Class 19

THE LAST ONE!!!!

Time flies when
you're having fun!

Plan for today

- Quick recap of the quarter
- Research talks!

What just happened?

- Techniques for analyzing randomized algorithms!
  - Linearity of expectation
  - Markov and Chebyshev
  - Chernoff bounds
  - "Poissonization"
  - Metric embeddings/JL transforms
  - The probabilistic method
    - Second moment method
    - LLL
    - Derandomization via conditional expectation
  - Markov chains
    - Mixing times and coupling
  - Martingales
    - Azuma-Hoeffding bound
    - Martingale stopping theorem
  - Pseudorandomness

Plus homework!

And quizzes!

And in-class work!

- Algorithms!
  - Polynomial identity testing
  - Perfect matchings
  - Karger's algorithm for minimum cut
  - Primality testing
  - Sampling-based median
  - Randomized routing
  - Load balancing and the power of 2 choices
  - Bourgain's embedding, and an approximate sparsest-cut algorithm
  - Locality sensitive hashing
  - Compressed sensing
  - Count-min sketch
  - Deterministic approximation algorithms for k-SAT, Max-Cut
  - Algorithmic LLL
  - Randomized algorithm for 2SAT
  - MCMC, sampling random colorings
  - Consensus algorithms
  - Extractors via expander walks for derandomizing randomized algorithms

Key take-aways

- Randomness is a powerful tool in computation.
- There's a lot of beautiful math that goes into analyzing it.

- I hope that now, you:
  - Are proficient with some techniques for the analysis of randomized algorithms.
  - Have seen enough examples of using these techniques that you can use them in your own work/research/life.

A few messages

Thanks to the CAs!

Guy

Pras

Ian

Preey

Thank You!!!!!

- For all your hard work, great questions, and valuable feedback!

Combinatorial Properties of Random(ish) Sets

Based on joint works with Ray Li, Venkat Guruswami, Jonathan Mosheiff, Nic Resch, Noga Ron-Zewi, Shashwat Silas, ….

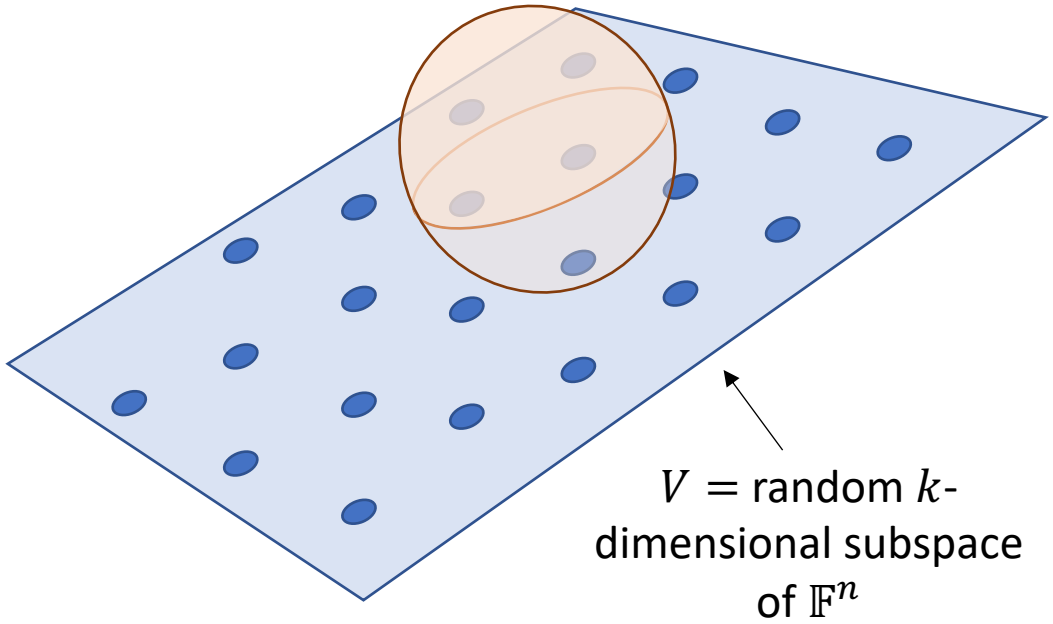
This talk is about

We will define all these terms in just a minute!

- What combinatorial properties are satisfied by random subspaces over finite fields?

The plan:

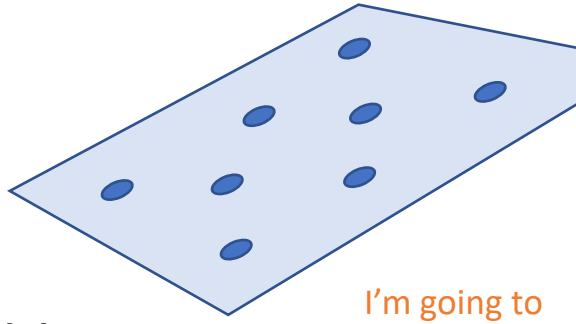
- Intro to these
- Why this question is hard
- Some of the approaches we have



Random subspaces and a few nice properties they might satisfy

Also some motivation from coding theory

Random subspaces over finite fields



I'm going to draw subspaces like this, but the picture is pretty misleading about the geometry.

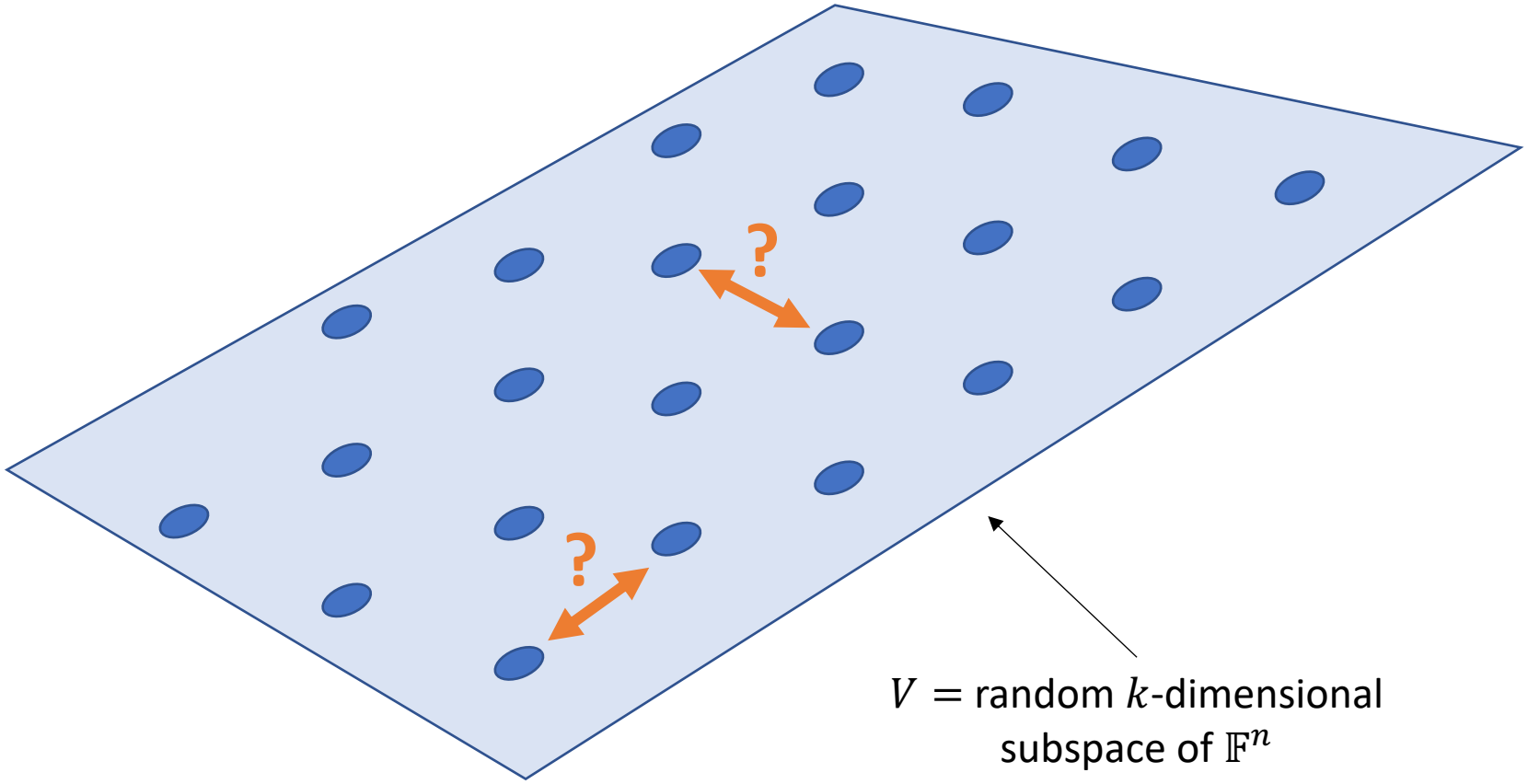
- A **finite field** $\mathbb{F}$ is a finite set where you can do arithmetic like you want to.
  - E.g., $\mathbb{F} = \mathbb{F}_2 = \{0,1\}$ with arithmetic mod 2.
- $V \subseteq \mathbb{F}^n$ is a subspace if it's closed under addition and scalar multiplication.
  - Eg., subspace of $\mathbb{F}_2^6$ of dimension 2:

$$\text{span}\left\{ \begin{pmatrix}1\\0\\1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\\0\\1\end{pmatrix} \right\} = \left\{ \begin{pmatrix}0\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\\0\\0\\1\end{pmatrix} \right\}$$

- A random subspace is…a (uniformly) random subspace.

