# Nearly Optimal Bounds for Sample-Based Testing and Learning of $k$-Monotone Functions

RANDOM 2024

Hadley Black
UC San Diego

# Property Testing [Rubinfeld-Sudan 96, Goldreich-Goldwasser-Ron 98]

- Universe of functions $F: X \to R$
- Property $P \subset F$
- Distance:

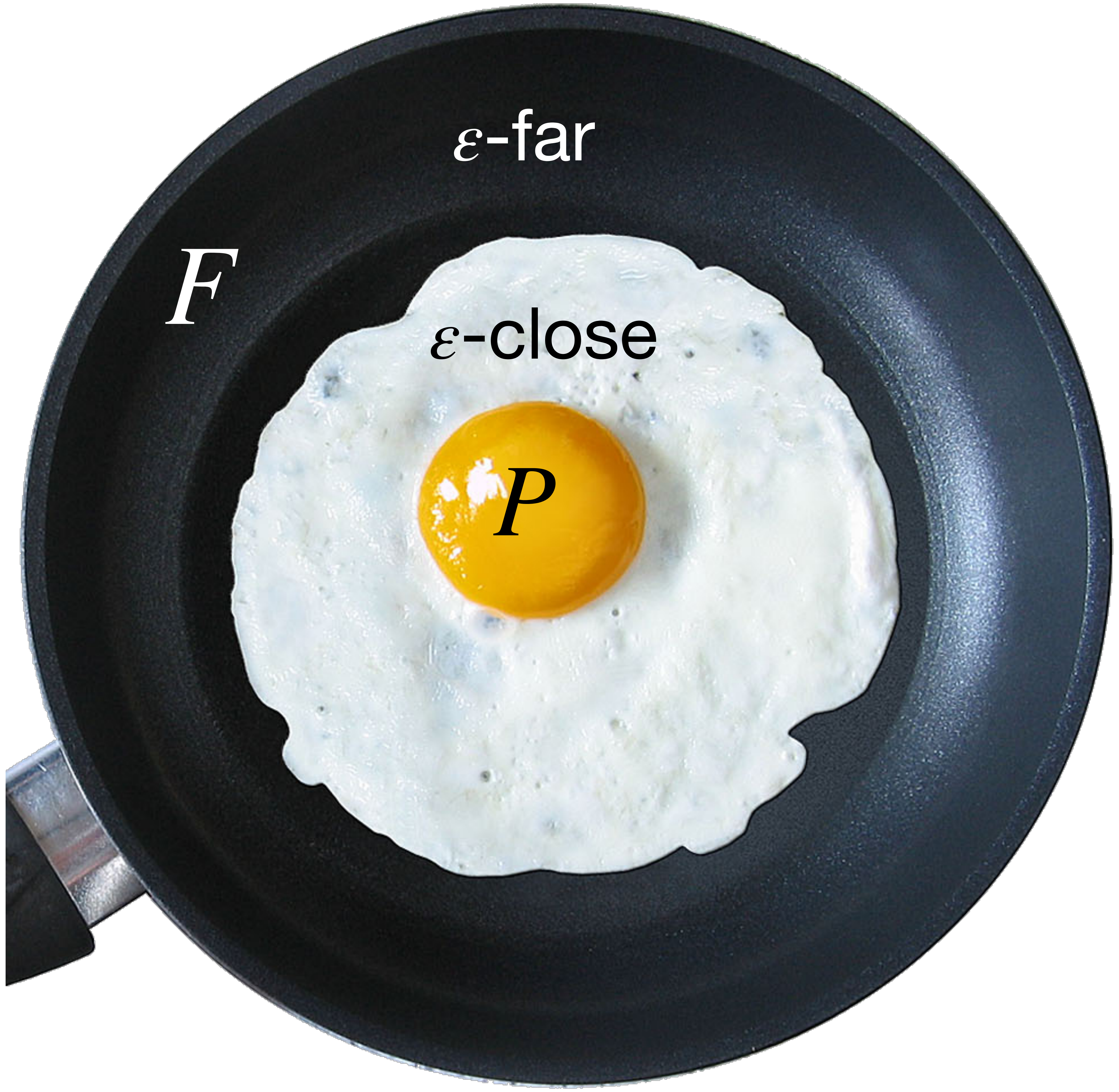
$$d(f,g) = \Pr_{x \in X}[f(x) \neq g(x)]$$

$$d(f,P) = \min_{g \in P} d(f,g)$$

**Property Tester**

Given oracle access to $f \in F$ and $\varepsilon > 0$:

1. if $f \in P$: **accept** with prob. $> 2/3$
2. if $d(f,P) > \varepsilon$: **reject** with prob. $> 2/3$

