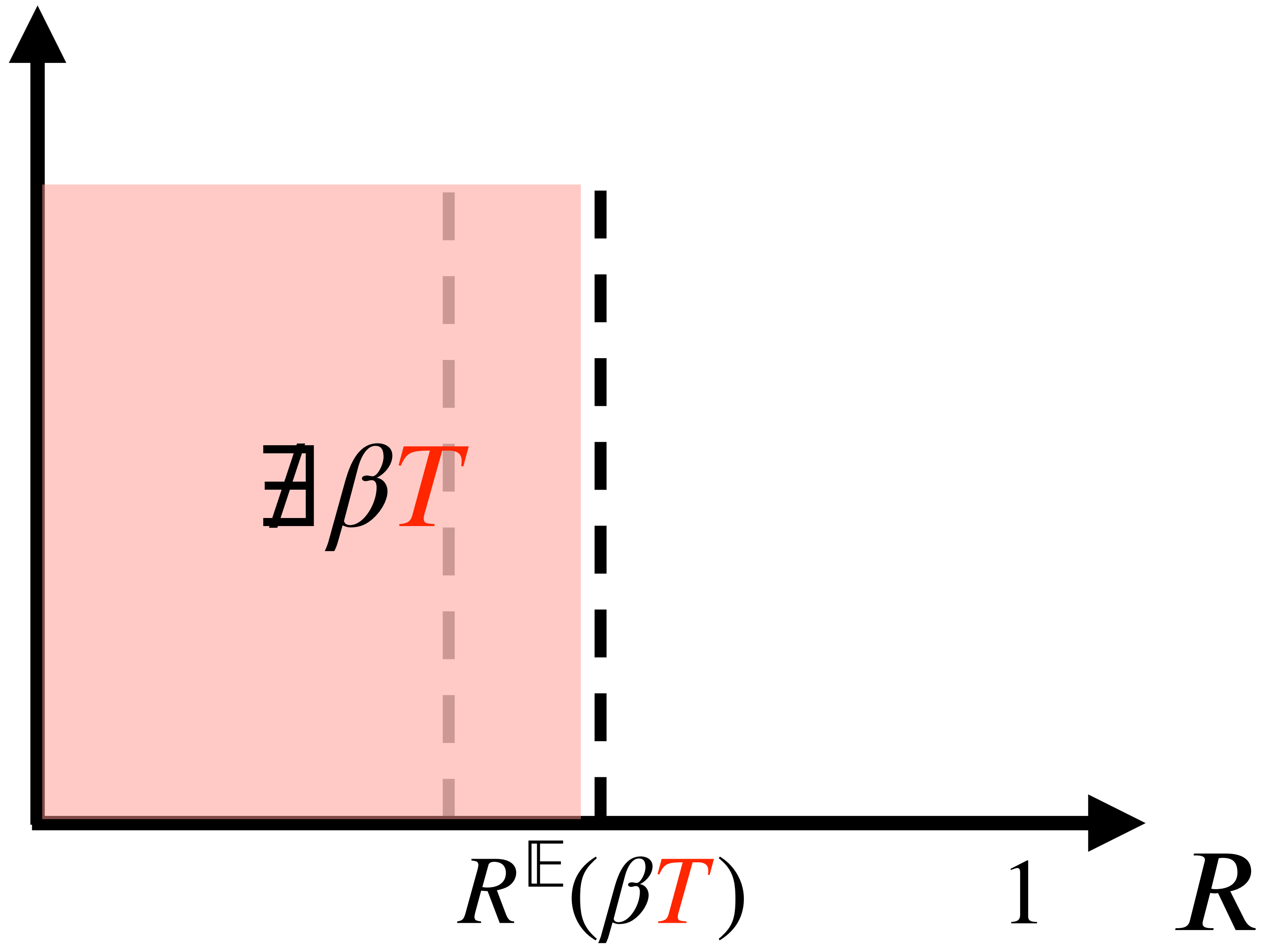
- Let C be a random linear code of rate R over $\mathbb{F}_q^n$
- If B is an $n \times \ell$ matrix of full rank, then $\Pr(B \subset C) = q^{-n\ell(1-R)}$
- Say that B had type β
- By union bound, $\Pr(\exists M \subset C \text{ of type } \beta) \leq q^{n(H_q(\beta)-(1-R)\ell)}$
- This is $o(1)$ if $R \leq 1 - \dfrac{H_q(\beta)}{\ell} - \varepsilon$ for $\varepsilon > 0$
- We define $1 - \dfrac{H_q(\beta)}{d(\beta)} = R^{\mathbb{E}}(\beta)$