Hamming distance between $x, y \in \mathbb{F}^n$ is $d(x,y) = |\{i : x_i \neq y_i\}|$

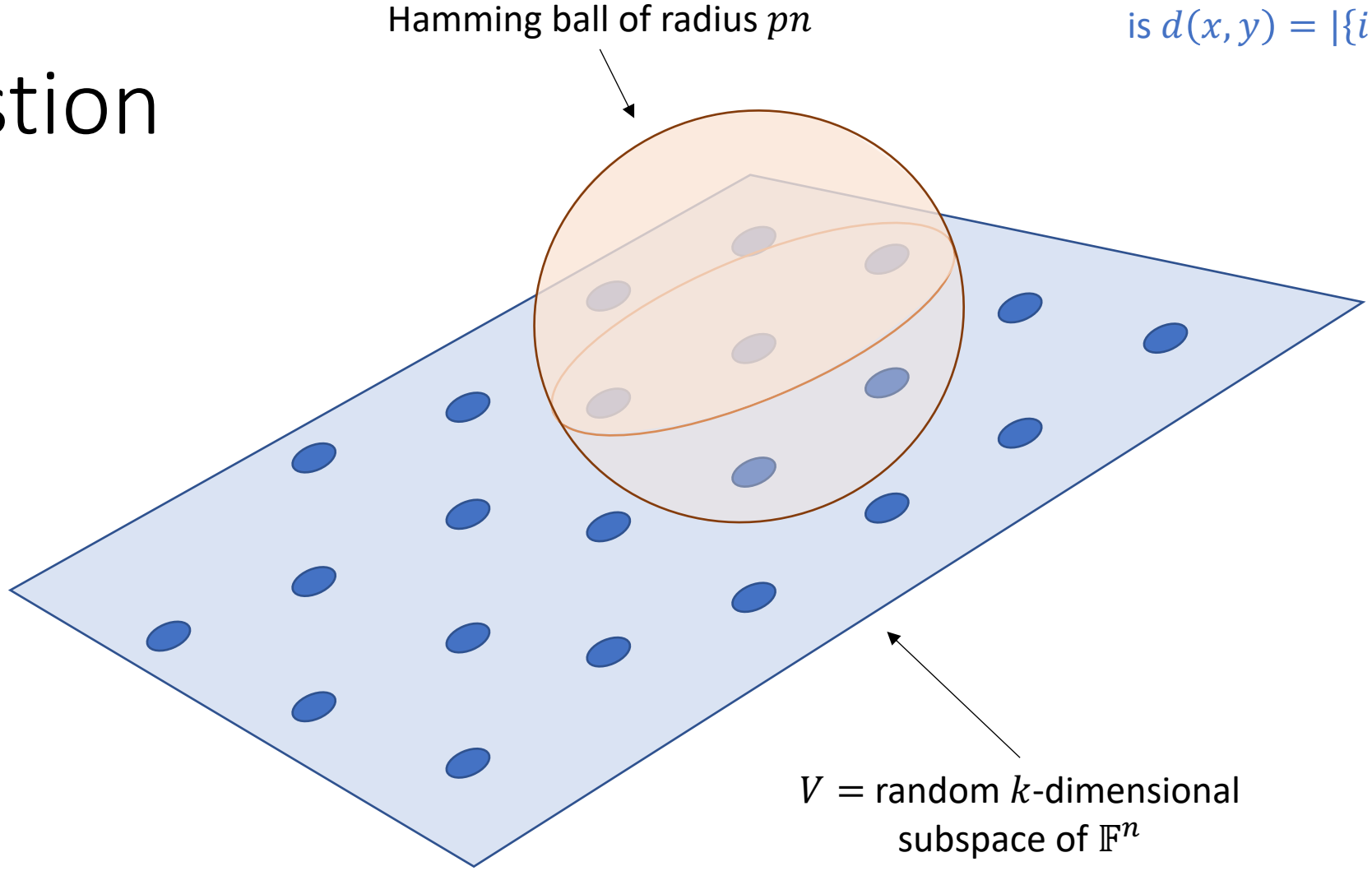
A Question



- What is the closest any two points in V can be to each other (in Hamming distance)?
- How does this depend on k and n?

A Question

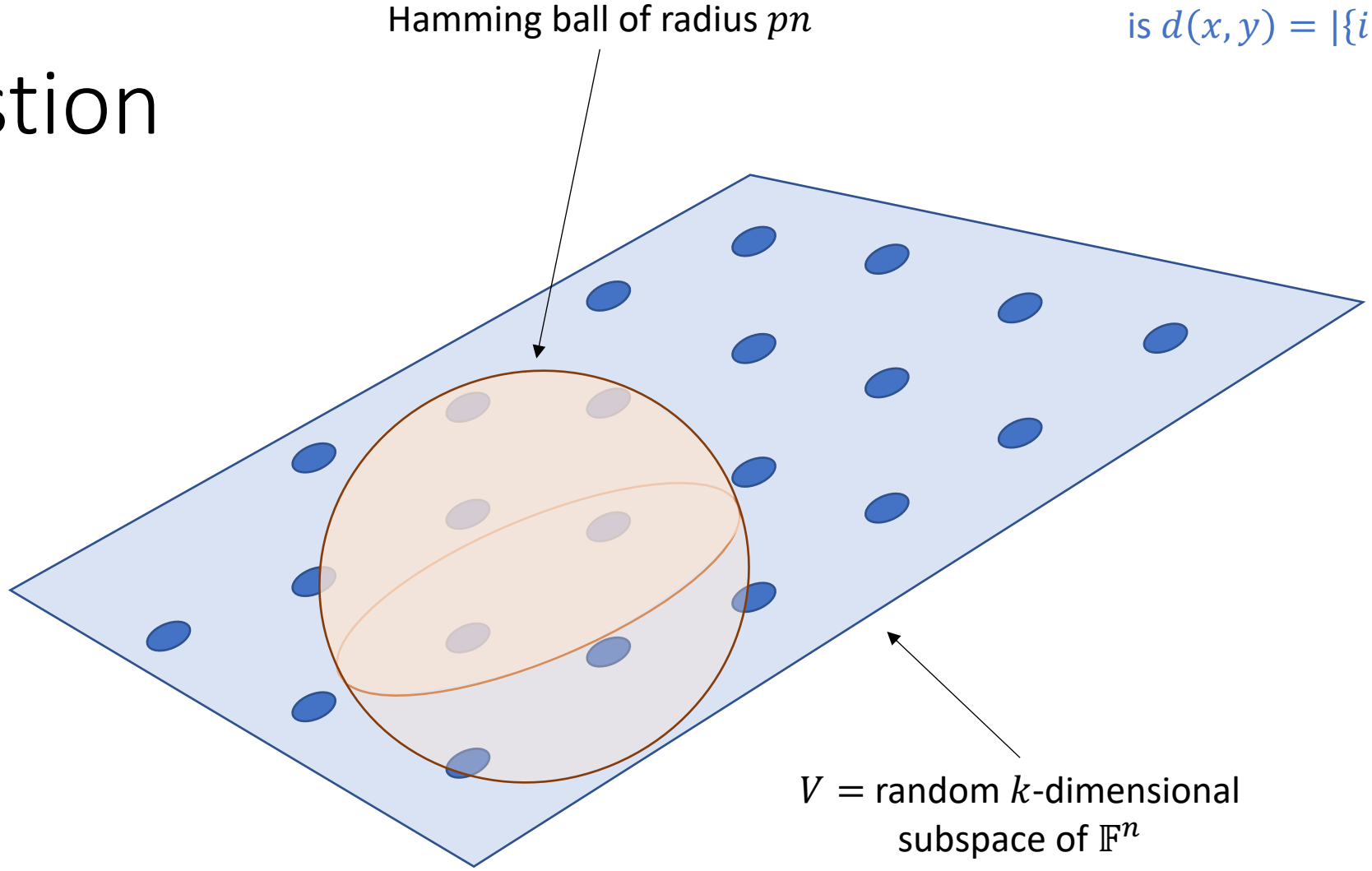
Hamming distance between $x, y \in \mathbb{F}^n$ is $d(x, y) = |\{i : x_i \neq y_i\}|$



- What is the maximum number of points of V that lie in any Hamming ball of radius pn?
- How does this depend on k and n?