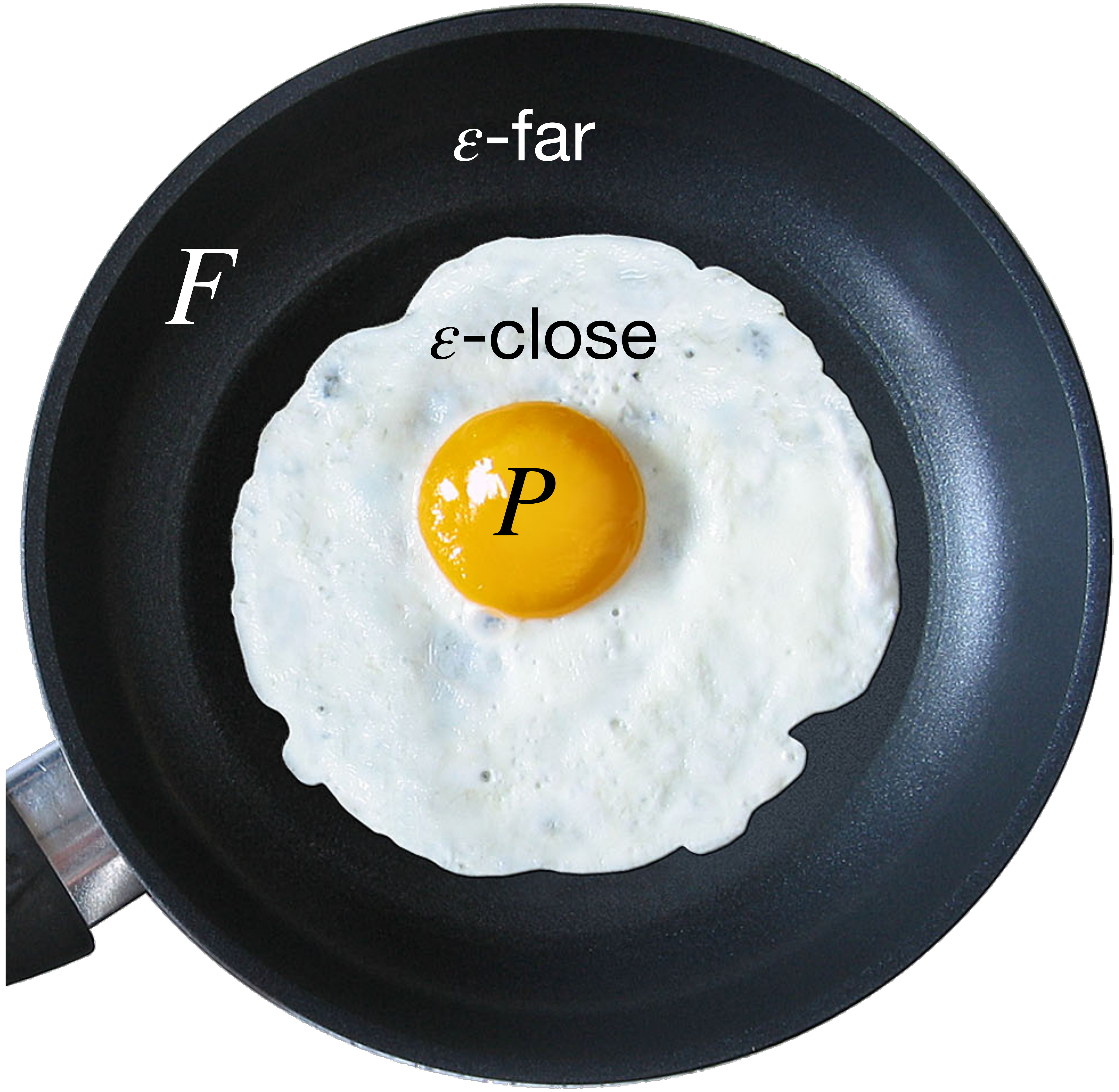


# PAC Learning [Valiant 84]

- Universe of functions $F: X \to R$
- Property $P \subset F$
- Distance:

$$d(f,g) = \Pr_{x \in X}[f(x) \neq g(x)]$$

$$d(f,P) = \min_{g \in P} d(f,g)$$




**PAC Learner**

Given oracle access to $f \in P$ and $\varepsilon > 0$:

- Output $h \in F$ such that $\mathbb{P}_h[d(f,h) \leq \varepsilon] \geq 2/3$

# When is testing easier than learning?

$q$ sample **learning** algorithm for $P$ $\Longrightarrow$ [GGR98] $\approx q$ sample **testing** algorithm for $P$

**Definition**

A **sample-based** algorithm is one that is only allowed to see the value of $f$ at uniform random points

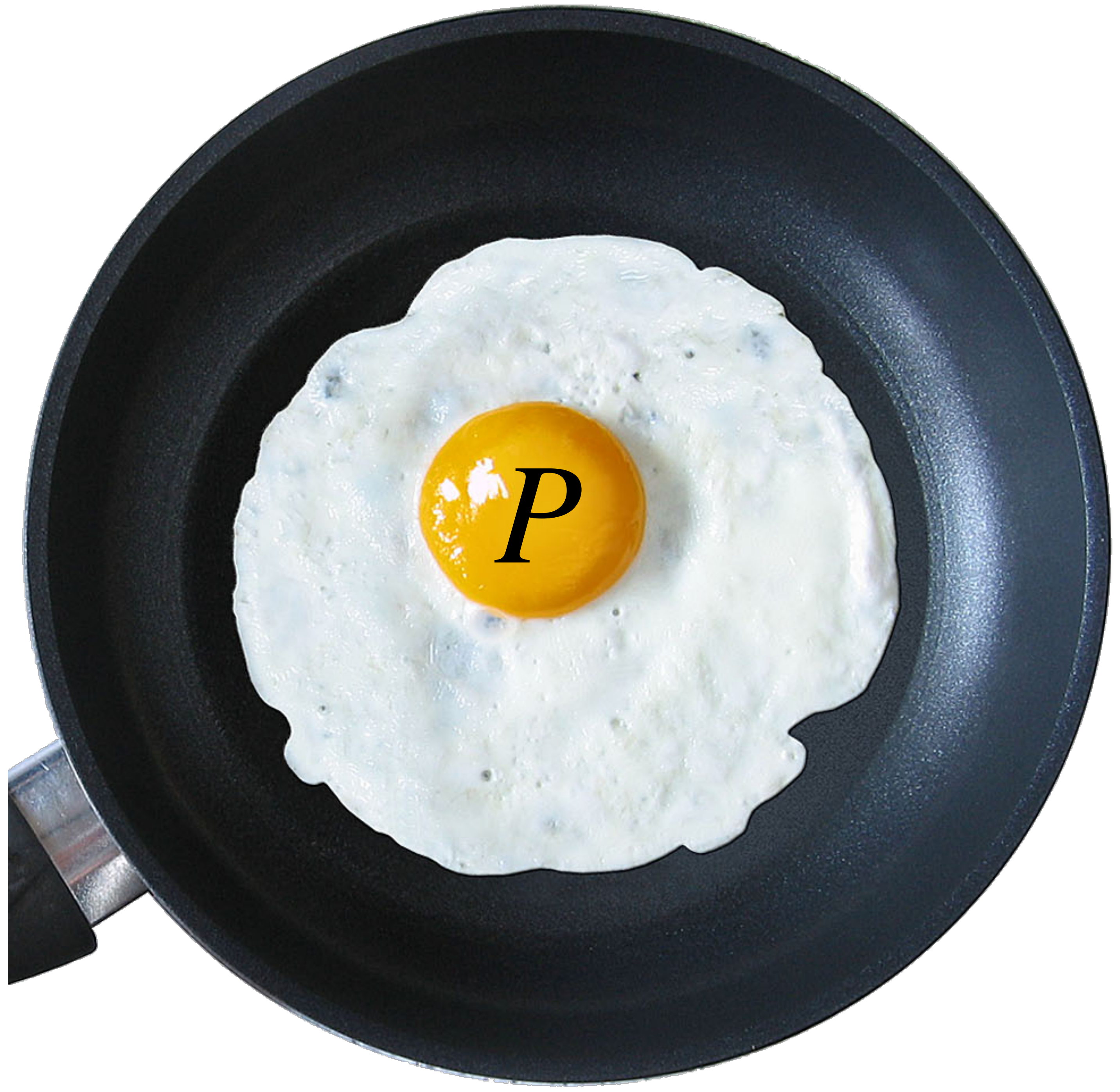
See $(x, f(x))$ where $x \sim \text{unif}(X)$

