k-Forrelation Optimally Separates Quantum and Classical Query Complexity

QIP 2021

Nikhil Bansal

CWI

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

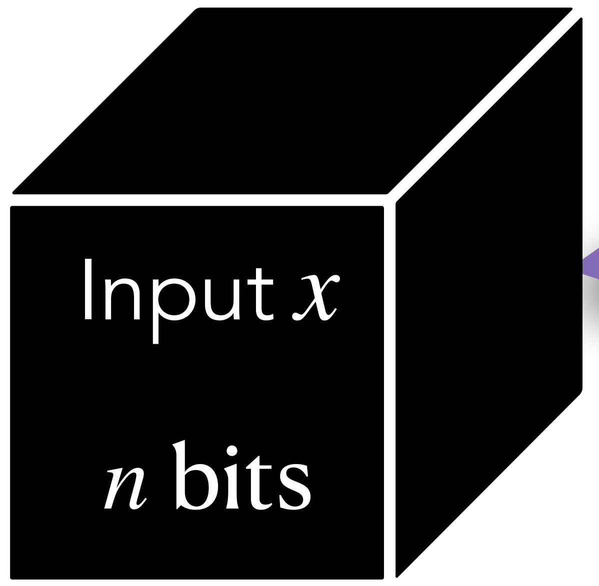
Makrand Sinha

CWI

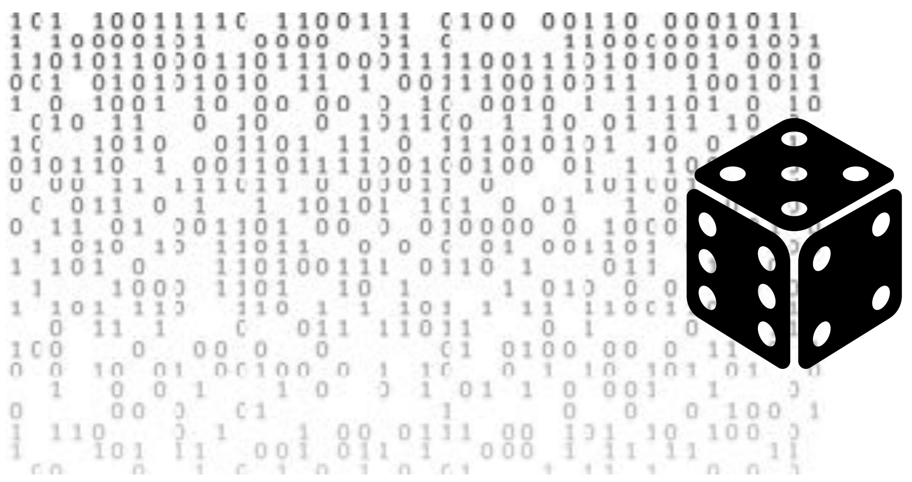
Quantum vs Classical Query Algorithms

Quantum Query Algorithm

Can query input bits in superposition



Choose bit to query at random



Randomized Query Algorithm

Probability Distribution over Decision Trees

Objective(s) Minimize queries

Quantum Speedup Question	What is the maximal separation? [Buhrman-Fortnow-Newman-Röhrig '03]

Total functions

Classical queries = (Quantum queries)^c

$c \leq 6$	[Beals-Buhrman-Cleve-Mosca-de Wolf '98]
$c \leq 4$	[Aaronson-Ben David-Kothari-Rao-Tal '21]
$c \geq 5/2$	[Aaronson-Ben David-Kothari '16]
$c \geq 8/3 - o(1)$	[Tal '20] Non-explicit

Partial functions

$O(\log^2 n)$ vs $\tilde{\Omega}(n^{1/2})$	[Simon '97] [Childs-Cleve-Deotto-Farhi-Gutmann-Spielman '03]
1 vs $\tilde{\Omega}(n^{1/4})$	[Beaudrap-Cleve-Watrous '02]
1 vs $\tilde{\Omega}(n^{1/2})$	[Aaronson-Ambainis '14]
$O(1)$ vs $\tilde{\Omega}(n^{2/3-\epsilon})$	[Tal '20] Non-explicit

Maximal Separation?

Is this optimal?

Theorem [Aaronson-Ambainis '14]

Every $\lceil k/2 \rceil$-query quantum algorithm with error $\frac{1}{2} - \delta$ can be simulated with error $\frac{1}{2} - \frac{\delta}{2}$ with $2^k \cdot n^{1-1/k} \cdot \delta^{-2}$ classical queries

$\tilde{O}(n^{1/2})$ classical queries
$k = 2$

$\tilde{O}(n^{0.999})$ classical queries
$k = 1000$

What is the task where quantum algorithms have the maximal advantage?

Conjecture [Aaronson-Ambainis '14]

$\boxed{k\text{-fold Forrelation problem}}$ gives a $\lceil k/2 \rceil$ vs $\tilde{\Omega}(n^{1-1/k})$ separation

Captures the maximal power of
- Quantum query algorithms?
- Quantum circuits

  BQP-complete for promise problems for $k \approx \log n$
  [Aaronson-Ambainis '14]

2-Fold Forrelation

Promise Problem Input $x := (x_1, x_2) \in \{\pm 1\}^{n+n}$

Distinguish these cases

$$\left|\frac{\langle x_1, \boxed{Hx_2}\rangle}{n}\right| \leq 0.01 \quad \text{vs} \quad \frac{\langle x_1, Hx_2\rangle}{n} \geq 0.1$$

Almost Orthogonal | **Far from Orthogonal**

[Aaronson '10]

Fourier transform

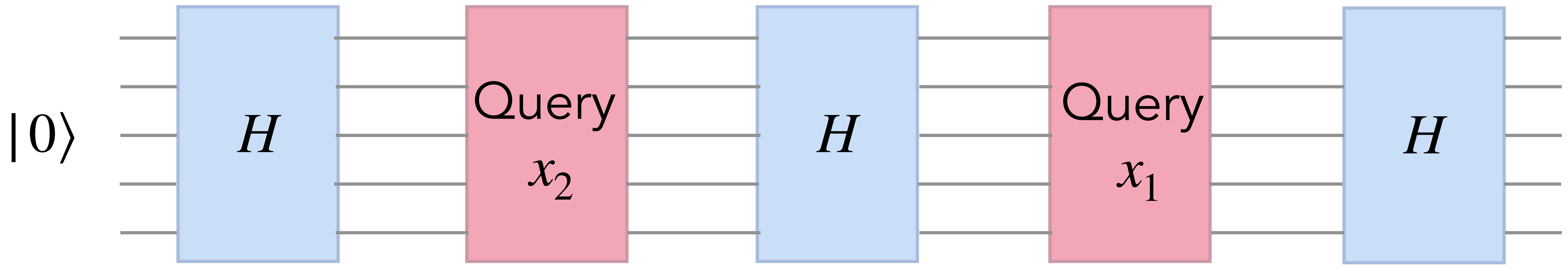
Can be solved with 1 query

Rotation

H = Hadamard/Fourier matrix

Each entry $\pm\frac{1}{\sqrt{n}}$

n



amplitude of $|0\rangle = \frac{1}{n}\sum_{i,j=1}^{n} x_1(i)H_{ij}x_2(j) = \frac{\langle x_1, Hx_2\rangle}{n}$