B. Threshold for containing a type

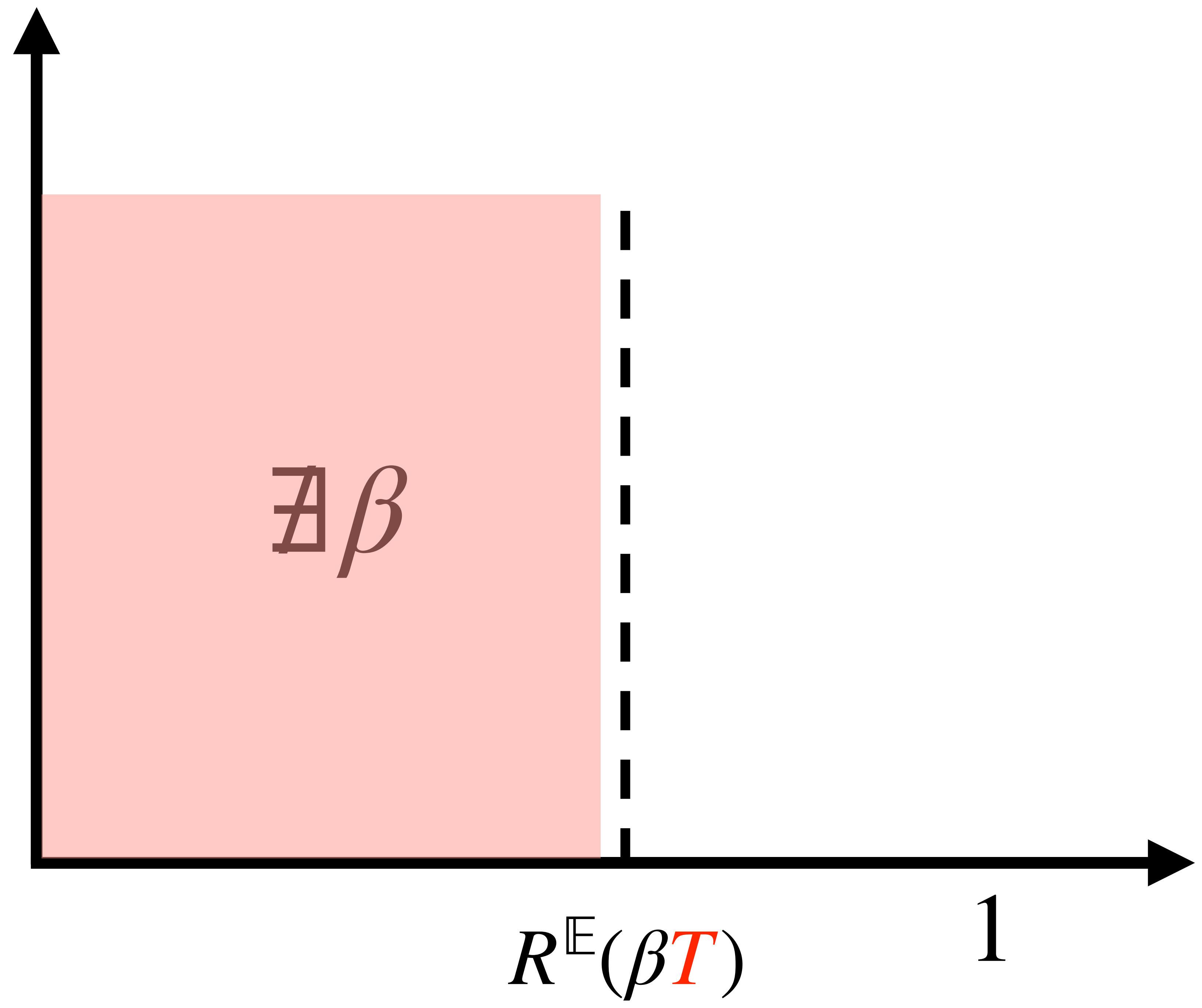
$$B = \begin{bmatrix} a & a & b \\ a & a & b \\ & \vdots & \\ a & a & b \\ c & b & b \\ & \vdots & \\ c & b & b \\ c & b & b \end{bmatrix} \text{ of type } \beta \quad \Longrightarrow \quad BT = \begin{bmatrix} a & a & b \\ a & a & b \\ & \vdots & \\ a & a & b \\ c & b & b \\ & \vdots & \\ c & b & b \\ c & b & b \end{bmatrix} \Big[\; T \;\Big] \text{ of type } \beta T$$