The standard access model for learning

**Question**

When does testing require fewer samples than learning?

[GGR98], [KR00], [BBBY12], [GR16], [BFH21]



# Monotone Functions

$f\colon \{0,1\}^d \to \{0,1\}$ **monotone** if $f(x) \le f(y)$ whenever $x \prec y$

Partial order: $x \preceq y$ iff $x_i \le y_i, \forall i \in [d]$

**Learning with samples:** $\exp\left(\widetilde{O}(d^{1/2}/\varepsilon)\right)$ samples [Bshouty-Tamon 96]

**Testing with queries:** $\widetilde{O}(d^{1/2}/\varepsilon^2)$ queries [Khot-Minzer-Safra 15]

Goldreich-Goldwasser-Lehman-Ron-Samorodnitsky 00, Fischer-Lehman-Newman-Raskhodnikova-Rubinfeld 02, Chakrabarty-Seshadhri 14, Chen-Servedio-Tan 15, Khot-Minzer-Safra 15, Chen-Waingarten-Xie 17

**What about testing with samples?**

[BT 96] + [GGR98]: $\exp\left(\widetilde{O}(d^{1/2}/\varepsilon)\right)$

$\Omega(\sqrt{\exp(d)/\varepsilon})$ when $\varepsilon \le d^{-3/2}$ [GGLRS 00]

**We prove:**

$\exp\left(\Omega(d^{1/2}/\varepsilon)\right)$ for all

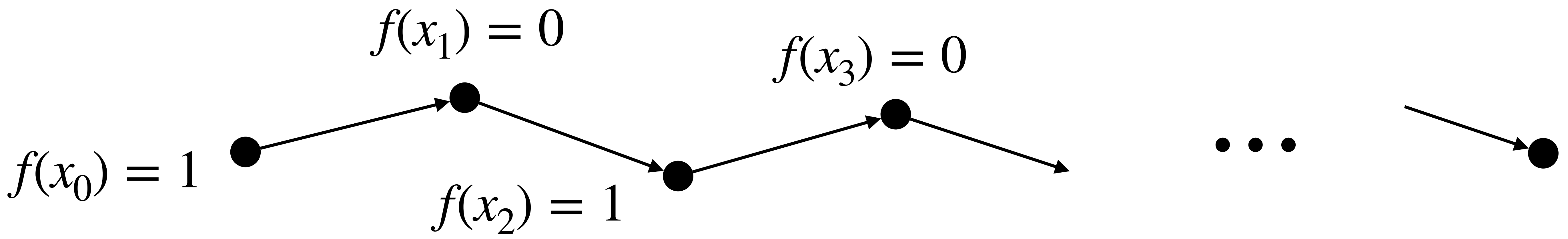
$d^{-1/2} \le \varepsilon \le c$ **[B 24]**

What about $\varepsilon \gg d^{-3/2}$?

In particular, $\varepsilon = \Omega(1)$?

# $(k)$-Monotone Functions

$f\colon \{0,1\}^d \to \{0,1\}$ is $k$-**monotone** if there does not exist $x_0 \prec x_1 \prec \ldots \prec x_k$ such that




**Learning with samples:** $\exp\left(\widetilde{O}(kd^{1/2}/\varepsilon)\right)$ [Blais-Cannone-Oliveira-Servedio-Tan 15]

**Testing with queries** (1-sided error, $k \geq 2$)**:** $\exp(\widetilde{\Theta}(d^{1/2}))$ [Grigorescu-Kumar-Wimmer 19], [Canonne-Grigorescu-Guo-Kumar-Wimmer 19]