2-Fold Forrelation

Promise Problem Input $x := (x_1, x_2) \in \{\pm 1\}^{n+n}$

Distinguish these cases

$|\mathrm{Forr}_2(x)| \leq 0.01$ **vs** $\mathrm{Forr}_2(x) \geq 0.1$

Can be solved with 1 query

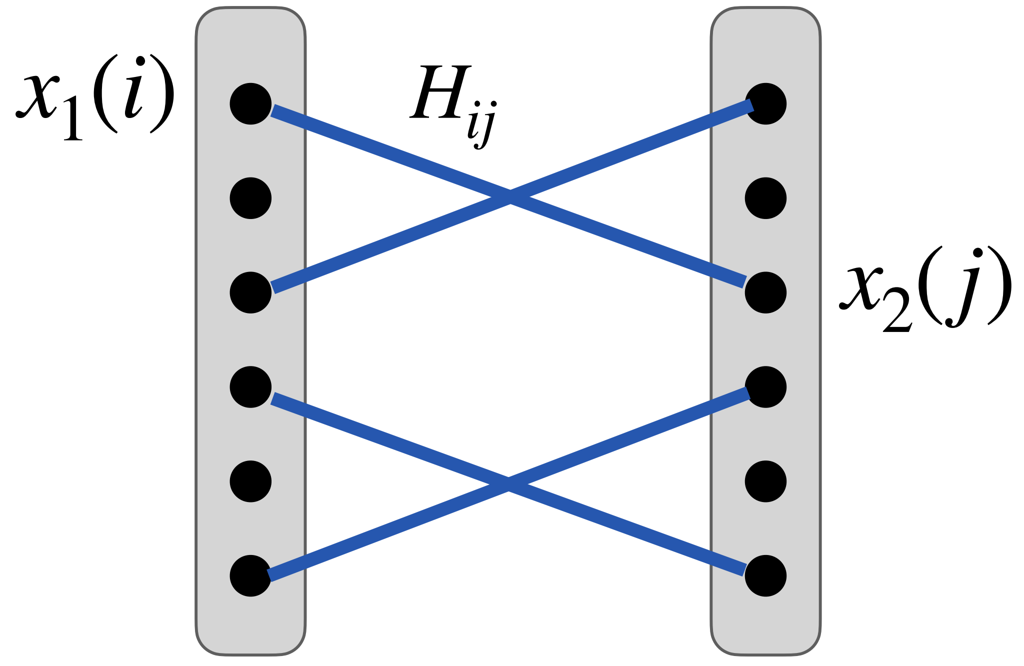
Rotation

H = Hadamard/Fourier matrix

Each entry

$\pm \frac{1}{\sqrt{n}}$

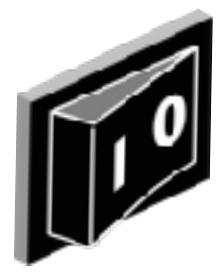
$\longleftarrow n \longrightarrow$



$$\mathrm{Forr}_2(x) = \frac{1}{n} \sum_{i,j=1}^{n} x_1(i) \cdot H_{ij} \cdot x_2(j) = \frac{\langle x_1, H x_2 \rangle}{n}$$

k-Fold Forrelation

Promise Problem Input $x := (x_1, \ldots, x_k) \in \{\pm 1\}^{kn}$



Distinguish these cases

$|\mathrm{Forr}_k(x)| \leq 0.01$ vs $\mathrm{Forr}_k(x) \geq 0.1$

Rotation

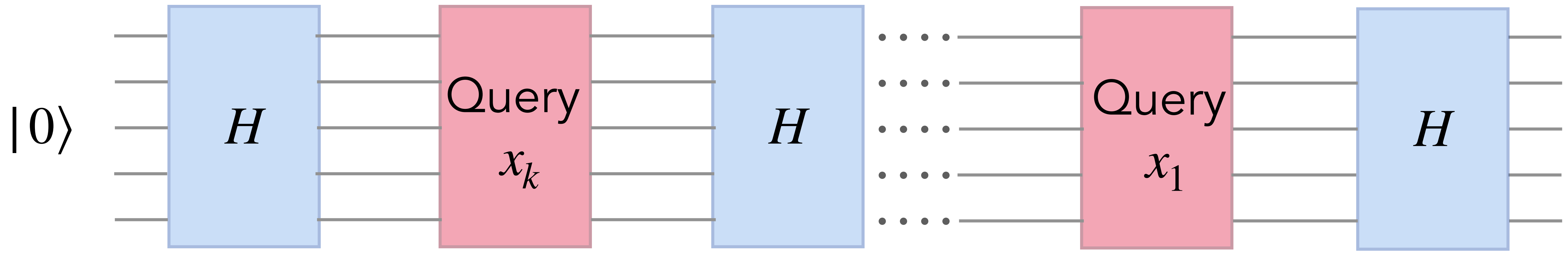
H = Hadamard/Fourier matrix

Each entry

$\pm\frac{1}{\sqrt{n}}$

n

Can be solved with $\lceil k/2 \rceil$ queries



$$\mathrm{Forr}_k(x) = \boxed{\frac{1}{n} \sum_{i_1, \ldots, i_k = 1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)}$$

amplitude of $|0\rangle$

k-Fold Forrelation

Promise Problem Input $x := (x_1, \ldots, x_k) \in \{\pm 1\}^{kn}$

Distinguish these cases

$|\mathrm{Forr}_k(x)| \le 0.01$ **vs** $\mathrm{Forr}_k(x) \ge 0.1$

Can be solved with $\lceil k/2 \rceil$ queries

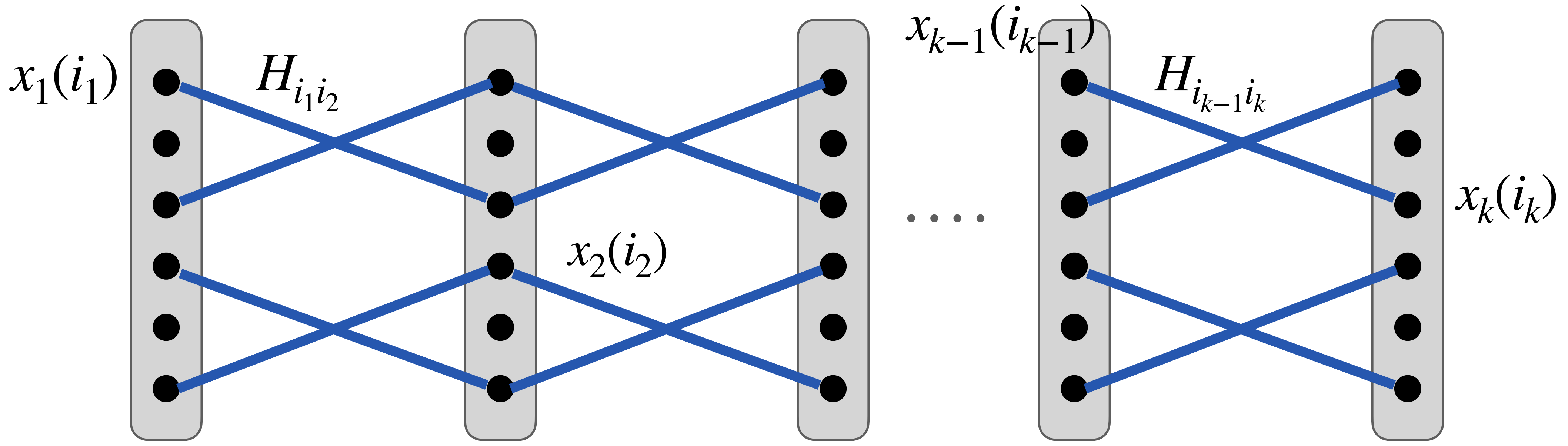
Rotation

H = Hadamard/Fourier matrix

Each entry

$\pm \frac{1}{\sqrt{n}}$

$\longleftarrow n \longrightarrow$



$$\mathrm{Forr}_k(x) = \frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

Our Results

Theorem

k-fold Forrelation problem gives a $\lceil k/2 \rceil$ vs $\tilde{\Omega}(n^{1-1/k})$ separation between quantum and classical query algorithms for advantage $\delta = 2^{-O(k)}$

500 vs $\tilde{\Omega}(n^{0.999})$

$k = 1000$

Main Contribution: classical lower bound

Previous lower bound: $\tilde{\Omega}(n^{1/2})$
[Aaronson-Ambainis '14]

Our proof also works for the non-explicit Rorrelation function introduced by [Tal '20]

Replace Hadamard with a Random Orthogonal matrix

$$\mathrm{Forr}_k(x) = \frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

Our Results

Theorem k-fold Forrelation problem gives a $\lceil k/2 \rceil$ vs $\tilde{\Omega}(n^{1-1/k})$ separation between quantum and classical query algorithms for advantage $\delta = 2^{-O(k)}$

500 vs $\tilde{\Omega}(n^{0.999})$
$k = 1000$

Main Contribution: classical lower bound

Previous lower bound: $\tilde{\Omega}(n^{1/2})$
[Aaronson-Ambainis '14]

Our proof also works for the non-explicit Rorrelation function introduced by [Tal '20]

Consequences

- **Query Complexity of Partial Functions with standard error**

  $O_\epsilon(1)$ vs $n^{1-\epsilon}$ separation for error $1/3$

- **Query Complexity of Total Functions with standard error**

  $\exists$total f Classical queries $\geq$ (Quantum queries)$^{3-o(1)}$

- Analogous separations in **Communicati**

Explicit

[SSW '21] + [Tal '20] rely on strong properties of random orthogonal matrices that do not hold for Hadamard matrix

Independent Work Analogous results for Rorrelation [Sherstov-Storozhenko-Wu '21] building on [Tal '20] Different Techniques

High-level Overview

Quantum vs Classical Query Algorithms

Fact Success probability of any d-query quantum or randomized algorithm is a degree $O(d)$ multilinear polynomial

$$f(z) = \sum_{S \subseteq [N]} \hat{f}(S) \cdot z_S \quad \text{where } z \in \{\pm 1\}^N \text{ and } z_S = \prod_{i \in S} z_i$$