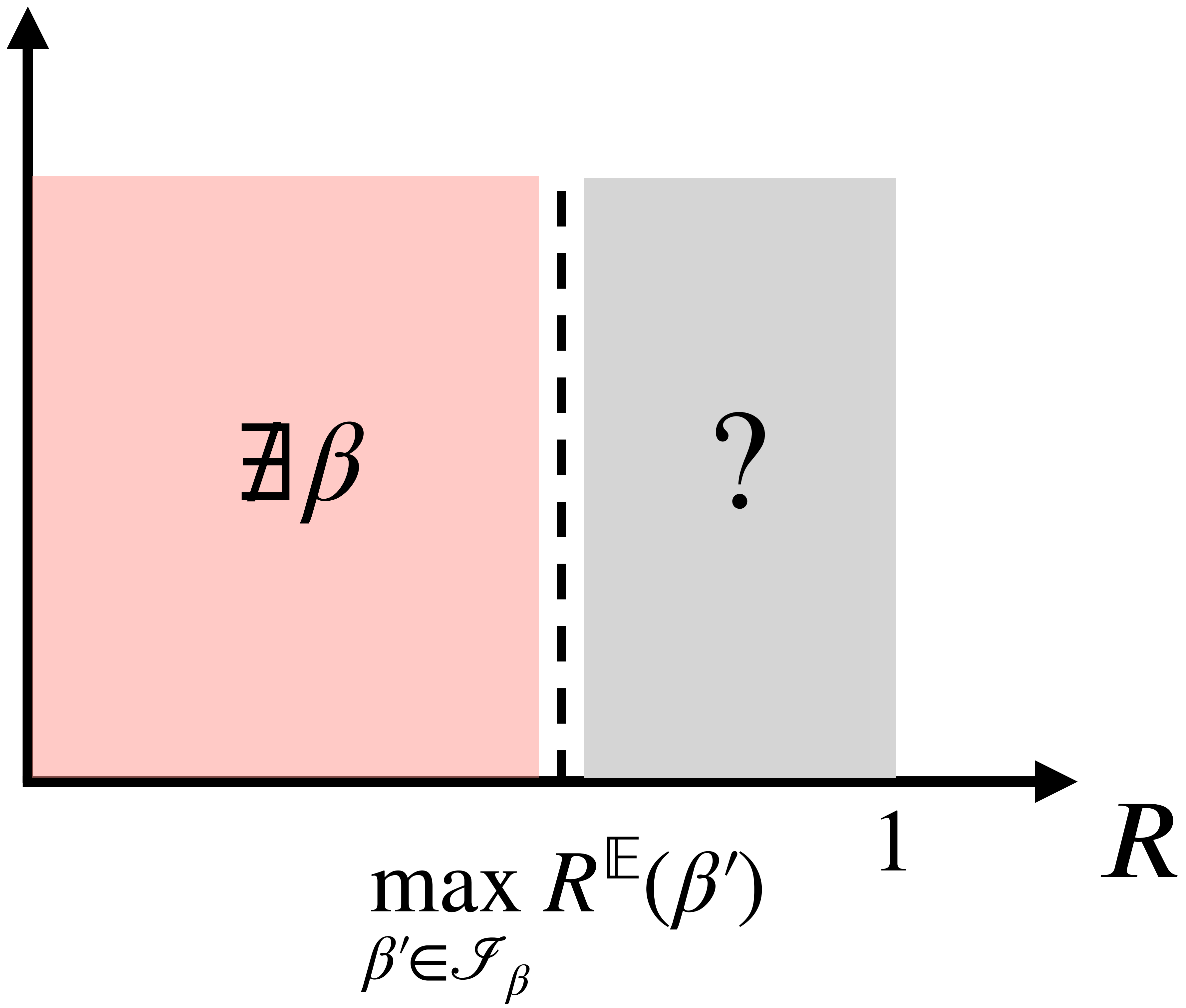
If you cannot find βT in the code, you certainly cannot find β in the code.