**What about testing with samples?**

$$f\colon \{0,1\}^d \to [r]$$

**We prove:** $\exp\left(\Omega(rkd^{1/2}/\varepsilon)\right)$ for all $d^{-1/2} \leq \varepsilon \leq c$, **[B 24]**

$\exp\left(\widetilde{O}(rkd^{1/2}/\varepsilon)\right)$ for learning **[B 24]**

# What can you hope to do with samples?

**Query-**based algorithms look for violations:

$x \prec y$ where $f(x) > f(y)$

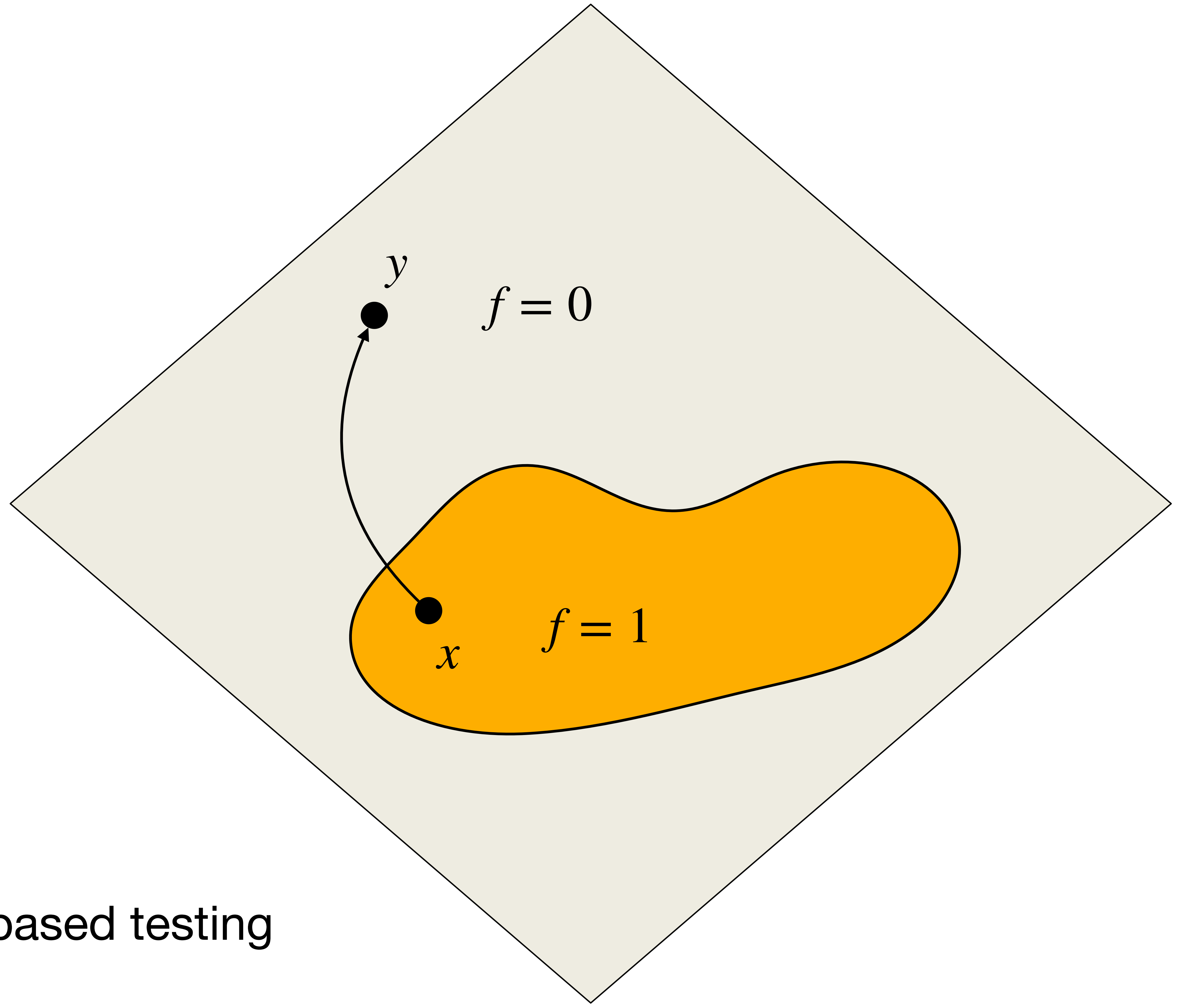
… if $x, y$ are **samples**, then

$$\Pr[x \prec y] = \prod_{i=1}^{d} \Pr[x_i \leq y_i] \; = (3/4)^d$$

$\Longrightarrow$ For $s$ samples, need $s^2 \geq (4/3)^d$ to see **even one** comparable pair of points. I.e. $s = \exp(\Omega(d))$

$\Longrightarrow$ $\exp(\Omega(d))$ lower bound for 1-sided error sample-based testing

**Our goal: 2-sided error lower bound**



# Lower bound for Boolean function monotonicity testing with samples

**Theorem [B 24]**

Testing monotonicity of $f\colon \{0,1\}^d \to \{0,1\}$ requires $\exp\left(\Omega(d^{1/2})\right)$ samples

We prove this for $\varepsilon \leq c$ for a sufficiently small constant $c \in (0,1)$

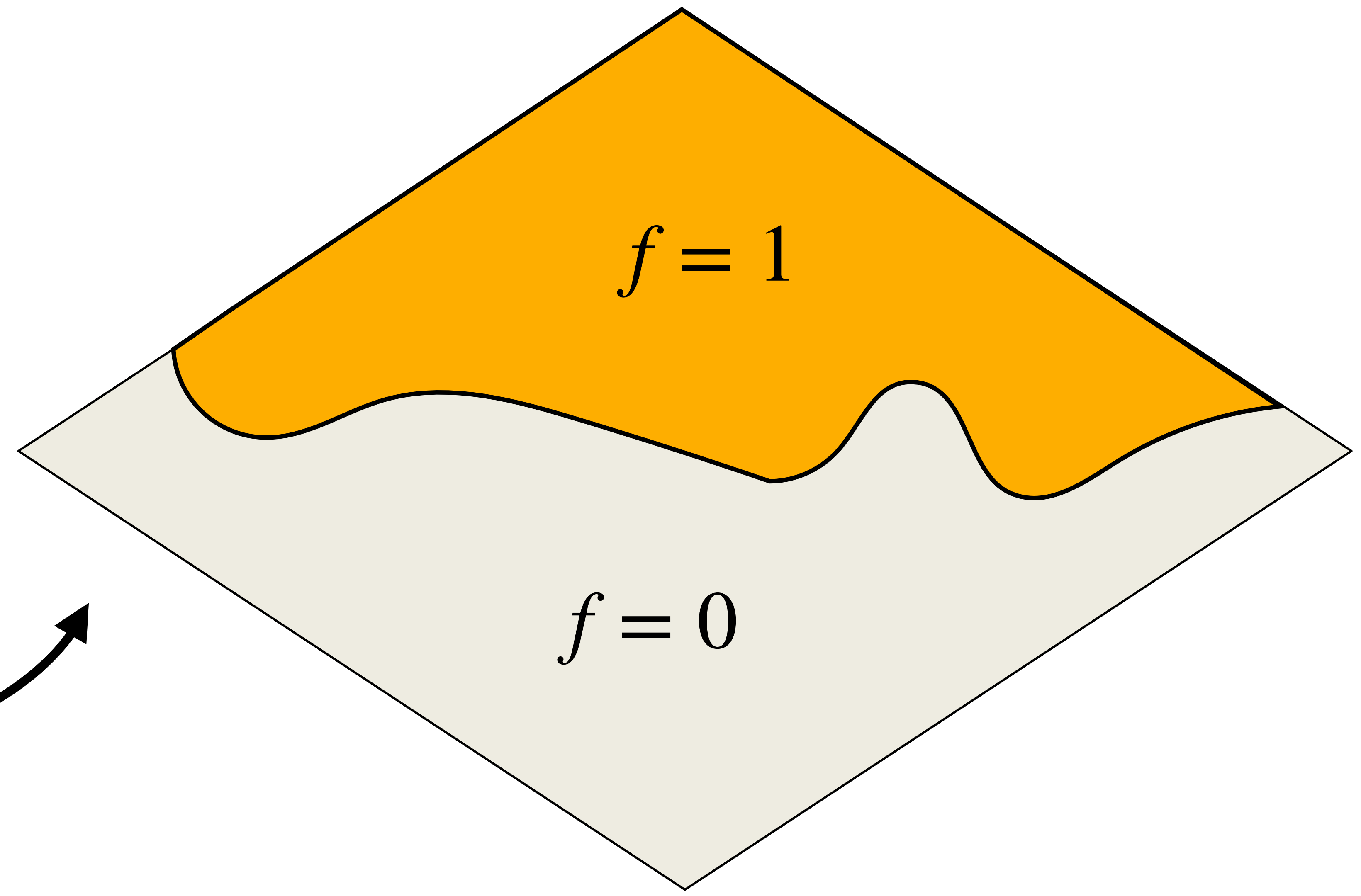
# High level view: $\mathscr{D}_{yes}$ and $\mathscr{D}_{no}$