Quantum Query Algorithm

Extremely good at computing dense polynomials with few queries

$$\text{Forr}_k(x) = \frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

Randomized Query Algorithm

Can only compute weakly-sparse polynomials [Tal '20] [SSW '21]

L_1-norm of coefficients of degree ℓ monomials $\ll$ number of degree ℓ monomials

$$\sum_{|S|=\ell} |\partial_S f(0)| = \sum_{|S|=\ell} |\hat{f}(S)| \le \sqrt{\binom{d}{\ell}} \ll \binom{d}{\ell} \quad \text{for all } \ell \le d$$

Observation [This Work]

$$\sum_{|S|=\ell} |\partial_S f(x)| \le \sqrt{\binom{d}{\ell}}$$

for any ℓ and any $x \in [-1,1]^N$

Rest of the talk

Multilinear polynomial p of degree $d \ll n^{1-1/k}$ **with small derivatives** cannot compute k-Fold Forrelation

We only need bound on derivatives of order $\le k^2$ which follow from [Tal '20]

Degree Lower Bounds

Multilinear polynomial p of degree $d \ll n^{1-1/k}$ **with small derivatives** cannot compute k-Fold Forrelation

Input $(x_1, \ldots, x_k) \in \{\pm 1\}^{kn}$

$$\frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

$\mathrm{Forr}_k(x)$

Show that such polynomials cannot distinguish distributions on 0 vs 1 inputs

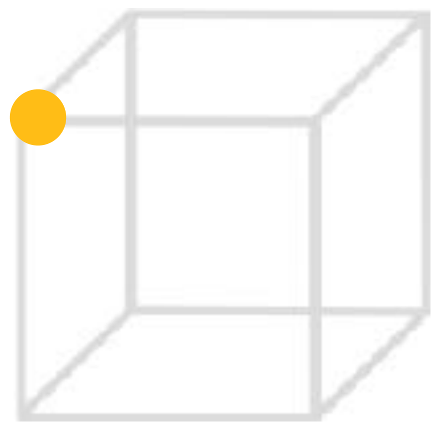
$|\mathrm{Forr}_k(x)| \leq 0.01$

vs

$\mathrm{Forr}_k(x) \geq 0.1$

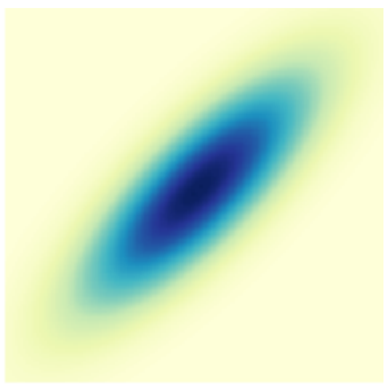
$\mathbb{E}[p(\mathscr{F}_k)] - \mathbb{E}[p(\mathscr{U})] \approx 0$

Uniform Distribution $\mathscr{U}$



Uniform Distribution on $\{\pm 1\}^{kn}$

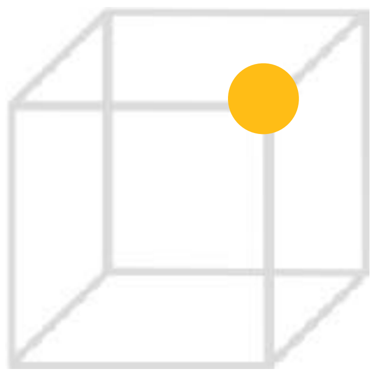
Pseudorandom Distribution $\mathscr{F}_k$



Sample from some distribution $\mathscr{P}_k$ over $\mathbb{R}^{kn}$

Rounding

e.g. take sign



Distribution on $\{\pm 1\}^{kn}$

Degree Lower Bounds via Interpolation

Multilinear polynomial p of degree $d \ll n^{1-1/k}$ **with small derivatives** cannot compute k-Fold Forrelation

Input $(x_1, \ldots, x_k) \in \{\pm 1\}^{kn}$

$$\frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

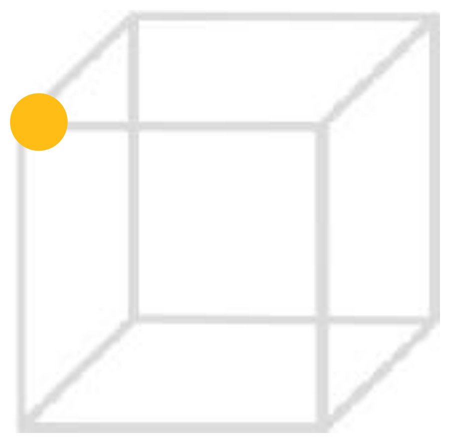
$\mathrm{Forr}_k(x)$

$$\mathbb{E}[p(\mathcal{F}_k)] - \mathbb{E}[p(\mathcal{U})] \approx 0$$

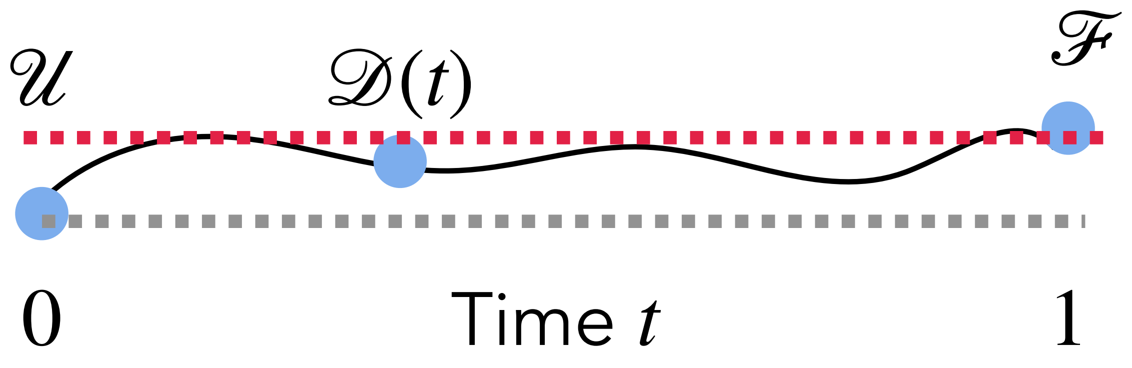
Labeled "Smart Path method" by Talagrand

Many applications in statistical physics, probability, convex geometry,…

Uniform Distribution $\mathcal{U}$

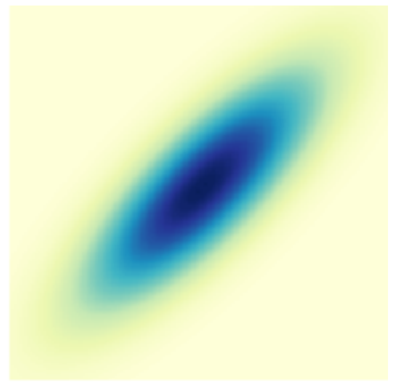


Uniform Distribution on $\{\pm 1\}^{kn}$



Choose smart path to control time derivative of $\mathbb{E}[p(\mathcal{D}(t))]$

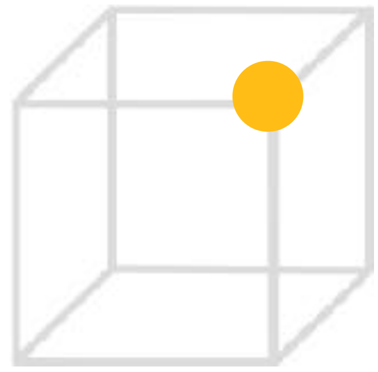
Pseudorandom Distribution $\mathcal{F}_k$



Sample from some distribution $\mathcal{P}_k$ over $\mathbb{R}^{kn}$

Rounding

e.g. take sign



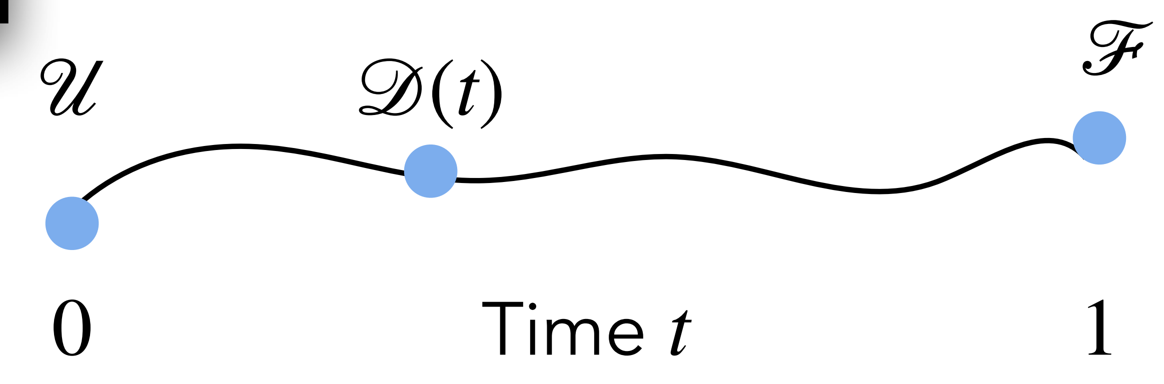
Distribution on $\{\pm 1\}^{kn}$

Our Main Technical Contribution

Multilinear polynomial p of degree $d \ll n^{1-1/k}$ **with small derivatives** cannot compute k-Fold Forrelation

$$\mathbb{E}[p(\mathcal{F}_k)] - \mathbb{E}[p(\mathcal{U})] \approx 0$$

Choose smart path to control time derivative of $\mathbb{E}[p(\mathcal{D}(t))]$



Lemma For every "time" t

$$\text{"Time derivative"} \leq \max_{x \in [-1,1]^{kn}} \sum_{\ell=k}^{k(k-1)} \left(\frac{1}{\sqrt{n}}\right)^{\ell(1-1/k)} \sum_{|S|=\ell} |\partial_S p(x)|$$