B. Threshold for containing a type (implied types)

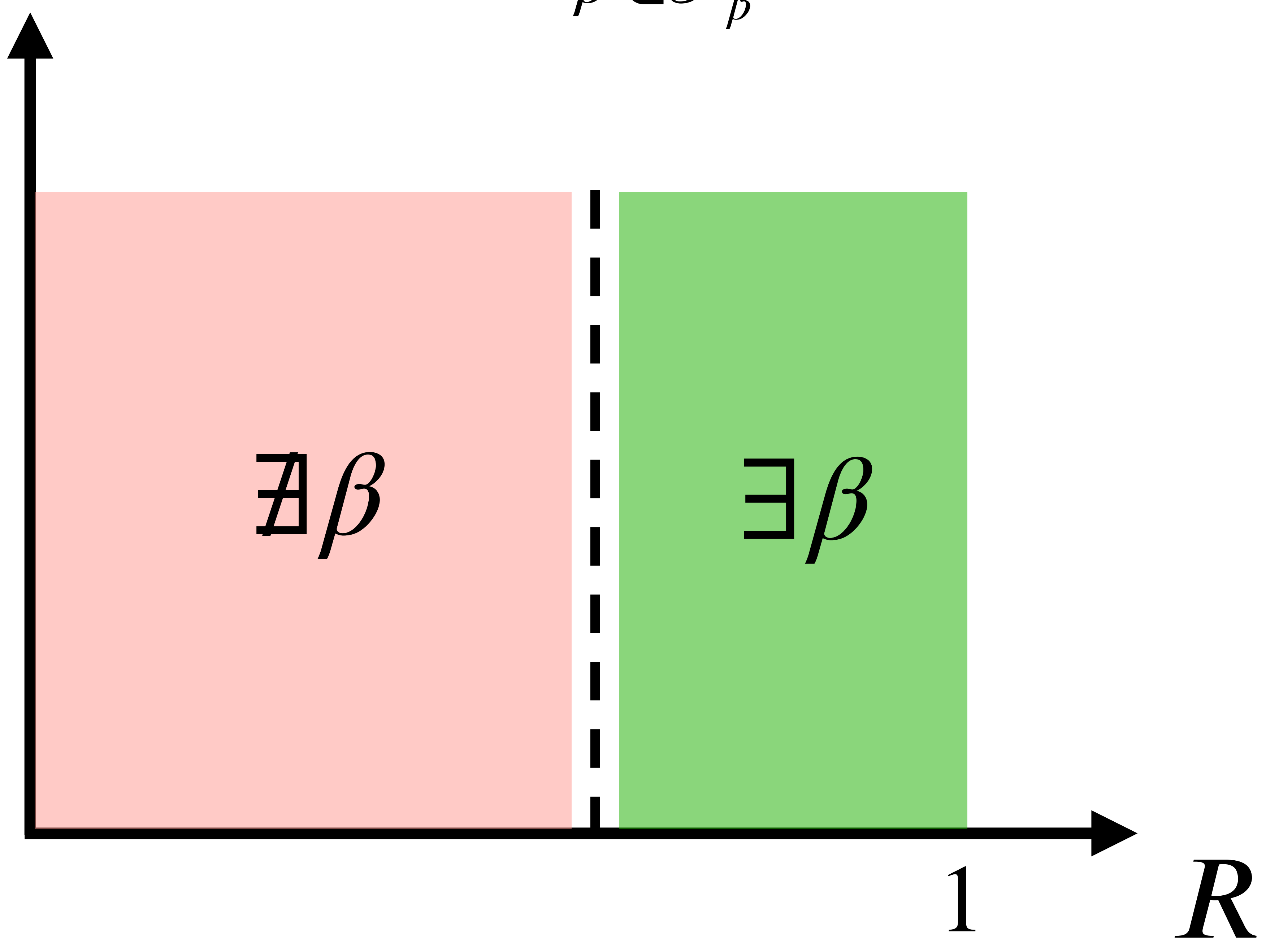


- If we want to compute the largest R such that β is unlikely to be in the code, we need at least to account for $R^{\mathbb{E}}(\beta T)$ for all T.
- We denote the set of all βT, which are the 'implied types of β', by $\mathcal{I}_\beta$.
- So β is unlikely to be in the code until rate at least $\max_{\beta' \in \mathcal{I}_\beta} R^{\mathbb{E}}(\beta')$.