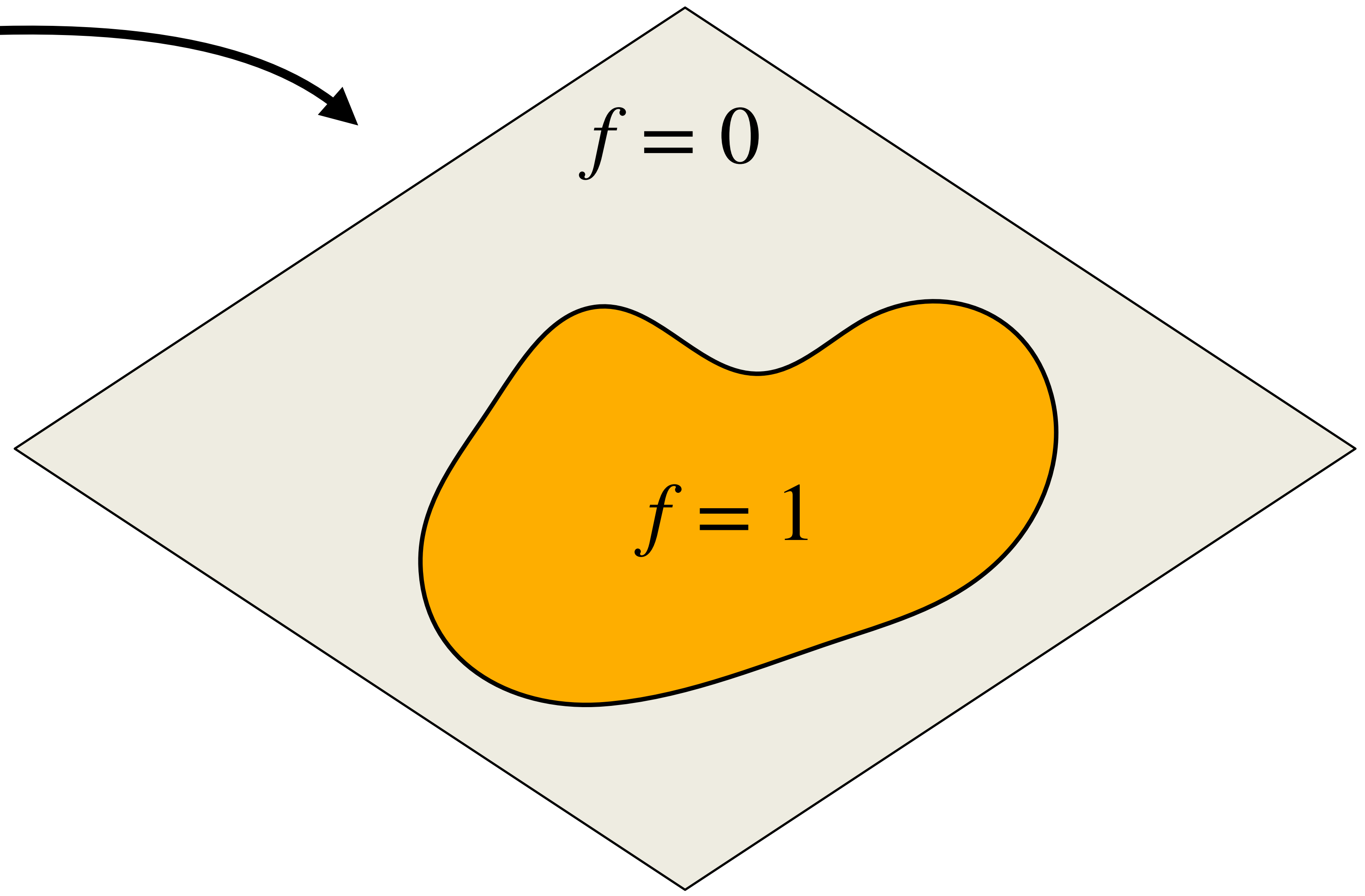
**Theorem [B 24]**

Testing monotonicity of $f\colon \{0,1\}^d \to \{0,1\}$ requires $\exp\left(\Omega(d^{1/2})\right)$ samples

Need to construct:

- $\mathscr{D}_{yes}$: supported over monotone $f$
- $\mathscr{D}_{no}$: outputting $f$ that is $\Omega(1)$-far with prob. $\Omega(1)$