Recall bound on derivatives [Tal '20] [SSW '21]

ℓ^{th}-order $\leq d^{\ell/2}$

+ Our Observation

2-Fold [Raz-Tal '18] [Wu '19]

$$\leq \max_{x \in [-1,1]^{kn}} \frac{1}{\sqrt{n}} \sum_{|S|=2} |\partial_S p(x)| \leq \frac{d}{\sqrt{n}}$$

2nd order

$\boxed{\textbf{Choose } d \approx n^{1/2}}$

3-Fold [This Work]

$$\leq \max_{x \in [-1,1]^{kn}} \frac{1}{n} \sum_{|S|=3} |\partial_S p(x)| + \frac{1}{n^2} \sum_{|S|=6} |\partial_S p(x)| \leq \frac{d^{3/2}}{n} + \frac{d^3}{n^2}$$

3rd order — 6th order

$\boxed{\textbf{Choose } d \approx n^{2/3}}$

Relies on stochastic calculus tools

- Gaussian Interpolation
- Gaussian Integration by Parts
- **Develop new Integration by Parts identities** for rounding

Proof Ideas

Degree Lower Bounds via Interpolation

Multilinear polynomial p of degree $d \ll n^{1-1/k}$ **with small derivatives** cannot compute k-Fold Forrelation

Input $(x_1, \ldots, x_k) \in \{\pm 1\}^{kn}$

$$\frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} x_1(i_1) H_{i_1 i_2} x_2(i_2) H_{i_2 i_3} \ldots H_{i_{k-1} i_k} x_k(i_k)$$

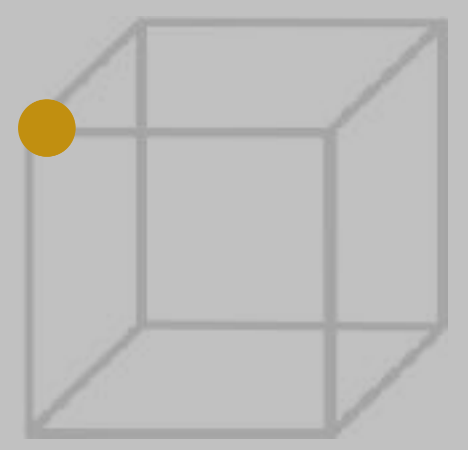
$\mathrm{Forr}_k(x)$

$$\mathbb{E}[p(\mathscr{F}_k)] - \mathbb{E}[p(\mathscr{U})] \approx 0$$

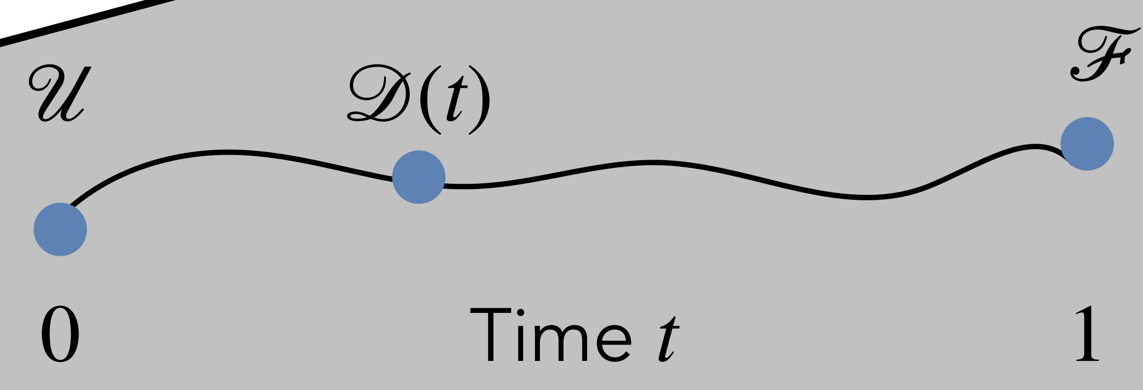
Recall

Small Value

Uniform Distribution $\mathscr{U}$



Uniform Distribution on $\{\pm 1\}^{kn}$



Choose smart path to control time derivative of $\mathbb{E}[p(\mathscr{D}(t))]$

Large Value

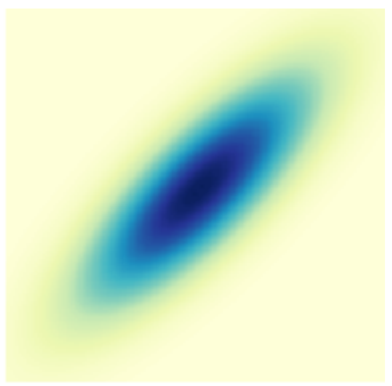
Pseudorandom Distribution $\mathscr{F}_k$