B. Threshold for containing a type (second moment method)

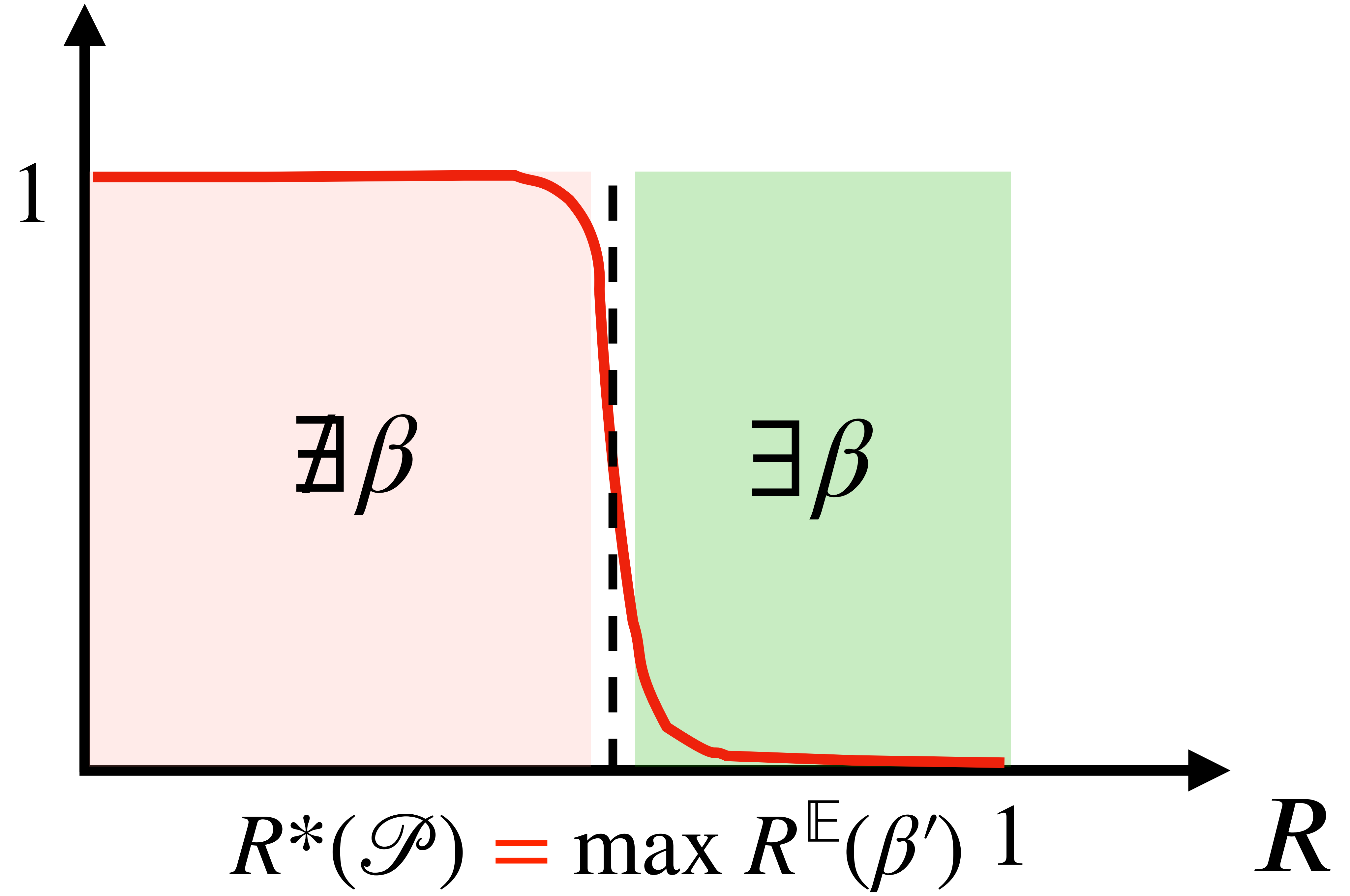


$$R^*(\beta) = \max_{\beta' \in \mathcal{I}_\beta} R^{\mathbb{E}}(\beta')$$



B. Threshold for containing a type

Suppose property $\mathscr{P}$ is satisfied $\iff$ no set of codewords with type β is in the code. Then we have computed $R^*(\mathscr{P})$.



$$R^*(\mathscr{P}) = \max_{\beta' \in \mathscr{I}_\beta} R^{\mathbb{E}}(\beta')$$

PART I: Informal results

A. Characterization theorems

B. Some applications

PART II: Proof outline for RLCs

A. Local properties

B. Threshold for containing a type

PART III: Formal results for RC and RLC

A. Characterization theorem for RLCs

B. Characterization theorem for RCs

PART IV: LDPC Codes

A. Definitions

B. Reduction

A. Characterization theorem for RLCs

- Given local property defined by exclusion of sets of ℓ vectors whose types lie in a set T

$$R^* = \min_{\tau \in T} \left(\max_{\tau' \in \mathcal{I}_\tau} R^{\mathbb{E}}(\tau') \right)$$

- If $R \leq R^* - \varepsilon$, then random linear code satisfies property w.h.p.
- If $R \geq R^* + \varepsilon$, then random linear code does not satisfy property w.h.p.

B. Characterization theorem for RCs

- Given symmetric property defined by exclusion of sets of ℓ vectors whose types lie in a set T

$$R^* = \min_{\tau \in T} R^{\mathbb{E}}(\tau)$$

- If $R \leq R^* - \varepsilon$, then random code satisfies property w.h.p.
- If $R \geq R^* + \varepsilon$, then random code does not satisfy property w.h.p.