A Question

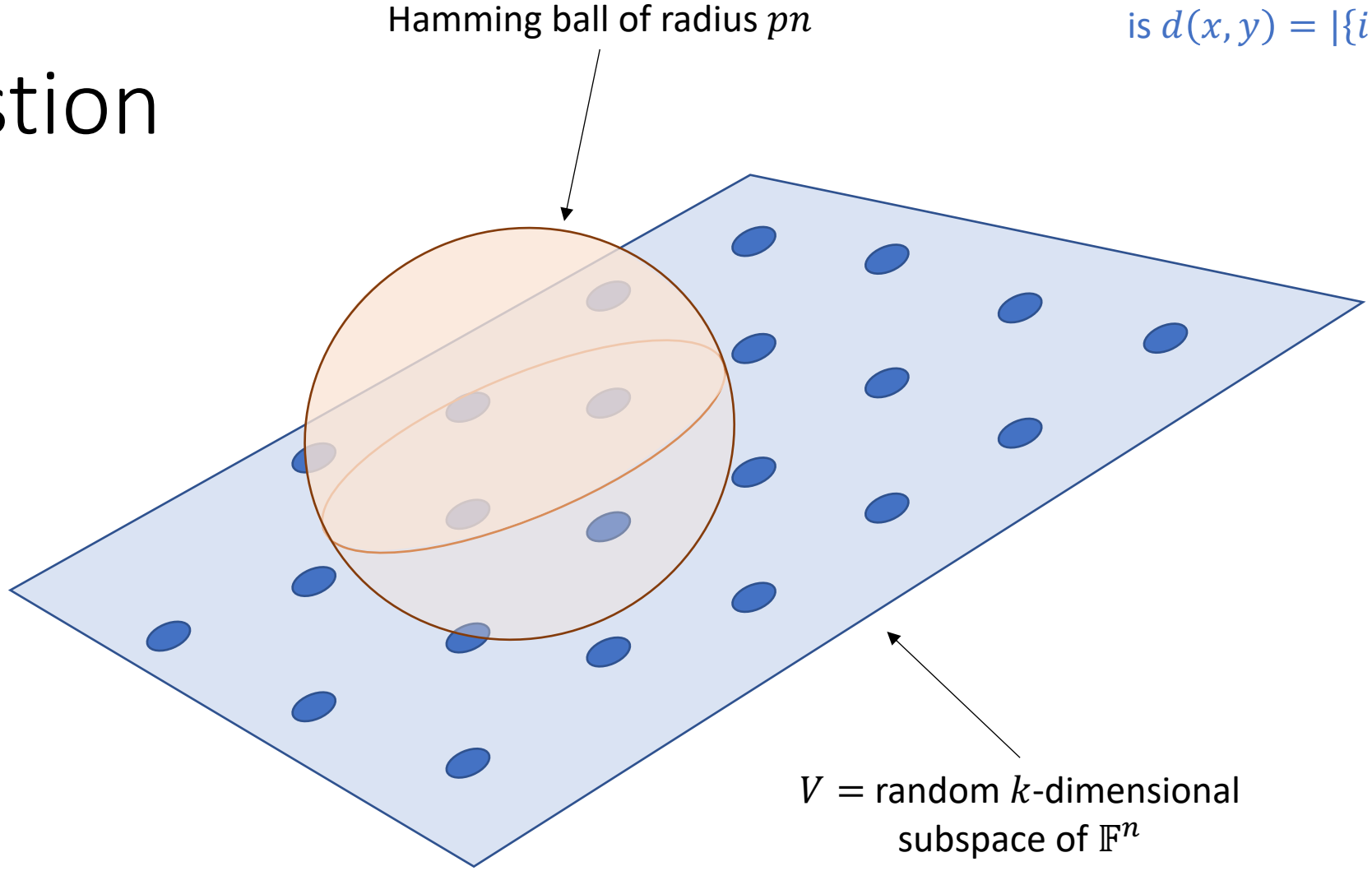
Hamming distance between $x, y \in \mathbb{F}^n$ is $d(x, y) = |\{i : x_i \neq y_i\}|$



- What is the maximum number of points of V that lie in any Hamming ball of radius pn?
- How does this depend on k and n?

A Question

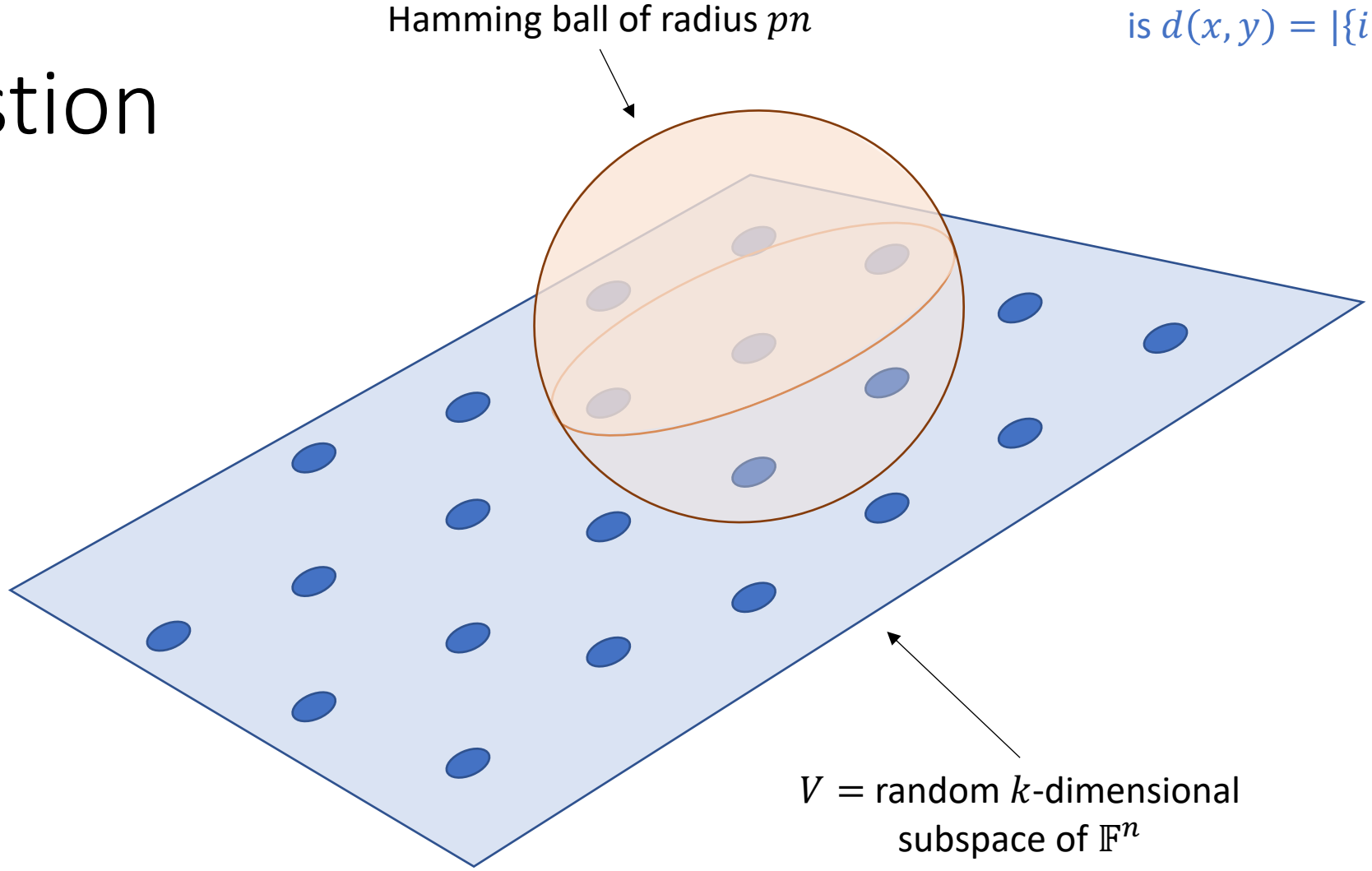
Hamming distance between $x, y \in \mathbb{F}^n$ is $d(x, y) = |\{i : x_i \neq y_i\}|$



- What is the maximum number of points of V that lie in any Hamming ball of radius pn?
- How does this depend on k and n?