Such that…

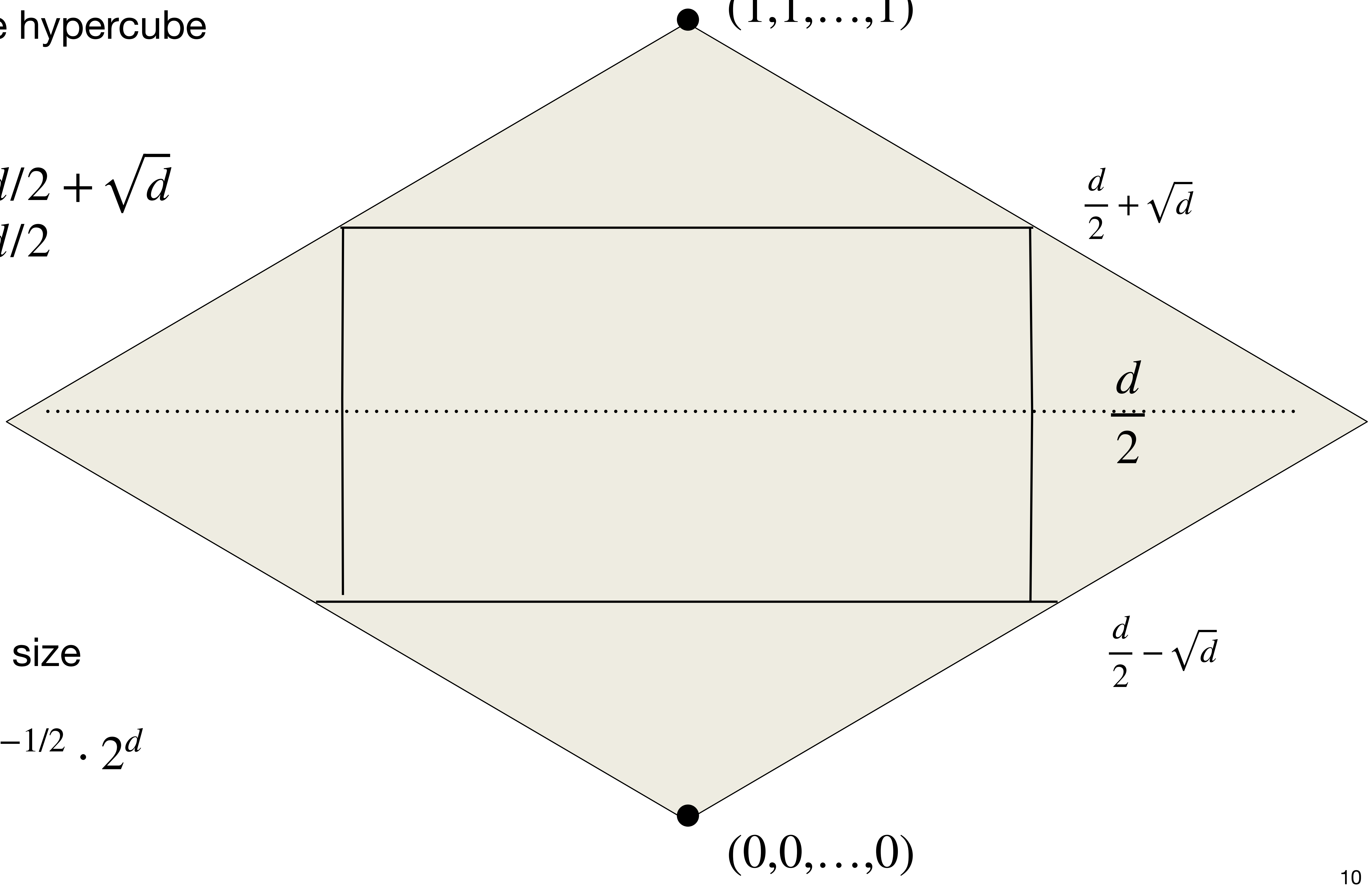
- A uniform random set $S$ of $\exp(o(d^{1/2}))$ points cannot tell if $f$ came from $\mathscr{D}_{yes}$ or $\mathscr{D}_{no}$

# First, some simplifications

1) Focus on upper middle layers of the hypercube

Functions we define will satisfy

- $f(x) = 1$ whenever $|x| > d/2 + \sqrt{d}$
- $f(x) = 0$ whenever $|x| < d/2$




2) Imagine middle layers are the same size

$$\ell \in [d^{1/2}] \implies \binom{d}{d/2 + \ell} \approx d^{-1/2} \cdot 2^d$$

# Talagrand's random DNF

[Talagrand 96]

Used by [Belovs-Blais 16, Chen-Waingarten-Xie 17]
[Chen-De-Li-Nadimpalli-Servedio 24]
[Black-Blais-Harms 24]