PART I: Informal results

A. Characterization theorems

B. Some applications

PART III: Formal results for RC and RLC

A. Characterization theorem for RLCs

B. Characterization theorem for RCs

PART II: Proof outline for RLCs

A. Local properties

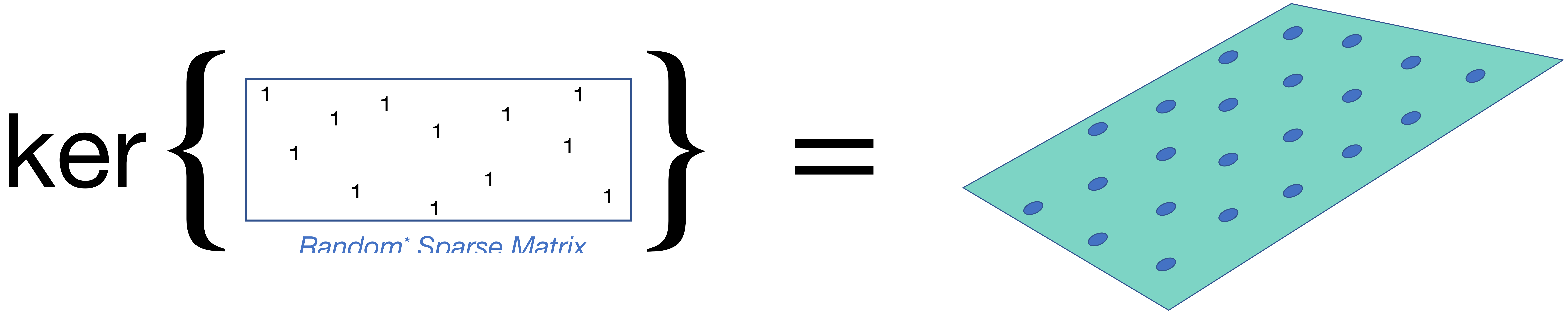
B. Threshold for containing a type

PART IV: LDPC Codes

A. Definitions

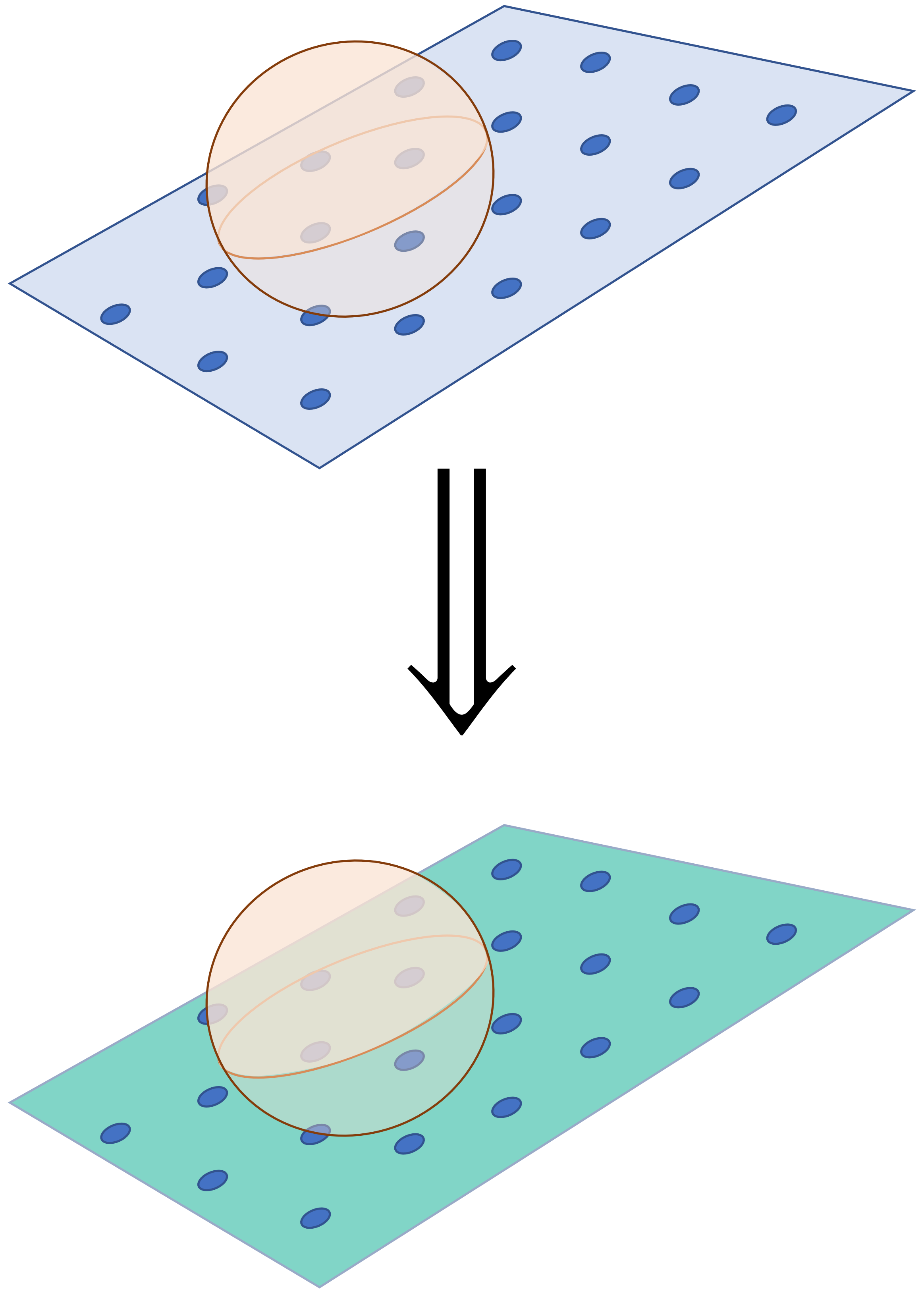
B. Reduction

A. Definitions



- Low-Density Parity-Check (LDPC) codes.
  - Very fast decoding algorithms.
  - Ubiquitous in theory and practice.
- Gallager showed that they achieve GV bound over binary alphabets (1960s).
- What about other combinatorial properties?
- Are they (combinatorially) list-decodable?