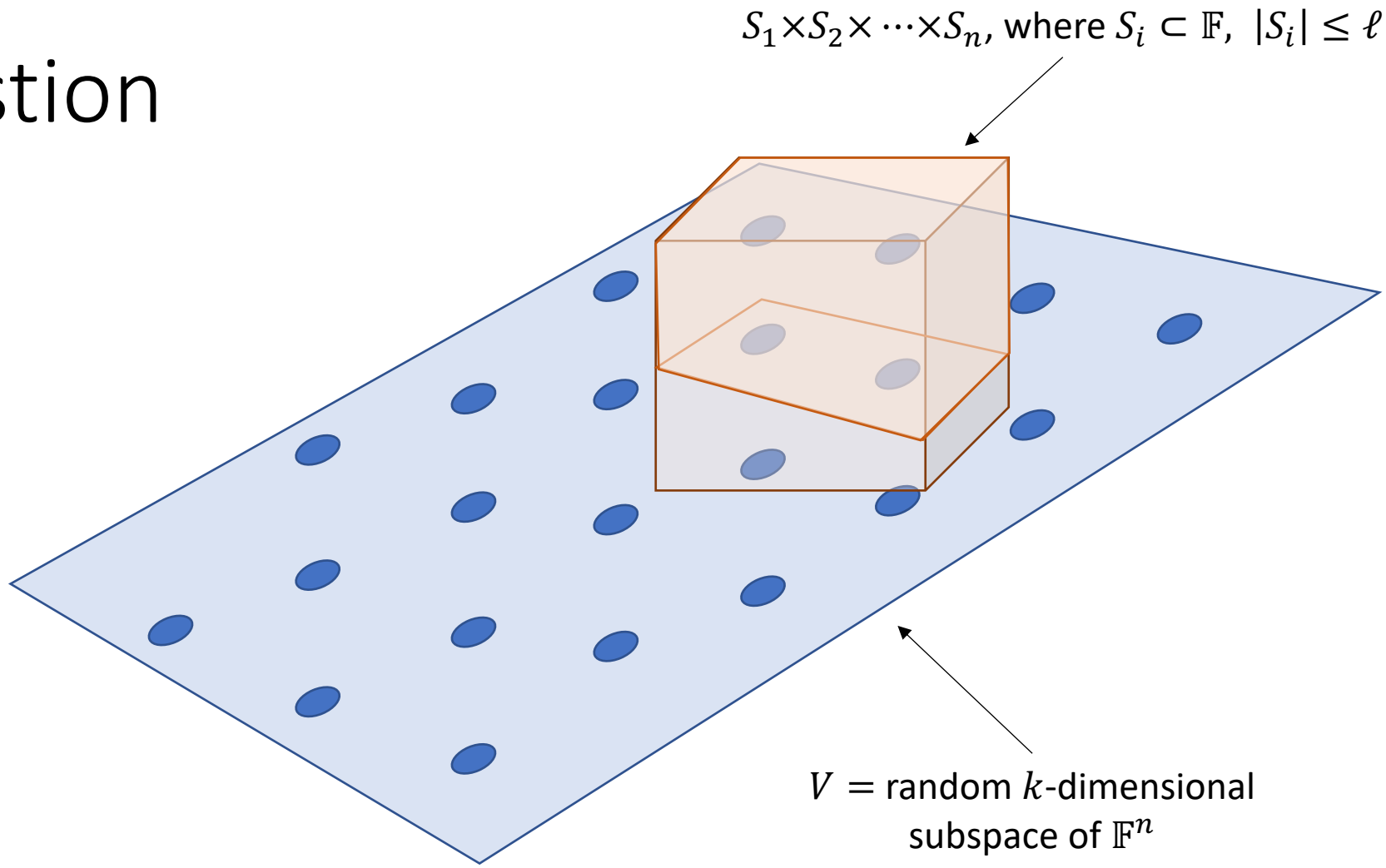
A Question

Hamming distance between $x, y \in \mathbb{F}^n$ is $d(x, y) = |\{i : x_i \neq y_i\}|$



- What is the maximum number of points of V that lie in any Hamming ball of radius pn?
- How does this depend on k and n?

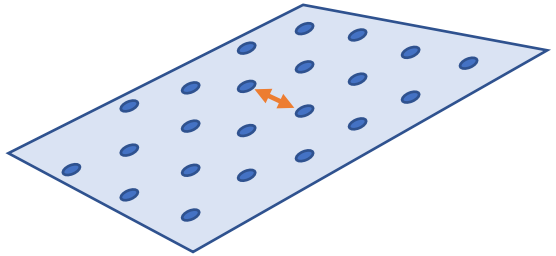
A Question



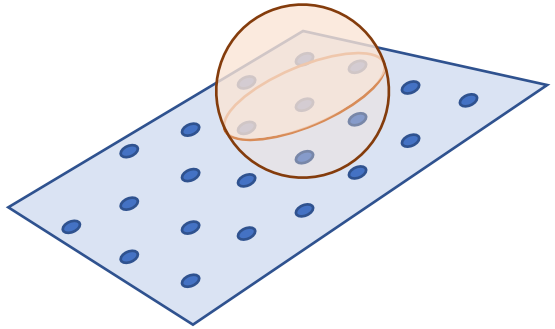
- What is the maximum number of points of V that lie in any Hamming ball of radius pn?
- How does this depend on k and n?

Why do we care?

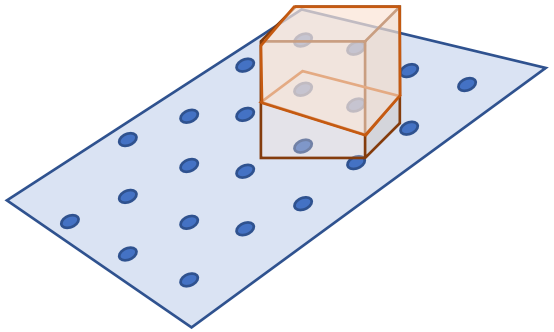
- Interesting combinatorial questions!
- They have applications in **error correcting codes.**



"Distance"



"List Decoding"



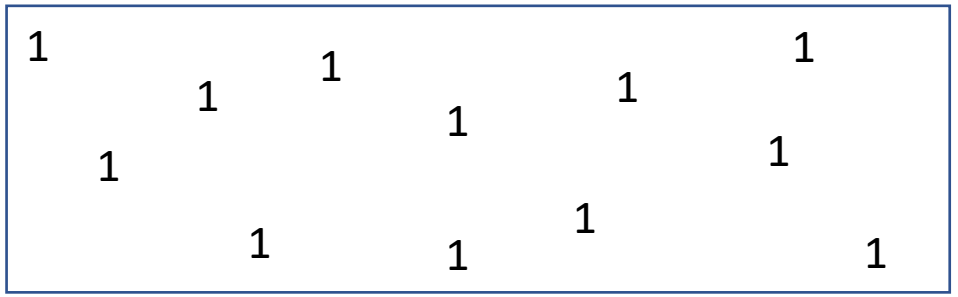
"List Recovery"

More questions

also motivated by error correcting codes

- What if we replace the uniformly random subspace with the kernel of a sparse random matrix?

$$V = \mathrm{Ker}$$



- What if we replace the uniformly random subspace with the image of a random Vandermonde matrix?

$$V = \mathrm{RowSpan}$$

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \dots & \alpha_n^2 \\ \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & & \alpha_n^3 \end{pmatrix}$$

Lots of work on these!

This is not a complete list and the dates might be a bit off…

Uniformly random subspaces:

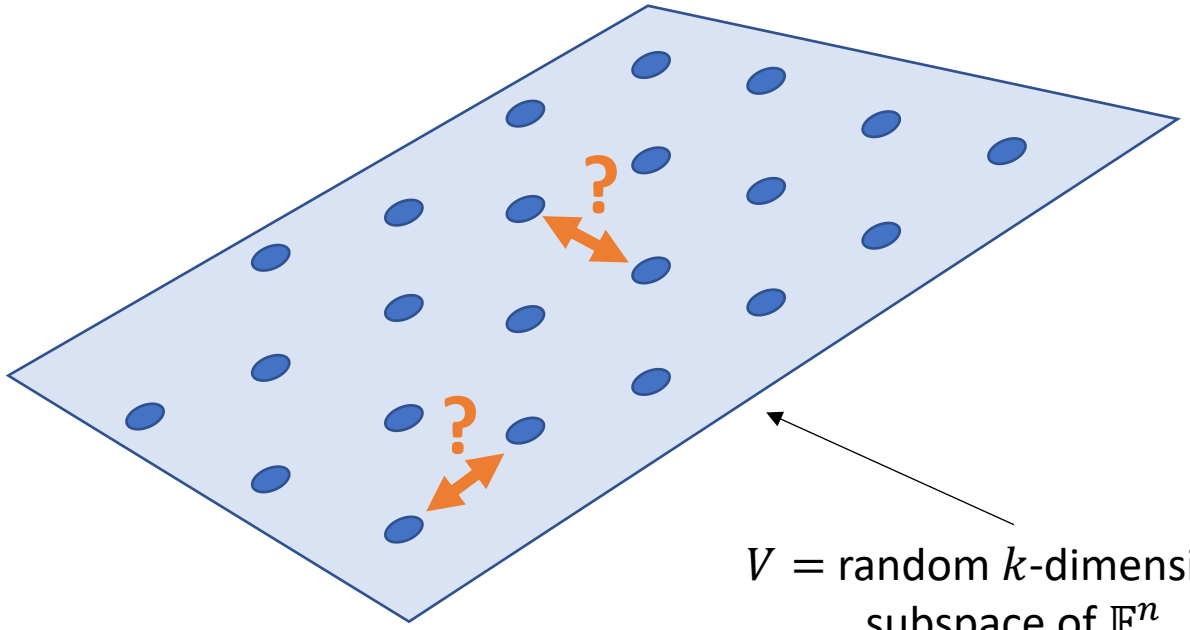
- [Zyablov, Pinsker, 1981]
- [Elias 1991]
- [Guruswami, Hastad, Kopparty 2011]
- [Cheraghchi, Guruswami, Velingker, 2013]
- [W. 2013]
- [Rudra, W., 2014]
- [Rudra, W., 2017]
- [Li, W. 2018]
- [Guruswami, Li, Mosheiff, Resch, Silas, W., 2020]

More structured randomized subspaces:

- [Rudra, W., 2014]
- [Rudra, W., 2015]
- [Mosheiff, Resch, Ron-Zewi, Silas, W., 2020]
- [Shangguan, Tamo 2020]
- [Goldberg, Shangguan, Tamo 2021]
- [Ferber, Kwan, Sauermann 2022]
- [Guo, Li, Shangguan, Tamo, W., 2022]
- [Guruswami, Mosheiff, 2022]
- [Brakensiek, Gopi, Makam 2022]
- …

Why these questions are hard

Actually, this one's not hard



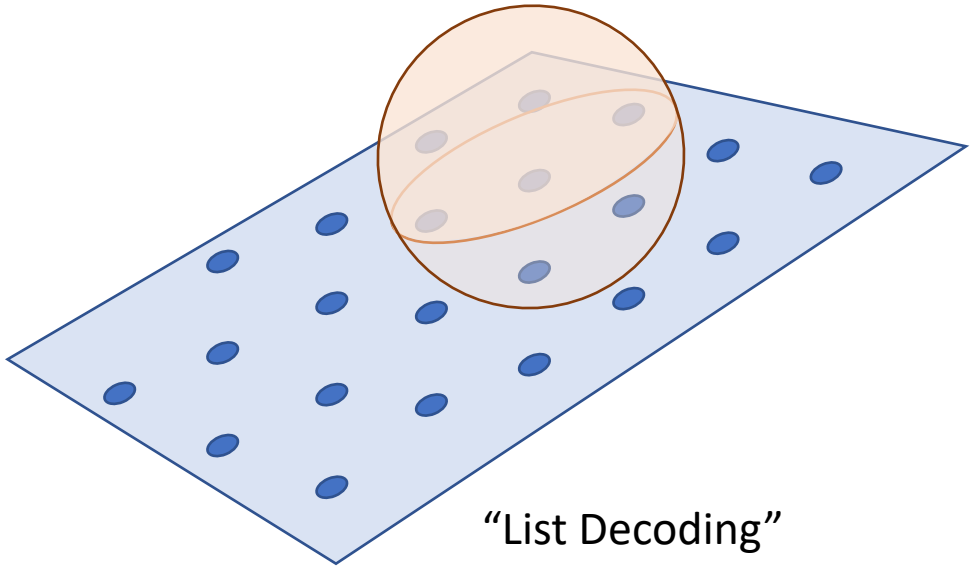
V = random k-dimensional subspace of $\mathbb{F}^n$

Fun exercise!

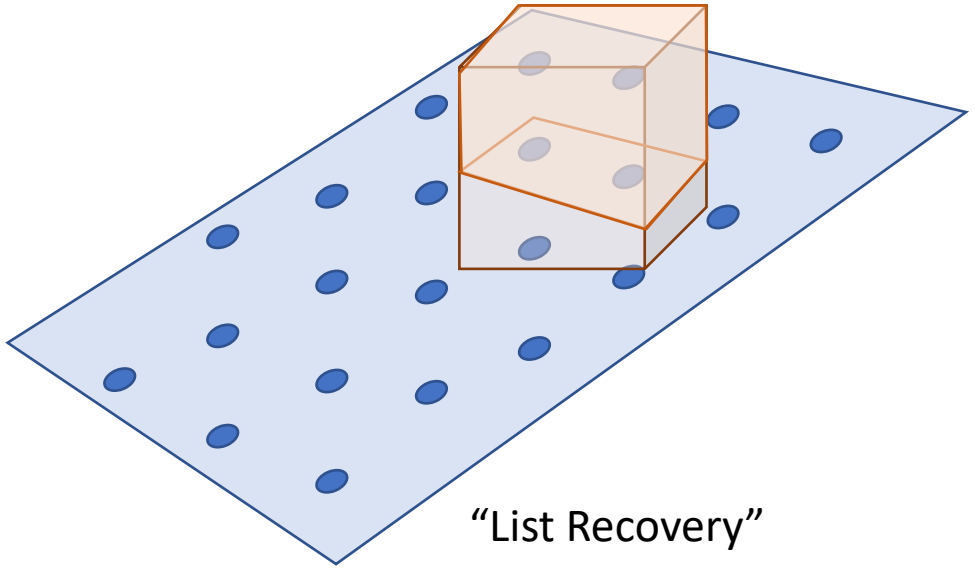
What should k be, in terms of n and p, so that no two points are closer than pn?

- What is the closest any two points in V can be to each other (in Hamming distance)?

But these are hard



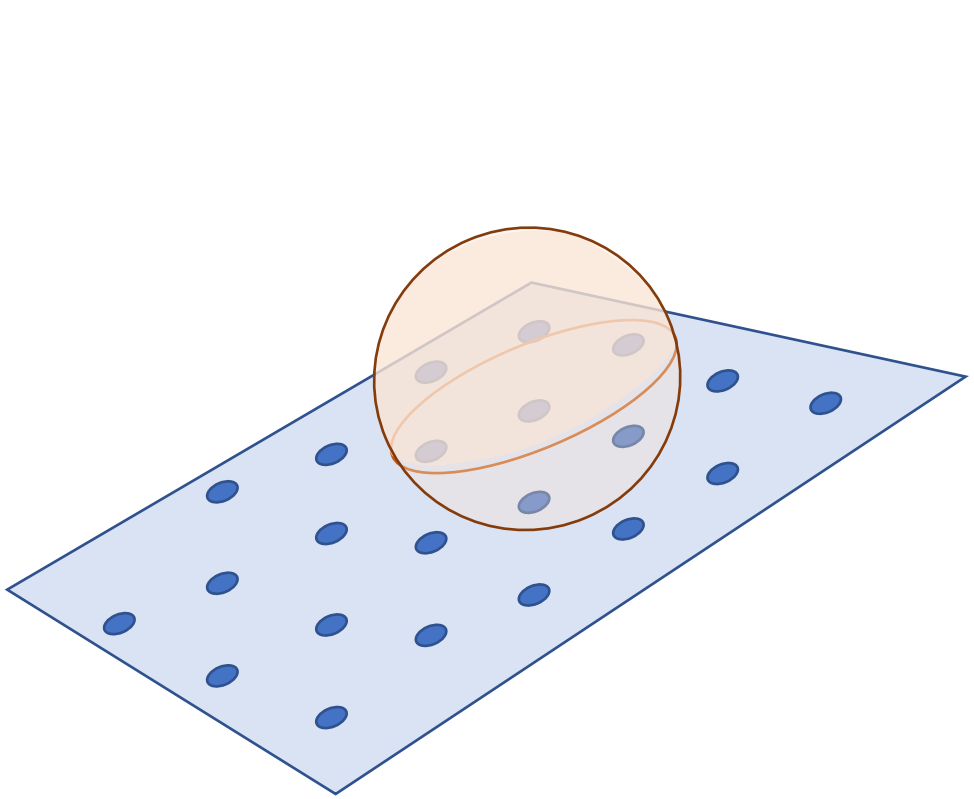
“List Decoding”



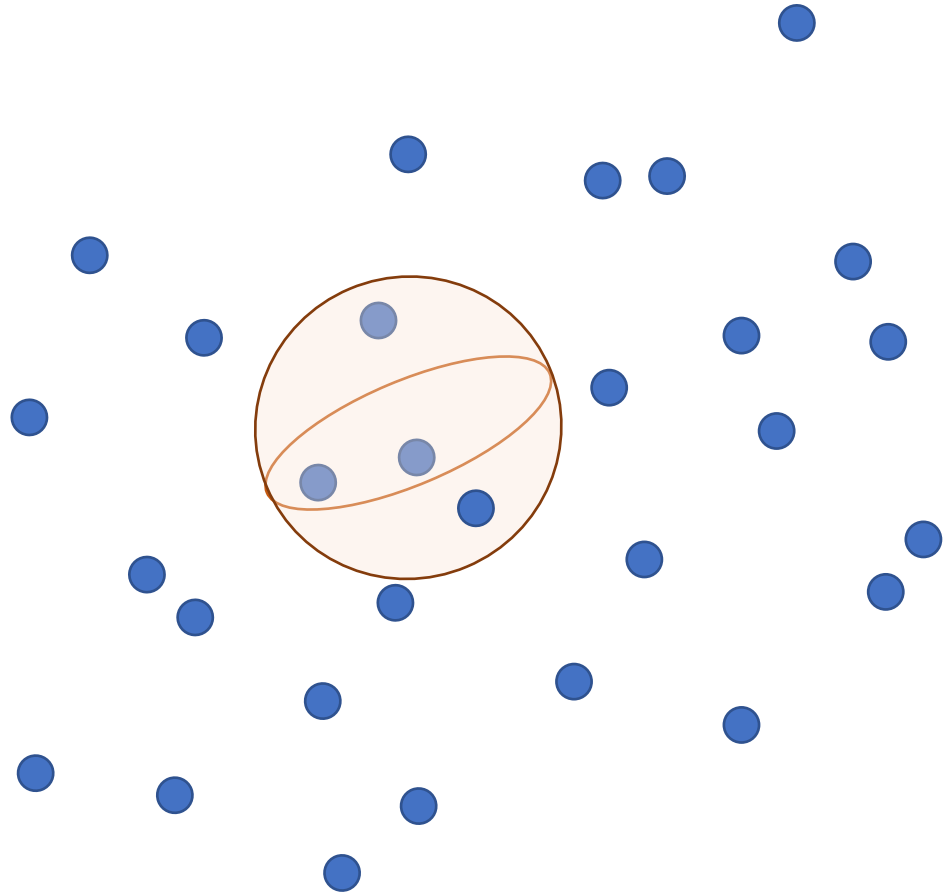
“List Recovery”