$N$ terms of width $w = o(d)$

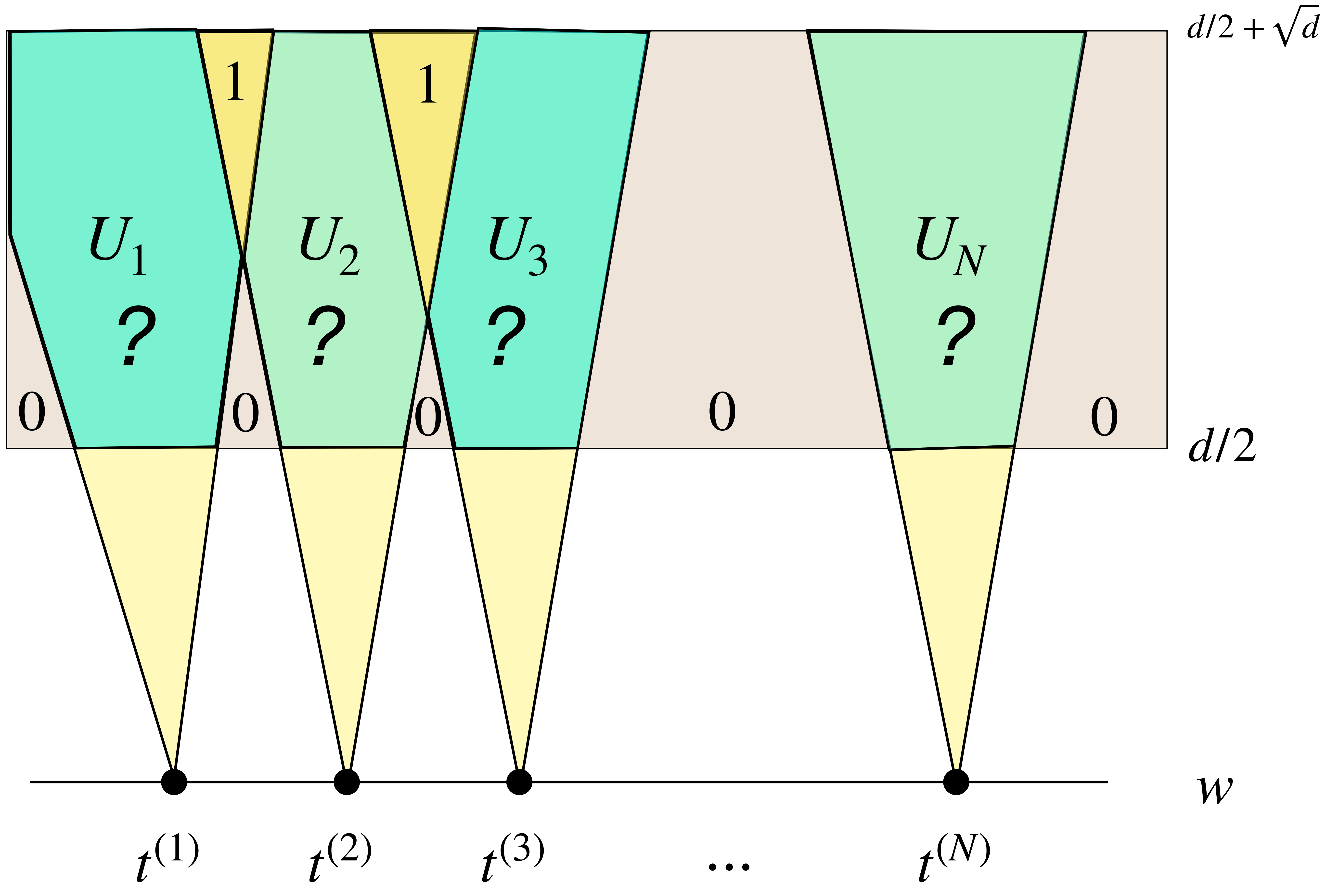
Draw $t^{(1)}, \ldots, t^{(N)} \in \{0,1\}^d$ with $|t^{(i)}| = w$

$x$ satisfies $t^{(i)}$ if $x \succeq t^{(i)}$

$U_i =$ all $x$ that satisfy $t^{(i)}$ **uniquely**

**Observation:**
points in $U_i$ and $U_j$ are incomparable

$\Longrightarrow$ embedding an arbitrary monotone function in each $U_i$ results in a monotone function



# $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$: what to put in each $U_i$?

- $\mathcal{D}_{yes}$: $f(U_i)$ is a **random constant** $\forall i$
- $\mathcal{D}_{no}$: $f(U_i)$ is **random** for every $\forall i$

**Observation 1**

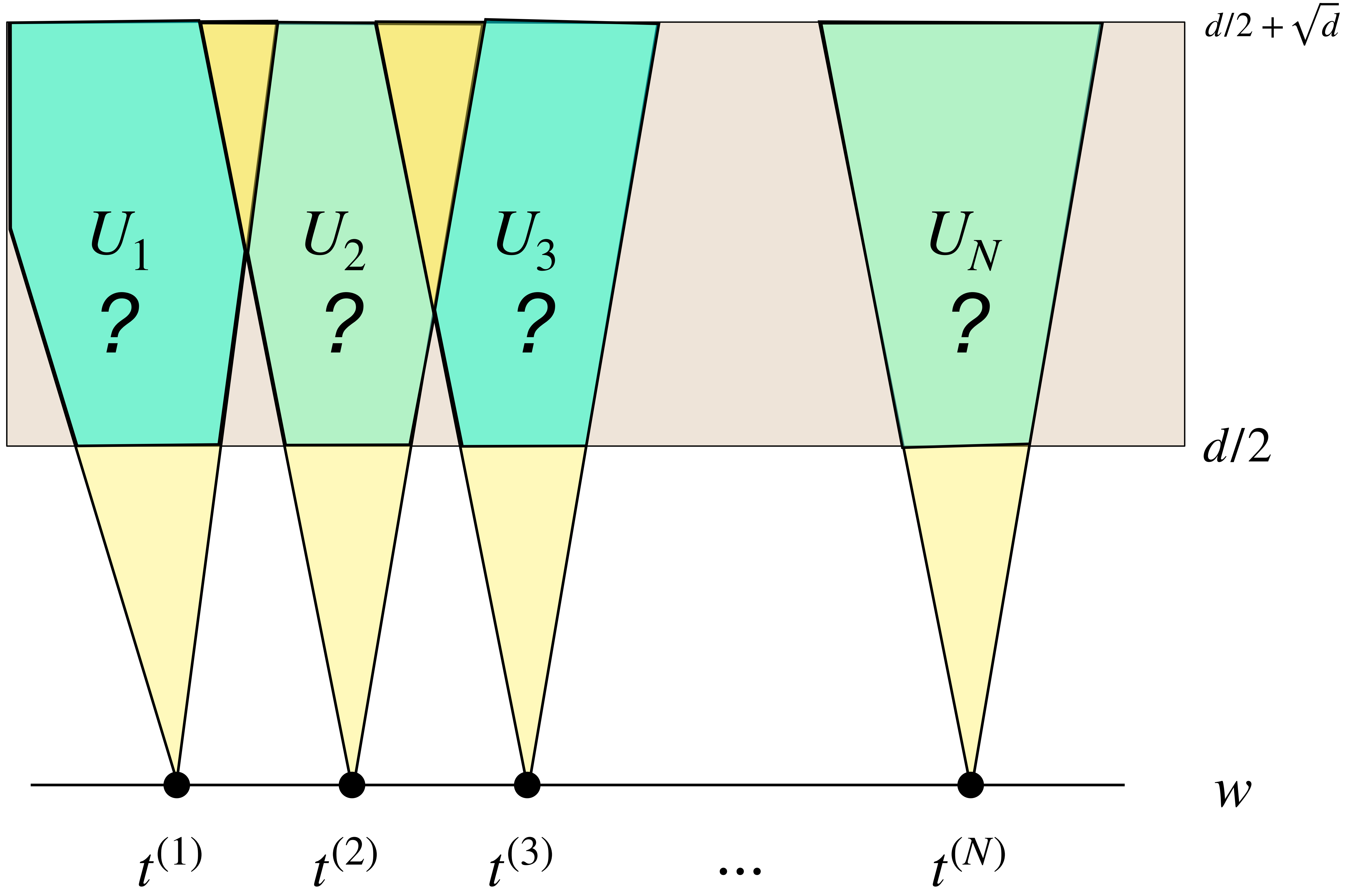
$S \subseteq \{0,1\}^d$ distinguishes $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ only if $|S \cap U_i| > 1$ for some $i$