B. Reduction



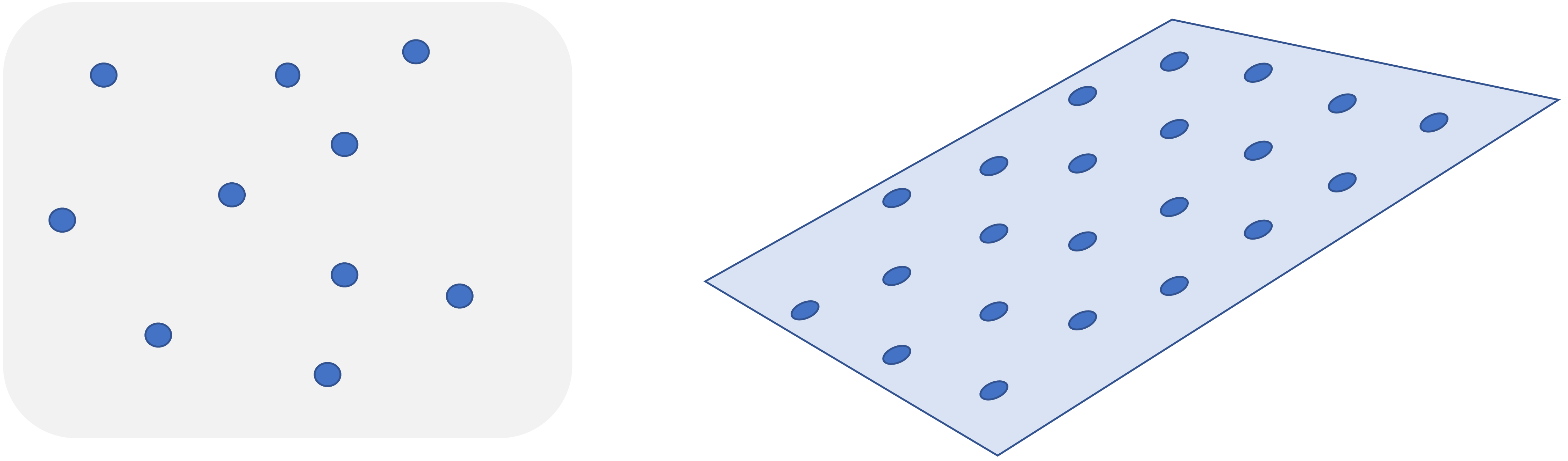
- If RLC of rate R satisfies a local property $\mathscr{P}$ w.h.p.
- Then LDPC code of rate R also satisfies $\mathscr{P}$ w.h.p.

LDPC codes achieve every local property RLCs achieve!

B. Reduction (proof idea)

- Let B be an $n \times \ell$ matrix of full rank and column distance δ
- For RLC of rate R, $\Pr(B \subset C) = q^{-n\ell(1-R)}$
- For any $\varepsilon > 0, \exists$ **L** such that **L**DPC code of rate R, $\Pr(B \subset C) = q^{-n\ell(1-\varepsilon)(1-R)}$
- **L** depends on $\varepsilon, \delta, q, \ell$

Conclusion

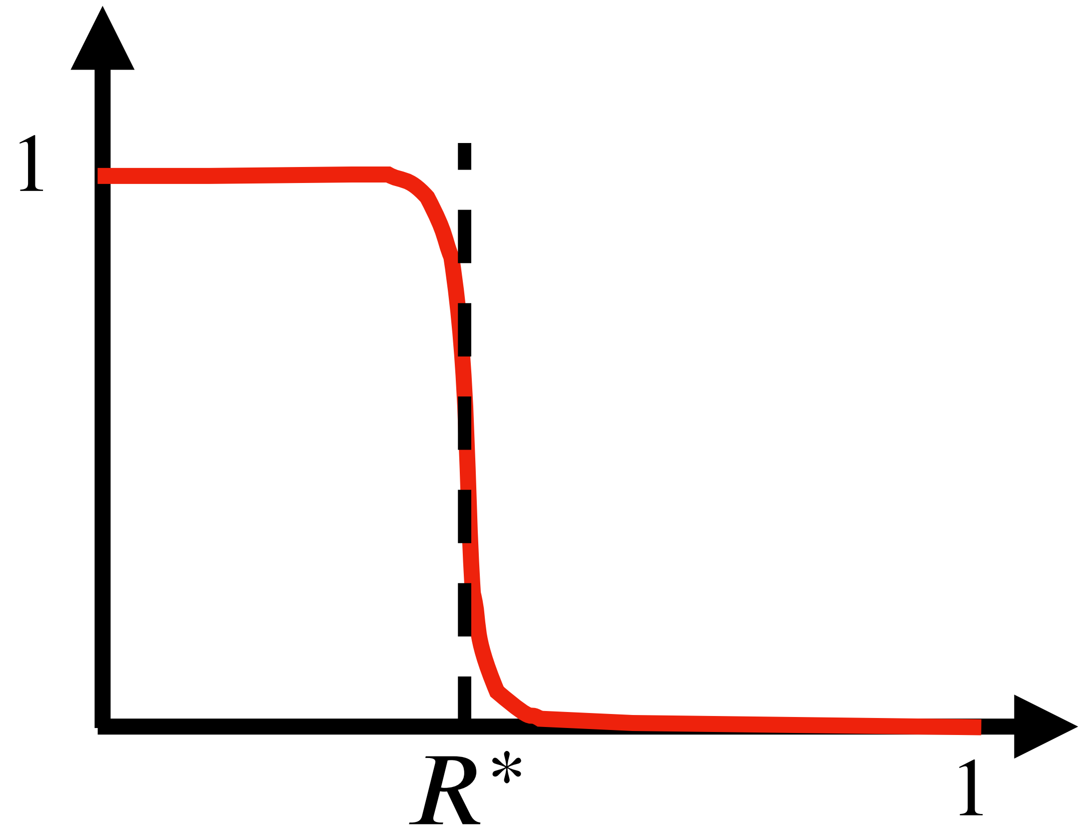


We wanted to understand the relation between combinatorial properties of random [linear] codes and their rate.