Sample from some distribution $\mathscr{P}_k$ over $\mathbb{R}^{kn}$

Distribution on $\{\pm 1\}^{kn}$

2-Fold Case: Pseudorandom Distribution

Pseudorandom Distribution $\mathscr{F}_k$



Sample from some distribution $\mathscr{P}_k$ over $\mathbb{R}^{kn}$

Roundin…

Not in this talk

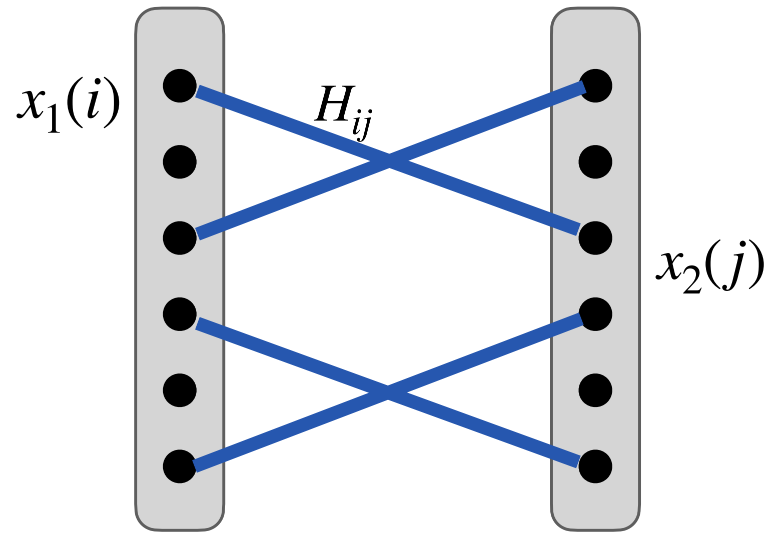
New Techniques in our work

Distribution on $\{\pm 1\}^{kn}$

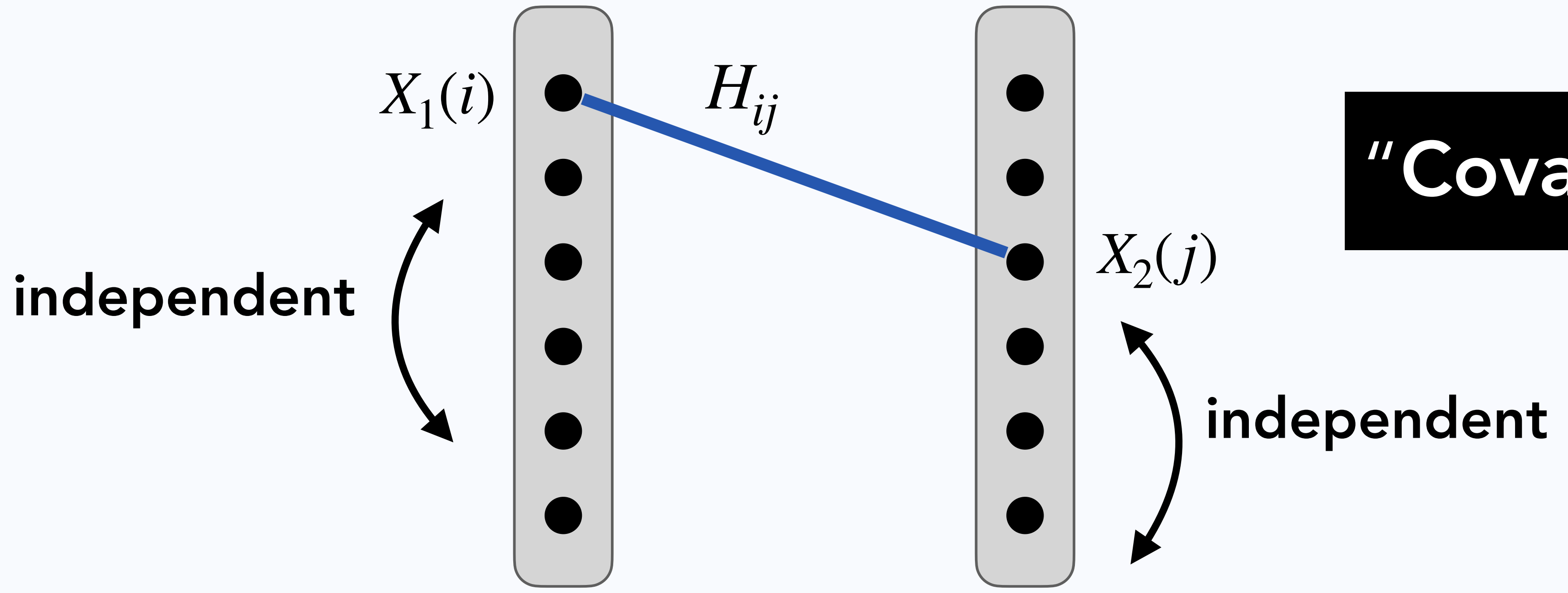
$\text{Forr}_k(x) \geq 0.1$

$k = 2$

Input $(x_1, x_2) \in \{\pm 1\}^{2n}$



$$\text{Forr}_2(x) = \frac{1}{n}\sum_{i,j=1}^{n} x_1(i) \cdot H_{ij} \cdot x_2(j)$$



"Covariance Graph"

$\mathbb{E}[X_1(i)X_2(j)] = H_{ij}$

$X := (X_1, X_2)$ Gaussian in $\mathbb{R}^{2n}$ with Covariance $\Sigma = \begin{pmatrix} I_n & H_n \\ H_n & I_n \end{pmatrix}$

[Aaronson-Ambainis '14]

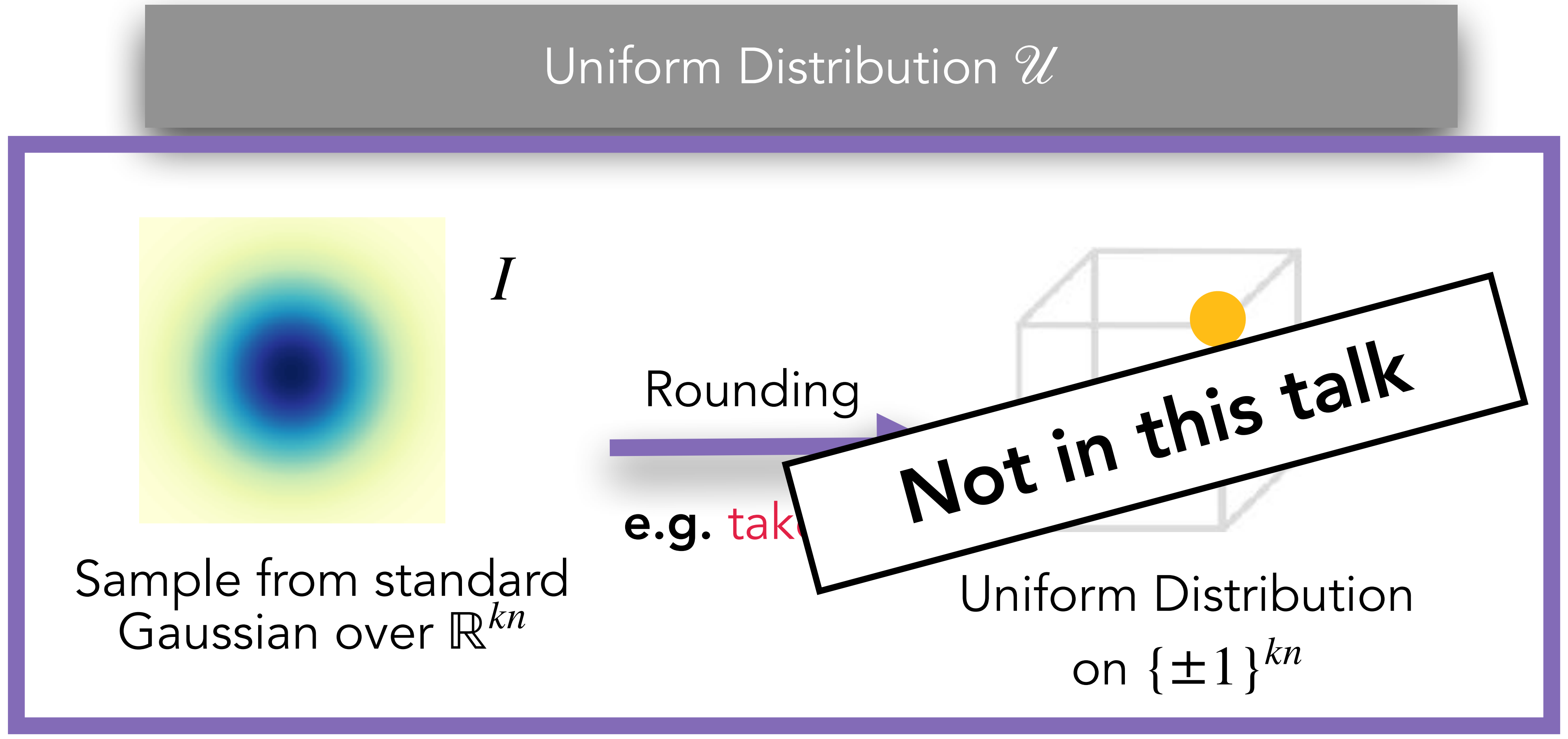
Intuition

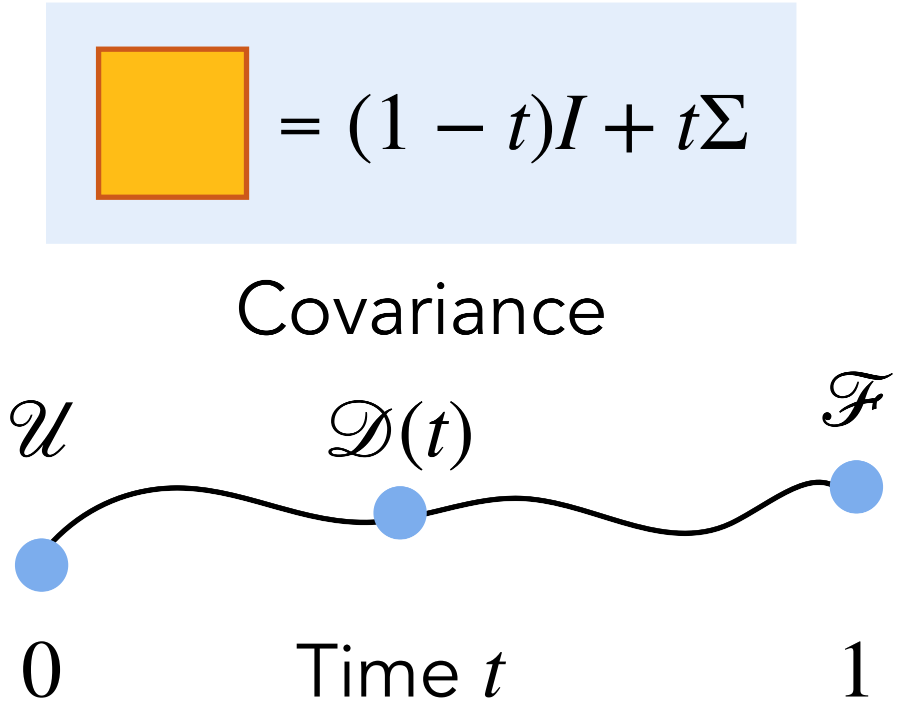
Decision tree or multilinear polynomial needs to compute all 2-wise correlations