Compare to an easy question

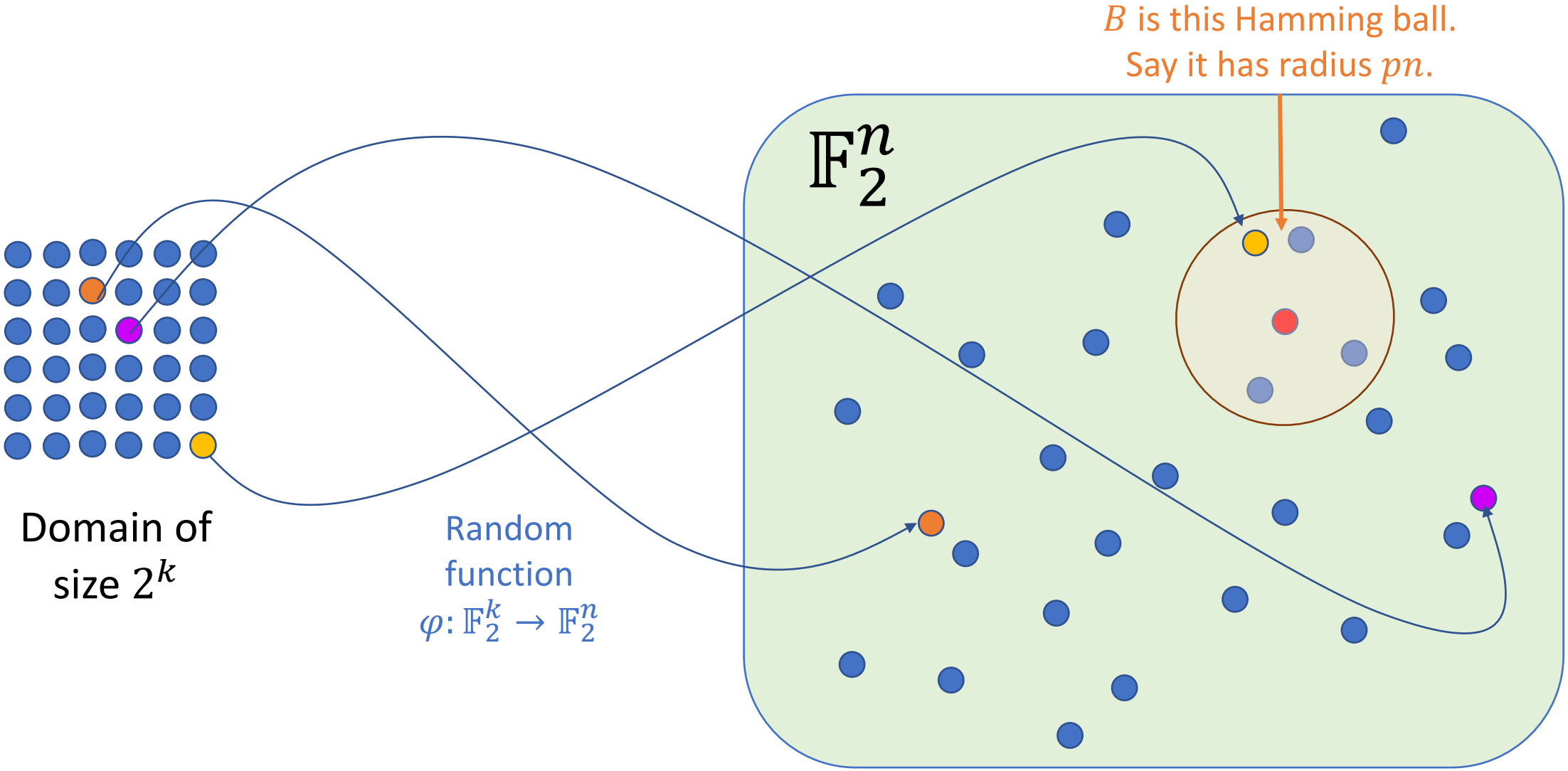
- Instead of a random **subspace**, consider a uniformly random **set**.



VS



What's the probability that some L points in a uniformly random **set** lie in a Hamming ball?

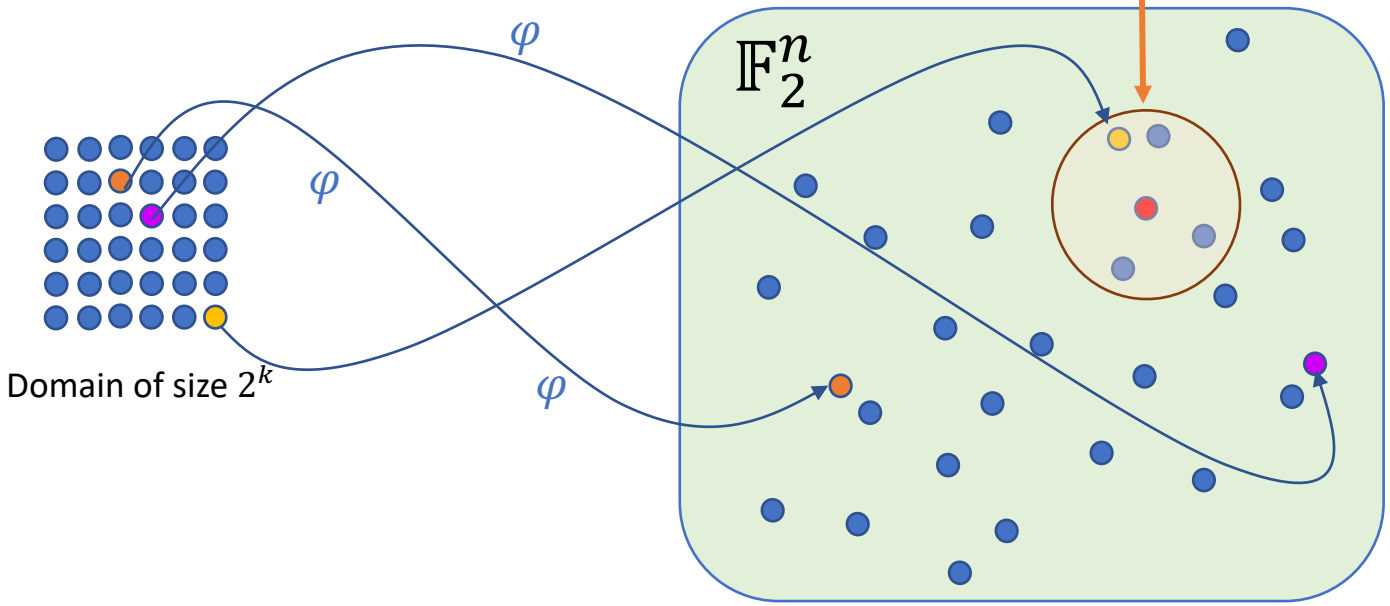


$$H(p) = p\log\left(\frac{1}{p}\right) + (1-p)\log\left(\frac{1}{1-p}\right)$$

Probability that a fixed L pts land in a fixed ball

- $\Pr[\,\varphi(x_1), \ldots, \varphi(x_L) \in B\,]$
- $= \left(\frac{\mathrm{Vol}(B)}{2^n}\right)^L$
- $\approx 2^{-n(1-H(p))L}$

B is this Hamming ball.
Say it has radius $p \cdot n$.



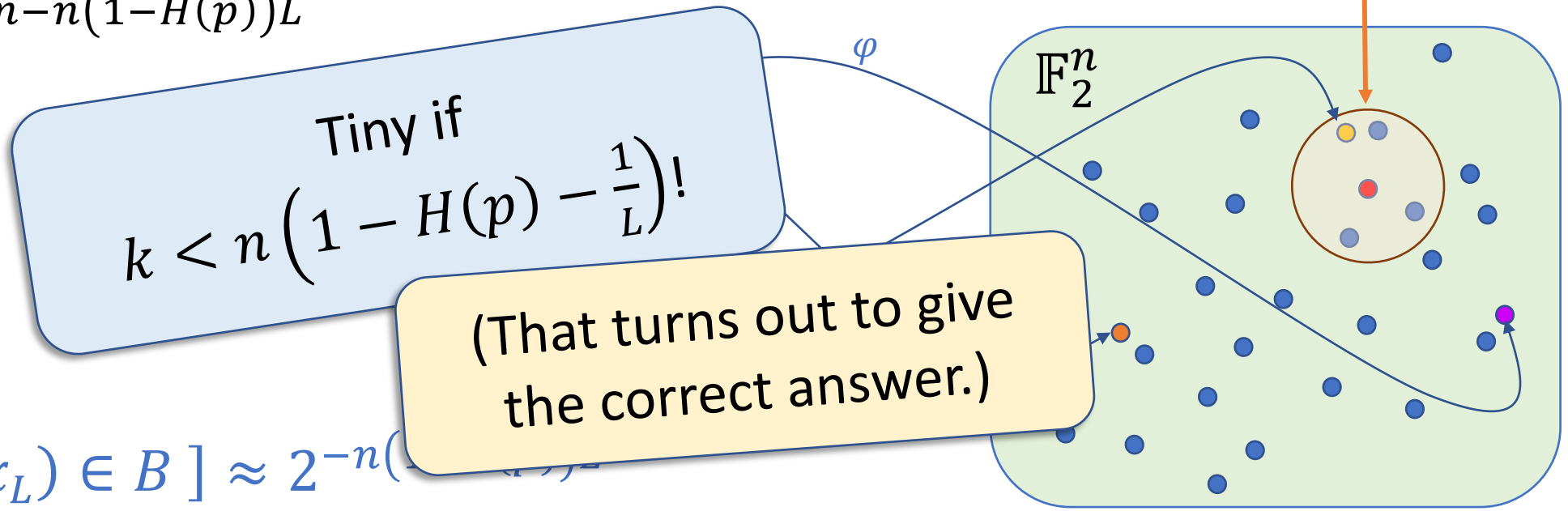
Probability that ~~a fixed~~ any L pts land in ~~a fixed~~ any ball