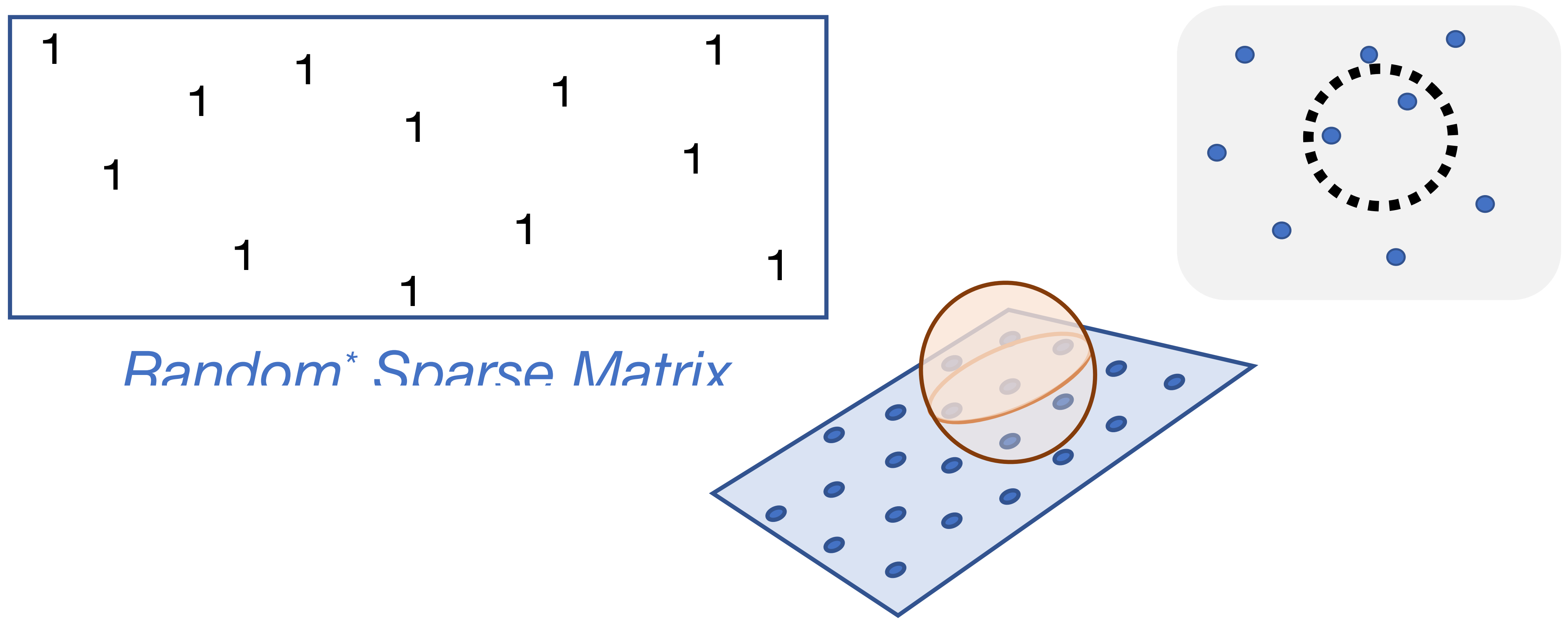
$$R^*_{RLC} = \min_{\tau\in T}\left(\max_{\tau'\in\mathcal{I}_\tau} R^{\mathbb{E}}(\tau')\right)$$

$$R^*_{RC} = \min_{\tau\in T} R^{\mathbb{E}}(\tau)$$

The threshold rate has a nice characterization.



Large classes of natural properties have threshold rates.



Applications to LDPC codes, list-sizes of RLCs and RCs, and other natural properties.

Open questions

1. Other applications of our characterization theorems?

2. Algorithms for list-decoding LDPC codes?

3. Many more…

Sharp threshold rates for random codes
Guruswami, Mosheiff, Resch, S., Wootters
ITCS 2021, arXiv:2009.04553

LDPC codes achieve list-decoding capacity
Mosheiff, Resch, Ron-Zewi, S., Wootters
FOCS 2020, arXiv:1909.06430

Bounds for list-decoding and list-recovery of random linear codes
Guruswami, Li, Mosheiff, Resch, S., Wootters
RANDOM 2020, arXiv:2004.13247

Questions?