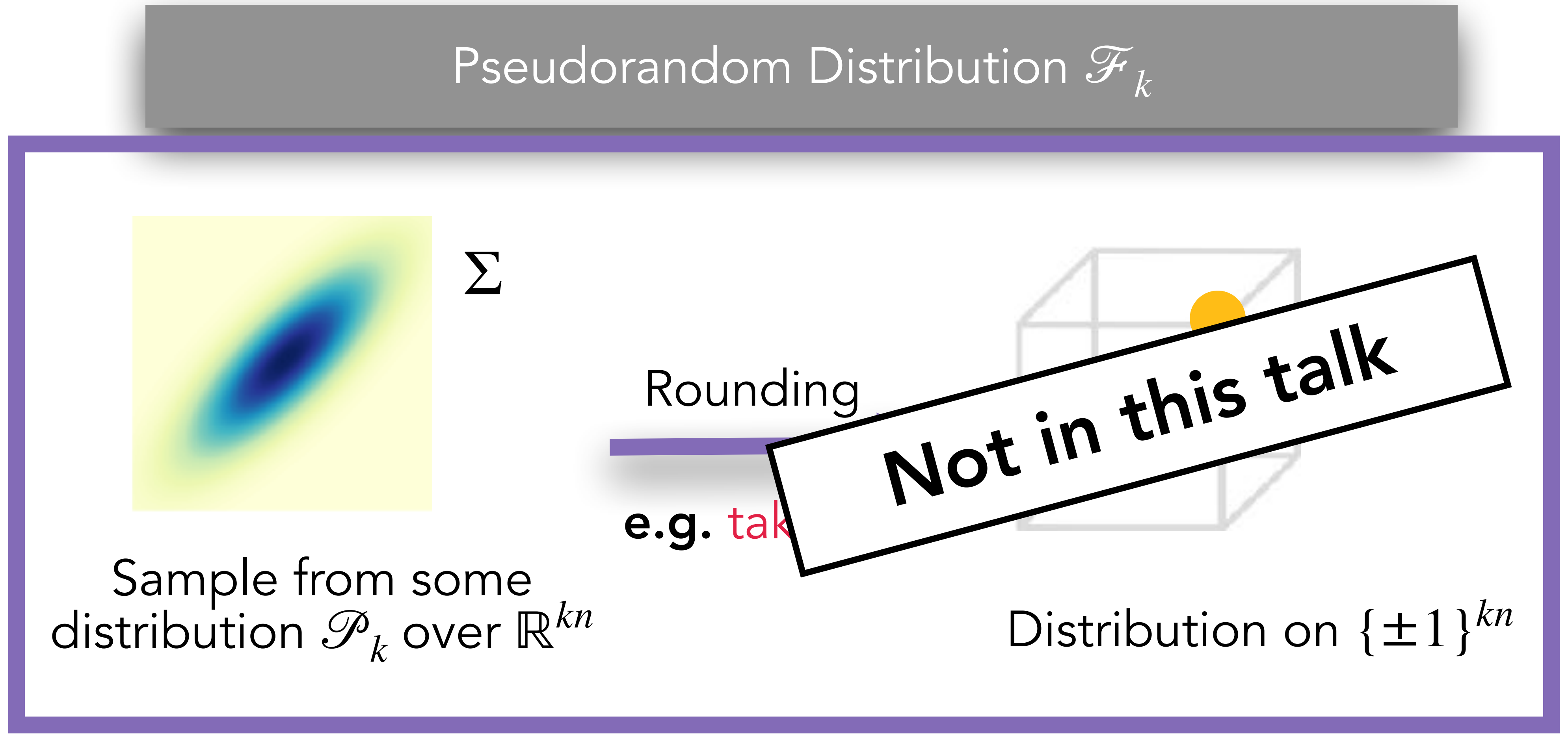
$$H_{ij} = \pm \frac{1}{\sqrt{n}}$$

$$\boxed{\mathbb{E}[\text{Forr}_2(X)] = \frac{1}{n}\sum_{ij} H_{ij}^2 = 1}$$

2-Fold Case: The Smart Path



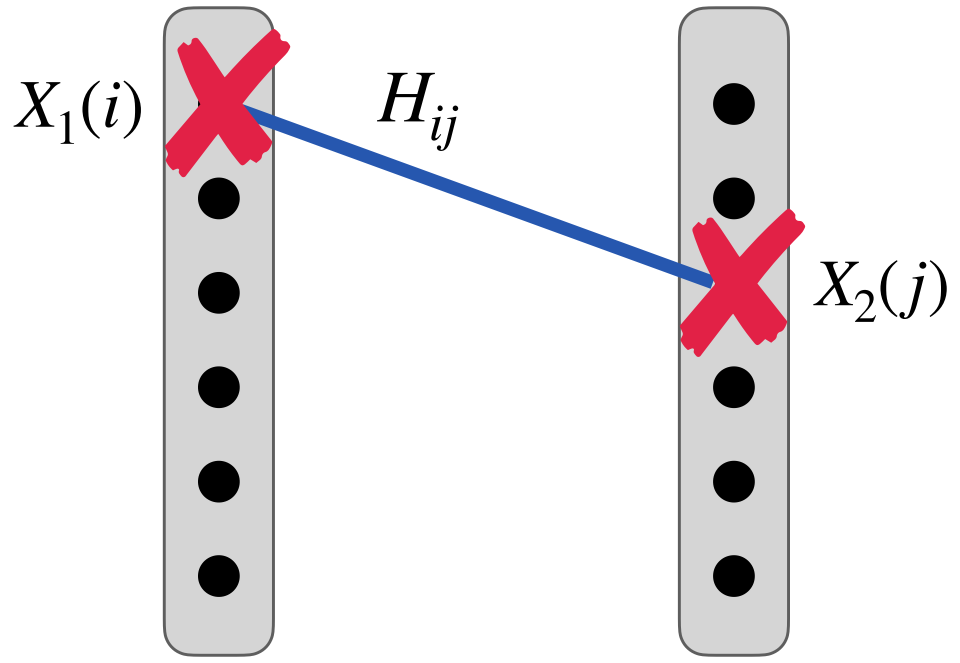




Can be directly handled using **Gaussian Interpolation Formula**

Multilinear polynomial p → Bound in terms of $\partial_{ij} p(x)$ and final covariance entries $H_{ij} = \pm \frac{1}{\sqrt{n}}$

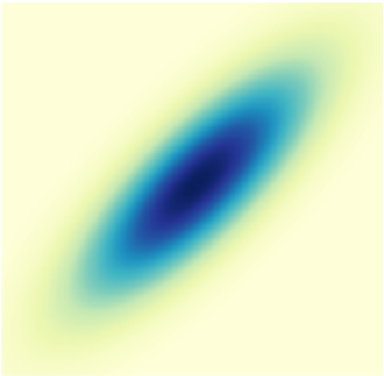
$$\boxed{\text{"Time derivative"} \leq \max_{x \in [-1,1]^{kn}} \frac{1}{\sqrt{n}} \sum_{ij} |\partial_{ij} p(x)|}$$

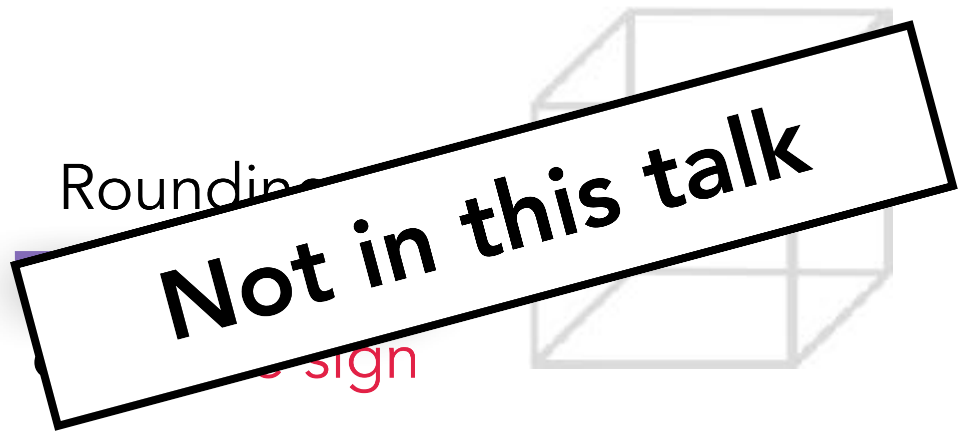


∂_{ij} corresponds to removing $x_i x_j$ **e.g.** $x_1 x_2 x_3 \cdots x_i x_j$

3-Fold Case: Pseudorandom Distribution

Pseudorandom Distribution $\mathscr{F}_k$





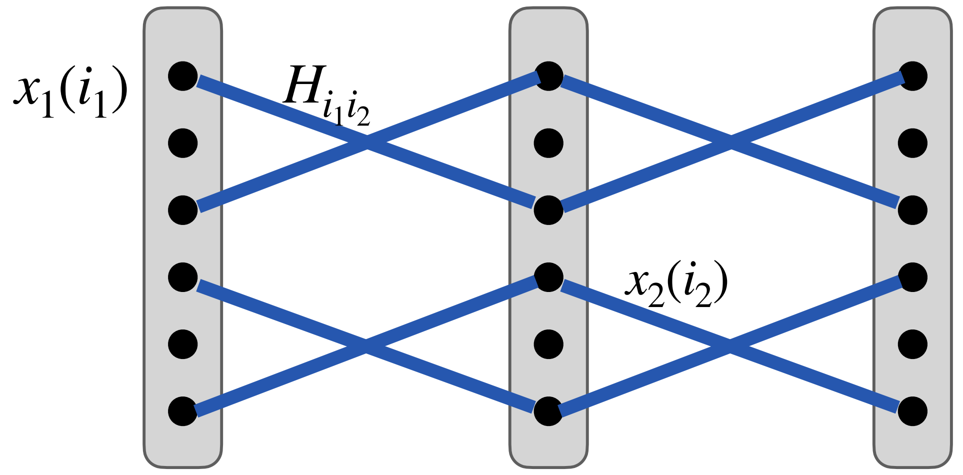
Sample from some distribution $\mathscr{P}_k$ over $\mathbb{R}^{kn}$