- Union bound!
- $\Pr[\ \exists x_1, \ldots, x_L, B \ s.t. \ \varphi(x_1), \ldots, \varphi(x_L) \in B]$
- $\leq \binom{2^k}{L} \cdot 2^n \cdot 2^{-n(1-H(p))L}$
- $\leq 2^{Lk+n-n(1-H(p))L}$

Tiny if
$k < n\left(1 - H(p) - \frac{1}{L}\right)!$

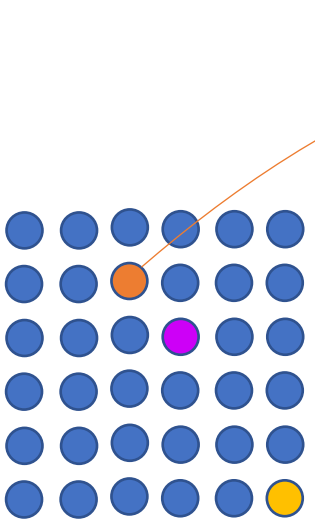
(That turns out to give the correct answer.)

B is this Hamming ball.
Say it has radius $p \cdot n$.



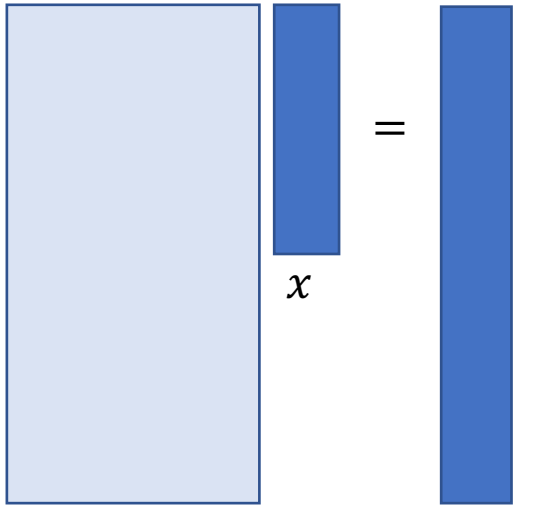
$\Pr[\ \varphi(x_1), \ldots, \varphi(x_L) \in B\ ] \approx 2^{-n(}$

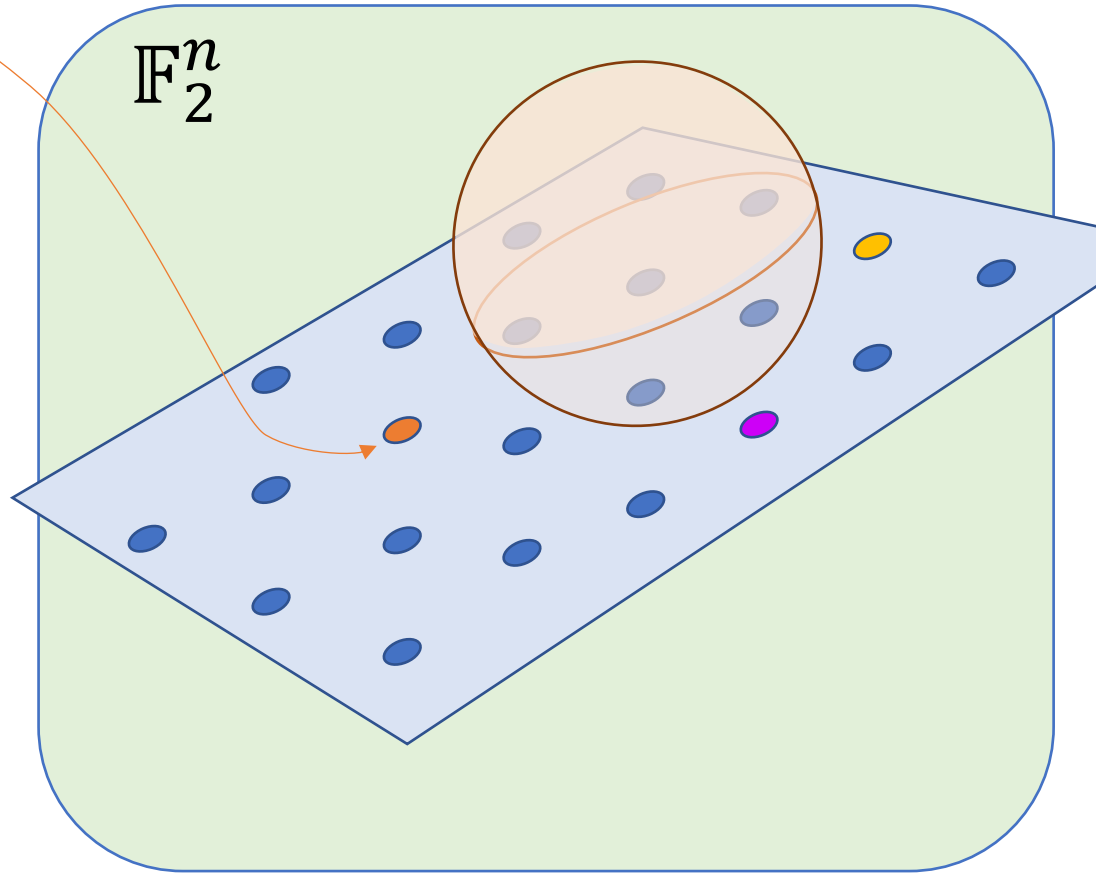
What goes wrong for a random subspace?



$\mathbb{F}_2^k$

The map $\varphi: \mathbb{F}_2^k \to \mathbb{F}_2^n$ is given by:



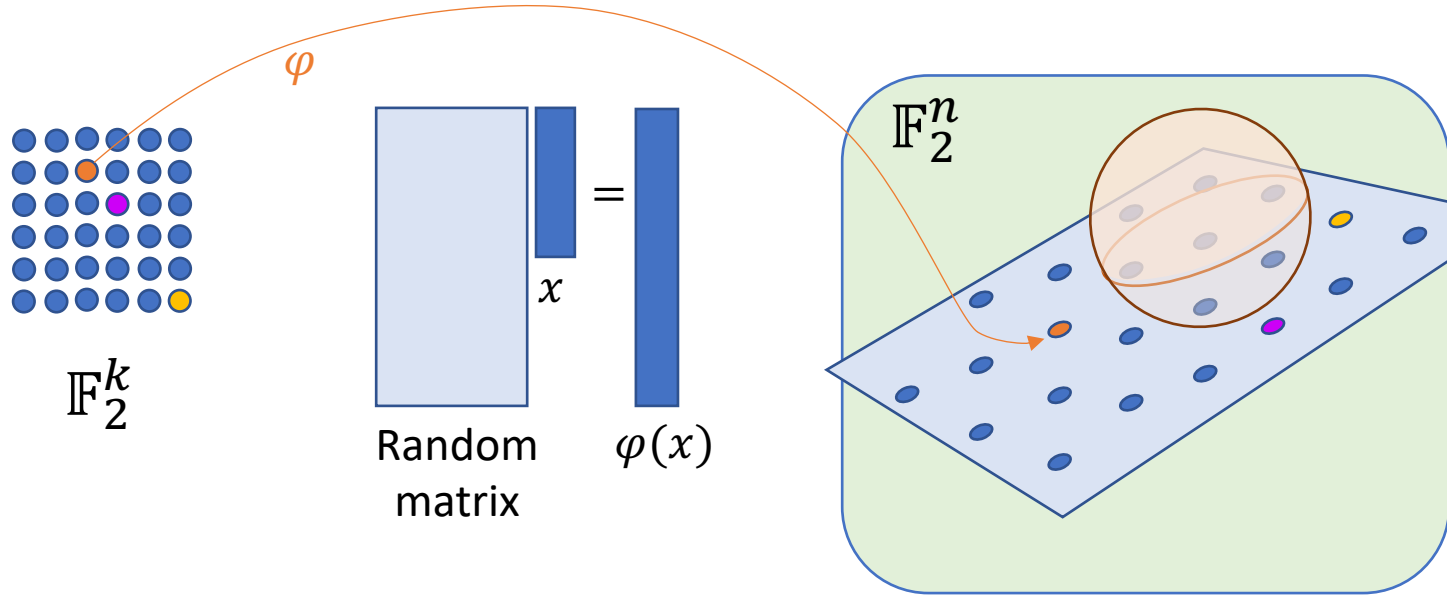


What goes wrong for a random subspace?

- $\Pr[\, \varphi(x_1), \ldots, \varphi(x_L) \in B \,]$
- $= \left(\frac{\mathrm{Vol}(B)}{2^n}\right)^L$
- $\approx 2^{-n(1-H(p))L}$

The $\varphi(x_i)$ aren't independent anymore!!
The union bound will fail ☹

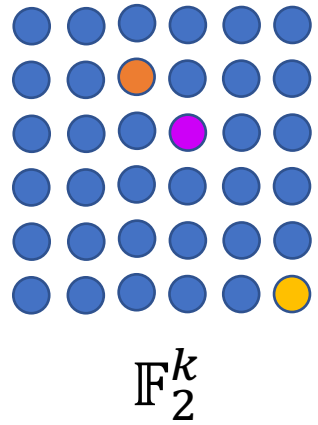
However, it turns out that $k \sim n\left(1 - H(p) - \frac{1}{L}\right)$ is still the right answer!



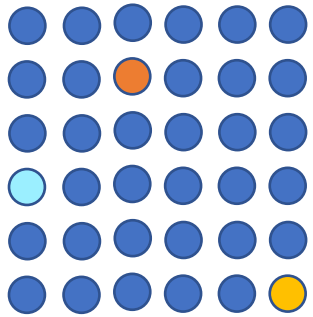
Techniques

Three ideas

Idea 1: Fancy union bounds

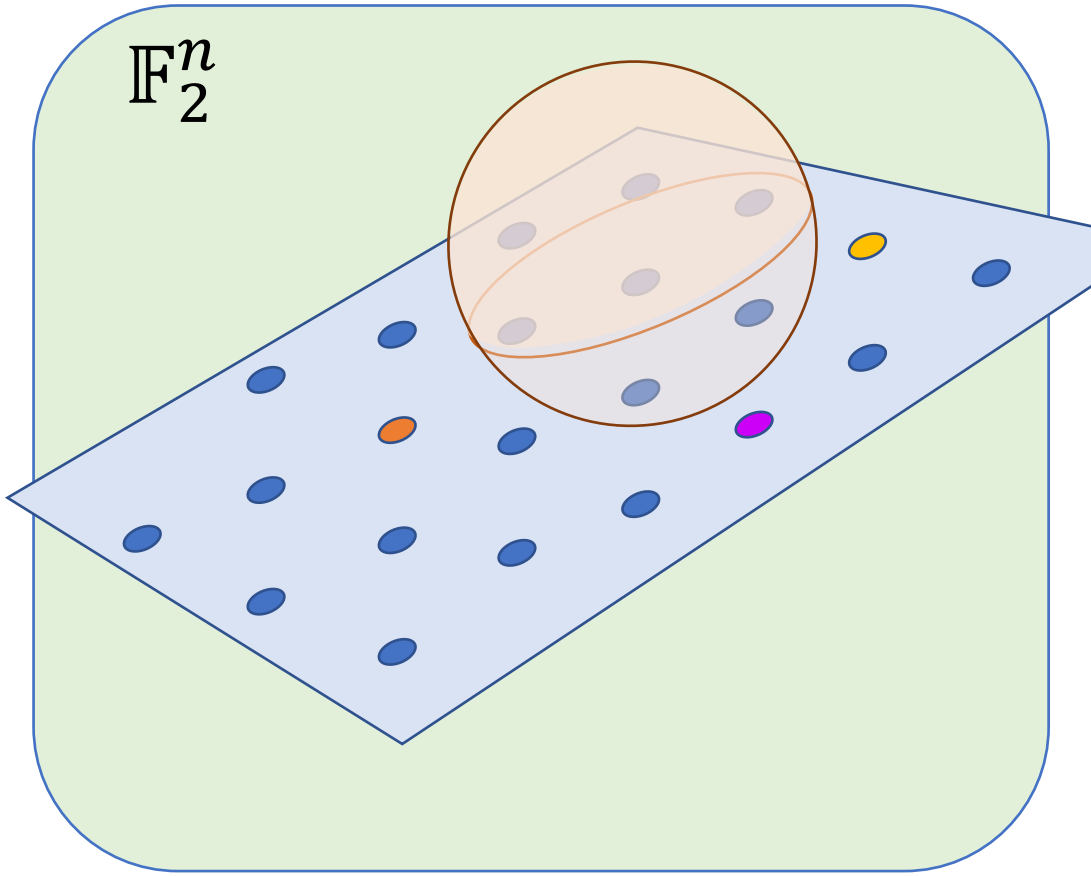


$\mathbb{F}_2^k$



$\mathbb{F}_2^k$

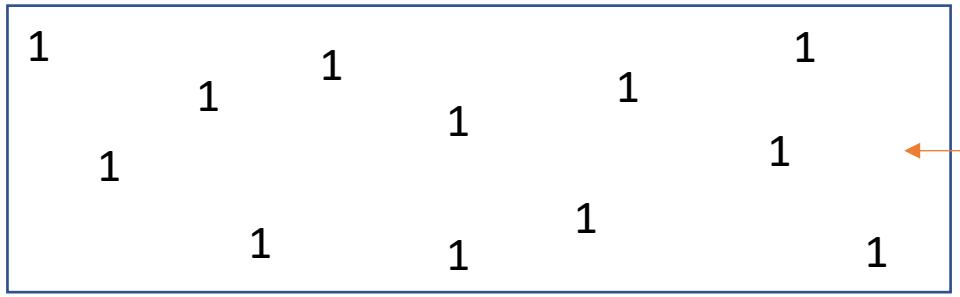
These two bad events are correlated!



Idea 2: Reduction to the random linear case

- It turns out that, e.g.,

$$V = \mathrm{Ker}$$



Random **sparse** matrix. (Constant number of 1's per row).

behaves enough like a completely random matrix that this problem is not actually harder!

- Proof crucially uses linearity of expectation and the second moment method!

Idea 3: Connections to graph theory/combinatorics

- Relevant for

$$V = \text{RowSpan} \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \dots & \alpha_n^2 \\ \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & & \alpha_n^3 \end{bmatrix}$$

but it's kind of hard to capture in one slide…

Related to the problem on HW1 where we used PIT to find perfect matchings!

- Attempt:
  - if too many vectors in V agree too much,
  - then you can cook up a matrix whose determinant should be zero
  - but by doing a combinatorial argument about the zero patterns in that matrix you can show that the determinant is unlikely to be zero.

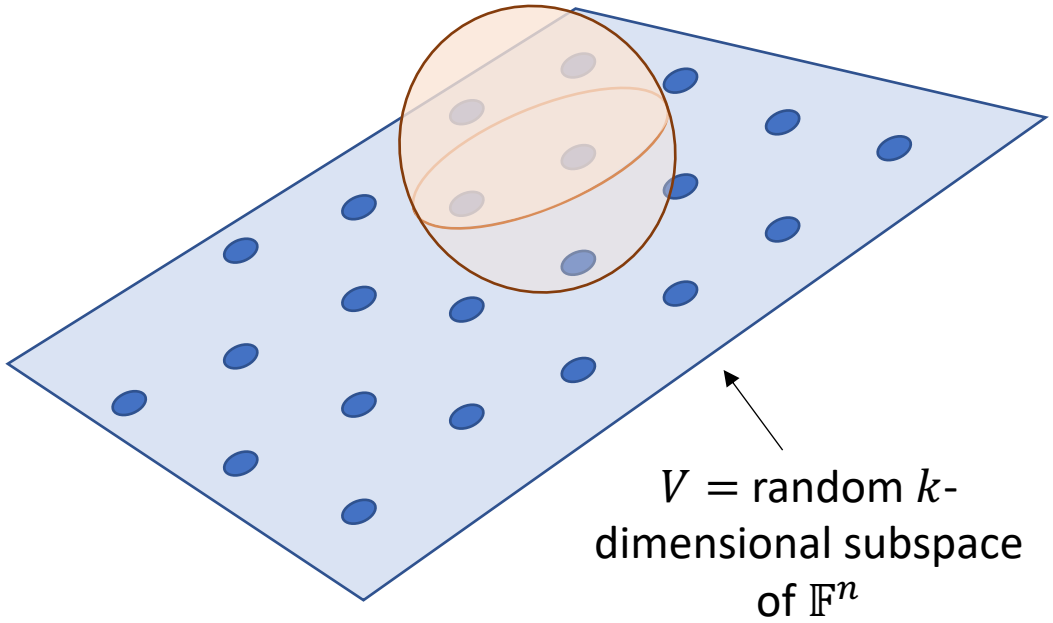
Conclusion

This talk was about

- Combinatorial properties of random subspaces over finite fields.

We saw:

- Intro to these
- Why this question is hard
- Some of the approaches we have



Moral of the story:

There are lots of really fun questions here! And CS265/CME309 material is relevant for them!

Thanks!

And thanks again for a great quarter!