Distribution on $\{\pm 1\}^{kn}$

$\text{Forr}_k(x) \geq 0.1$

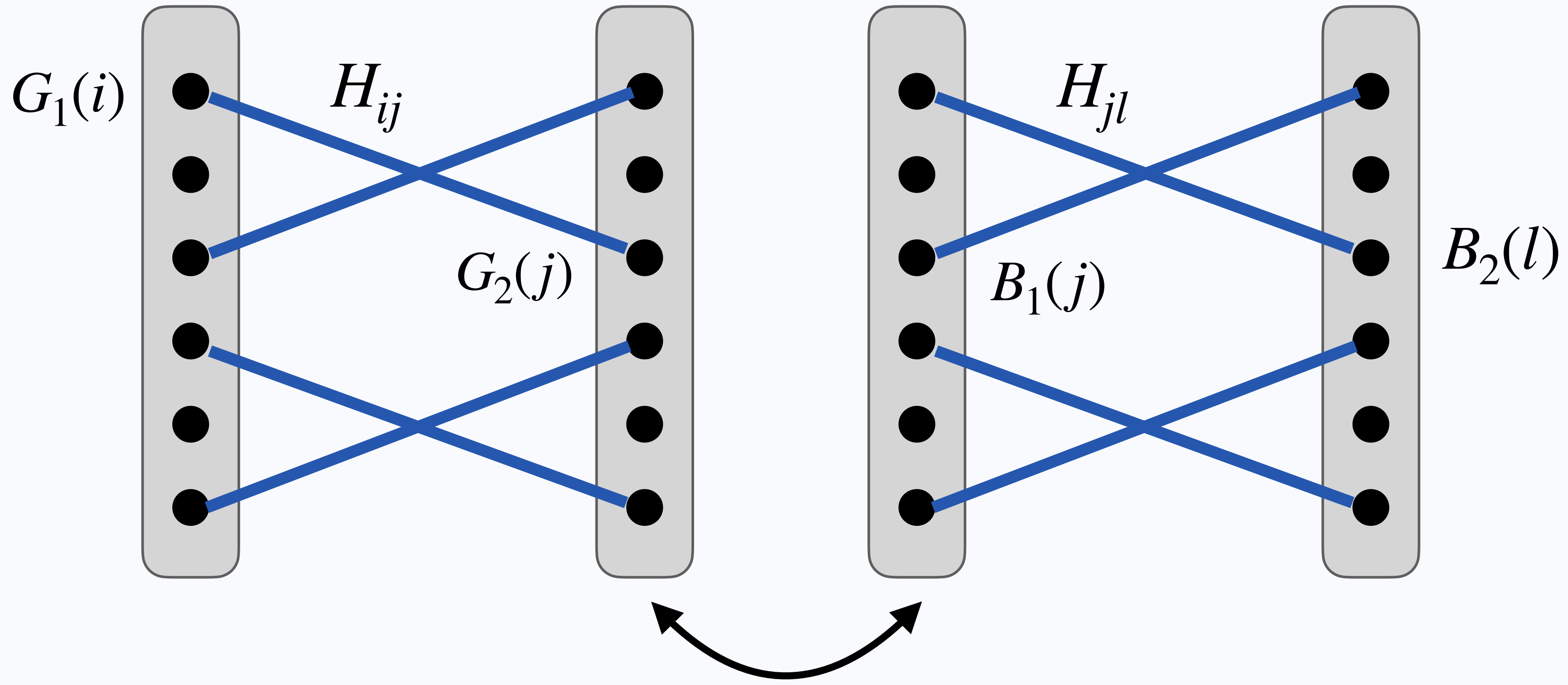
$k = 3$

Input $(x_1, x_2, x_3) \in \{\pm 1\}^{3n}$



$$\text{Forr}_3(x) = \frac{1}{n} \sum_{i,j,l}^{n} x_1(i) H_{ij} x_2(j) H_{jl} x_3(l)$$

G, B independent Gaussians in $\mathbb{R}^{2n}$ **with covariance** $\Sigma = \begin{pmatrix} I_n & H_n \\ H_n & I_n \end{pmatrix}$



Take product of these random variables

$X := (G_1, G_2 \odot B_1, B_2) \in \mathbb{R}^{3n}$ Entry-wise product

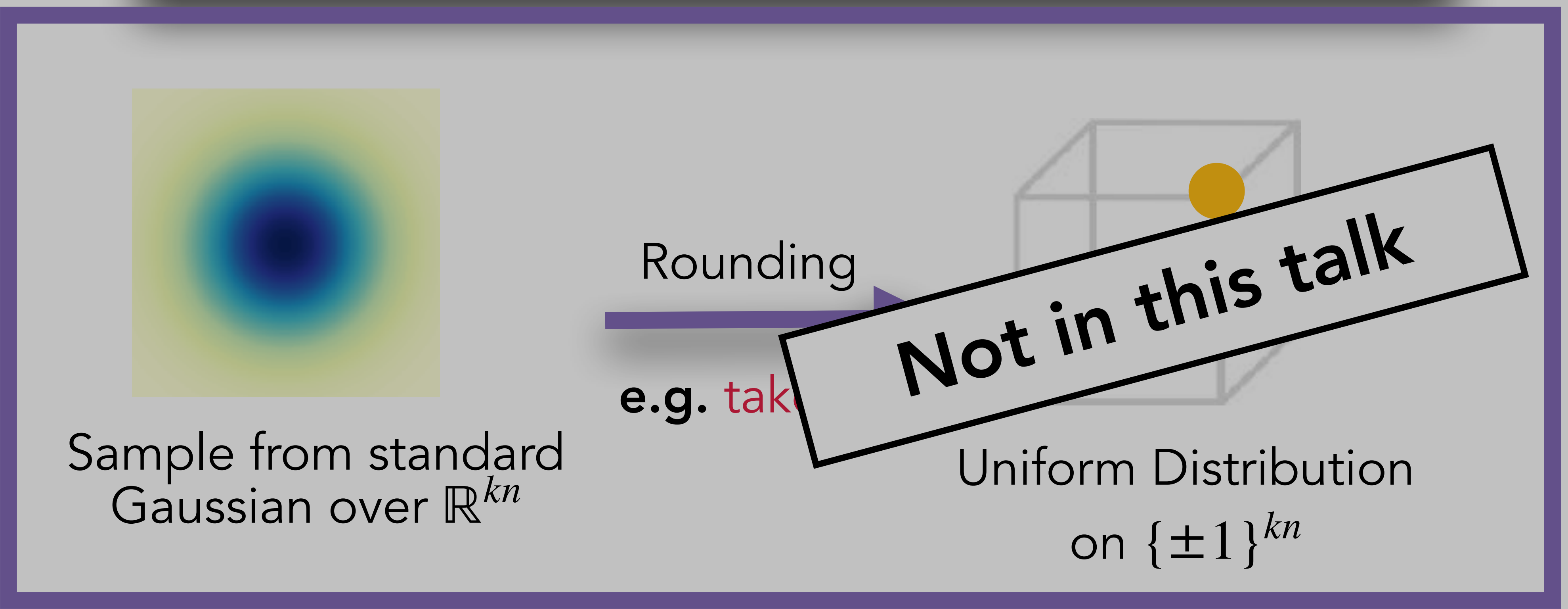
[Tal '20]

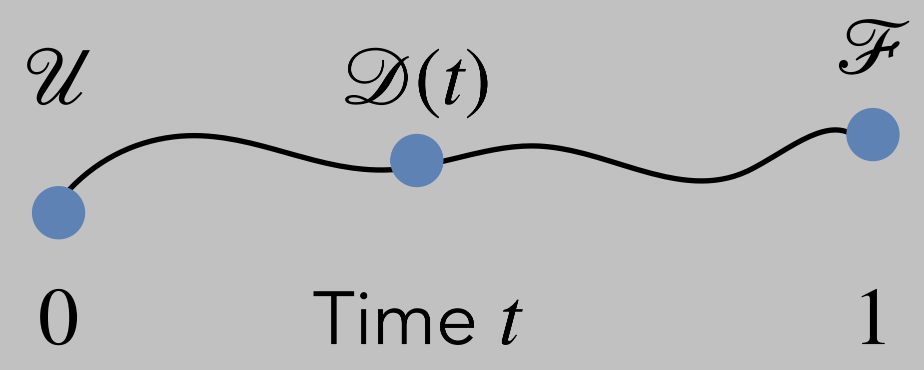
$\mathbb{E}[\text{Forr}_3(X)] = 1$

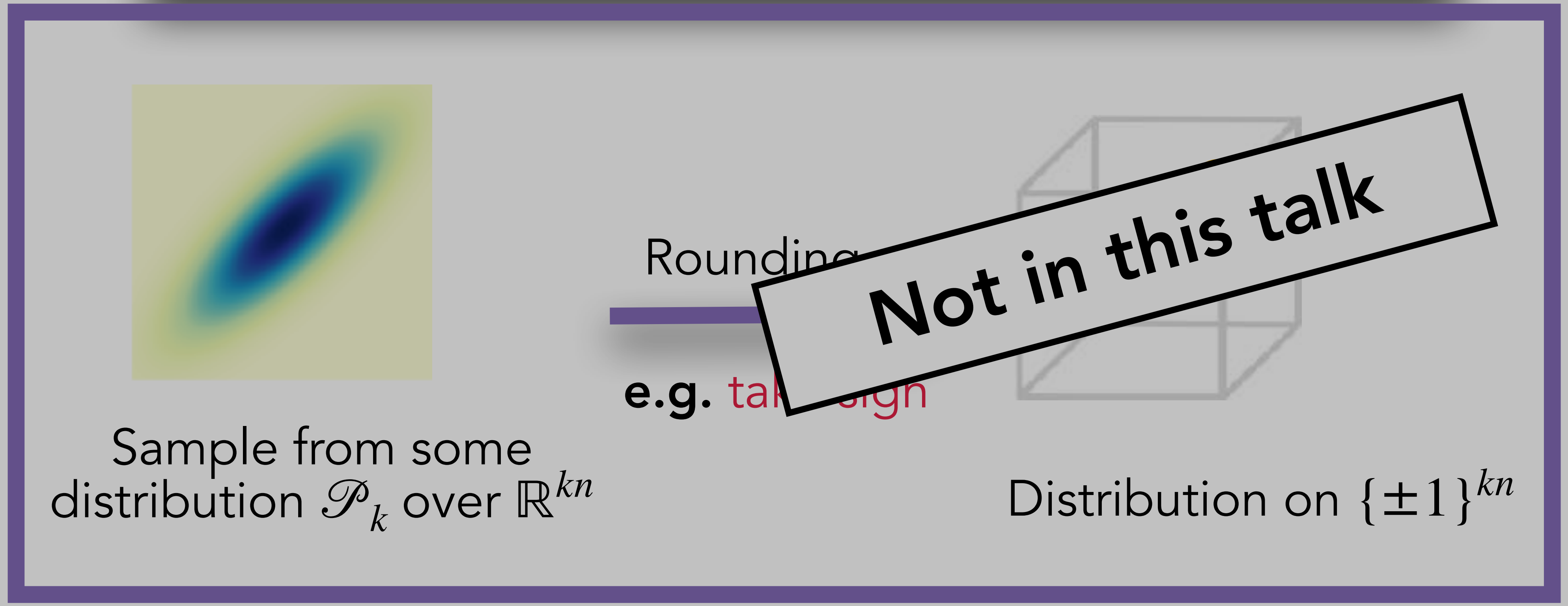
Intuition

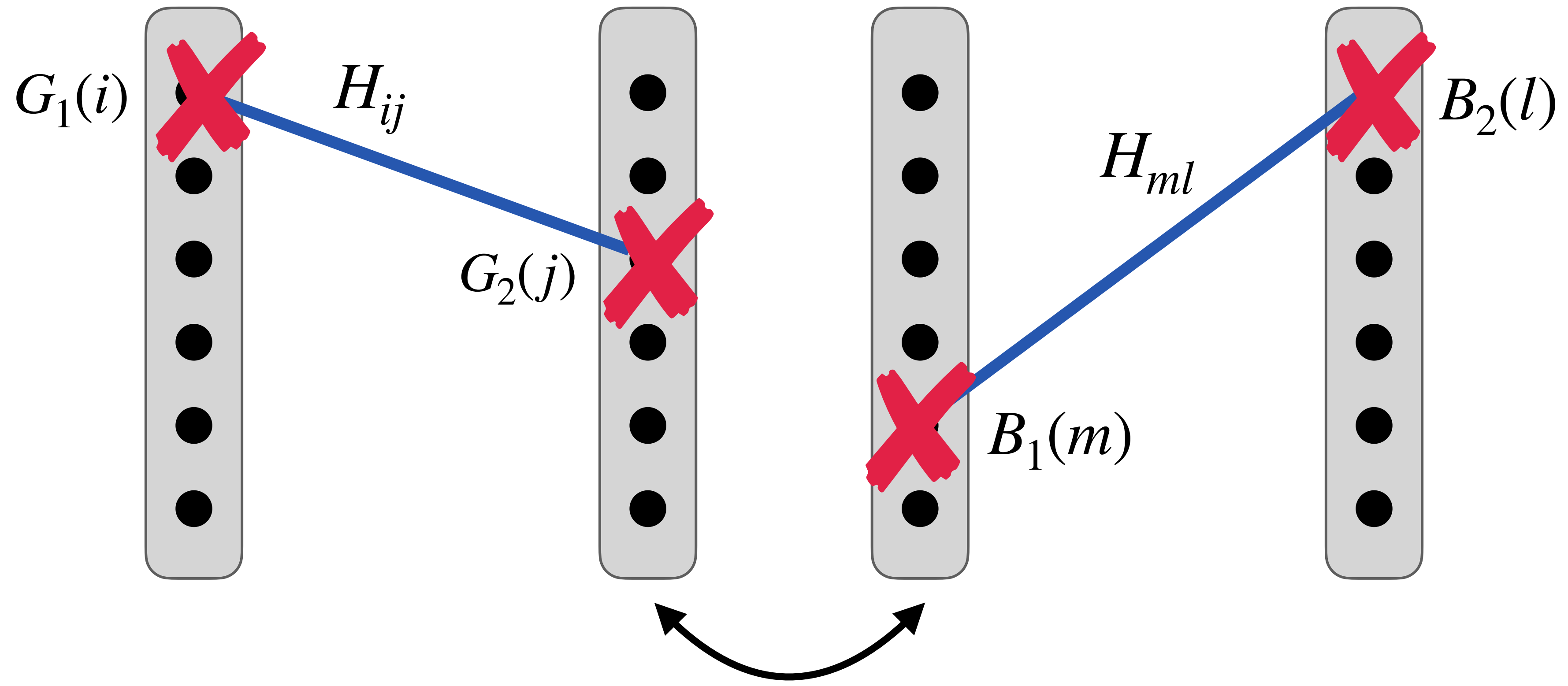
Decision tree or multilinear polynomial needs to compute all three-wise correlations now

3-Fold Case: The Smart Path









Take product of these random variables

$X := (G_1, G_2 \odot B_1, B_2) \in \mathbb{R}^{3n}$

Interpolate G and B separately

Want bounds in terms of $\partial_{ij\ell} p$ and sixth order derivatives

$z_1 = g_1,\ z_2 = g_2 \cdot b_2,\ z_3 = b_3$

$p(z) = \cdots + \cdots\ \ z_1 z_2 z_3\ \ \cdots + \cdots$

$||$

$g_1 g_2 \cdot b_2 b_3$

Multilinear polynomial p

Other stochastic calculus tools **e.g. Gaussian Integration by Parts** to relate derivatives after substitution

Summary and Open Problems

Theorem k-fold Forrelation problem gives a $\lceil k/2 \rceil$ vs $\tilde{\Omega}(n^{1-1/k})$ separation between quantum and classical query algorithms for advantage $\delta = 2^{-O(k)}$

Optimal Separation

Relies on stochastic calculus tools

- Gaussian Interpolation
- Gaussian Integration by Parts
- **Develop new Integration by Parts identities** for rounding

Open Problem Quantum vs Classical Communication Complexity of **Total Functions**

Are these polynomially related?