$\Longrightarrow$ $\Omega(\sqrt{N})$ to distinguish by birthday paradox

**Observation 2**

if $|U_1 \cup \cdots \cup U_N| = \Omega(2^d)$, then $f$ will be $\Omega(1)$-far from monotone who

$\Longrightarrow$ $\Omega(\sqrt{N})$ lower bound



# For what $N$ can we get $f \sim \mathscr{D}_{no}$ to be $\Omega(1)$-far?

Terms $\boldsymbol{t} = (t^{(1)}, \ldots, t^{(N)})$ of width $|t^{(i)}| = w$

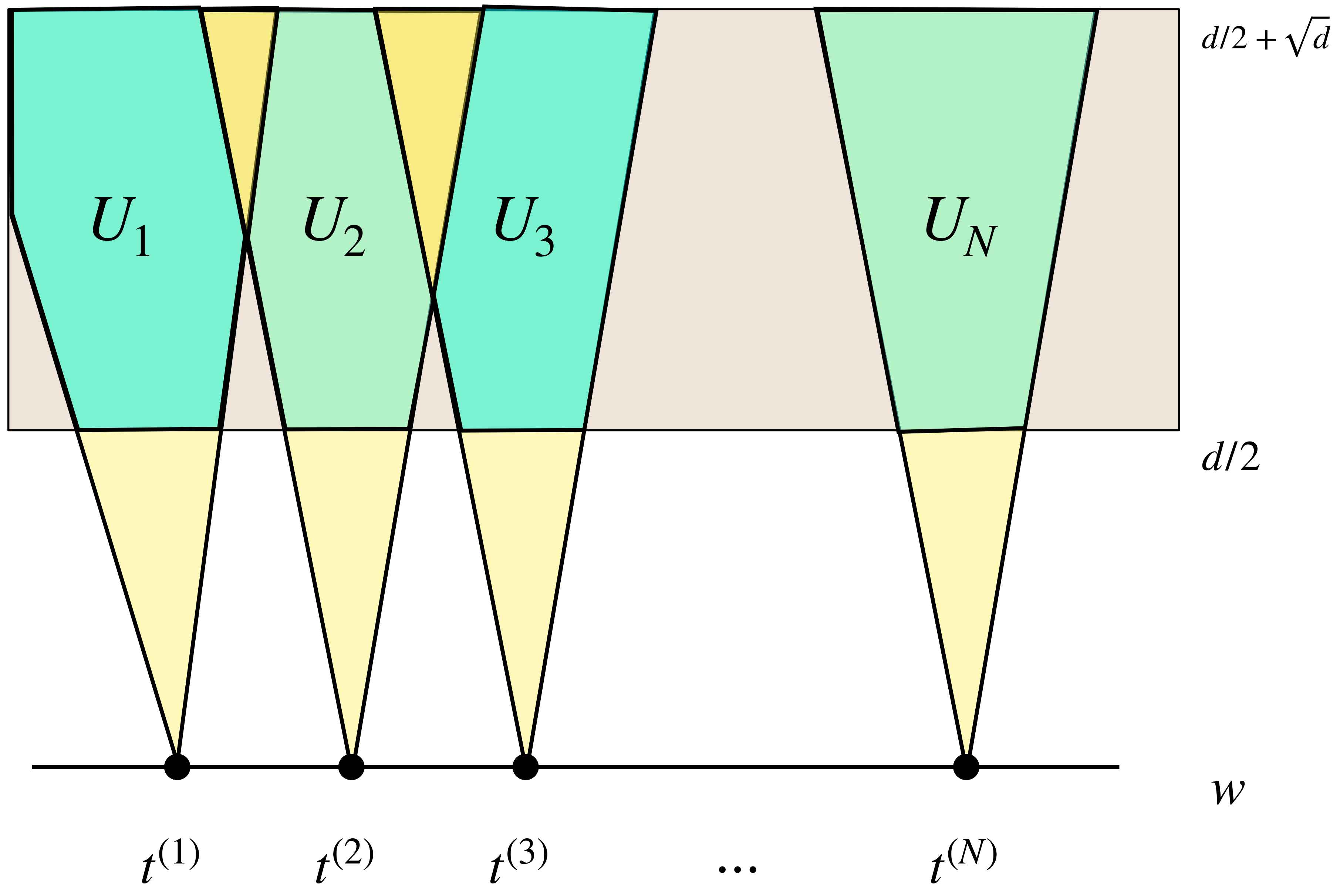
$U_i =$ all $x$ that satisfy $t^{(i)}$ **uniquely**

$U = U_1 \cup \cdots \cup U_N$ … can we get $|U| = \Omega(2^d)$?

- We need $\Pr[x \in U] = \Omega(1)$ for all $x$ in upper middle

If $|x| = d/2 \ldots$

$$\mathbb{E}_{\boldsymbol{t}}[\#i : t^{(i)} \preceq x] = N \cdot (|x|/d)^w$$
$$= N \cdot 2^{-w}$$
$$\approx 1 \quad \text{if } N \approx 2^w$$



# For what $N$ can we get $f \sim \mathcal{D}_{no}$ to be $\Omega(1)$-far?

Terms $\boldsymbol{t} = (t^{(1)}, \ldots, t^{(N)})$ of width $|t^{(i)}| = w$

$U_i$ = all $x$ that satisfy $t^{(i)}$ **uniquely**

$U = U_1 \cup \cdots \cup U_N$ ... can we get $|U| = \Omega(2^d)$?

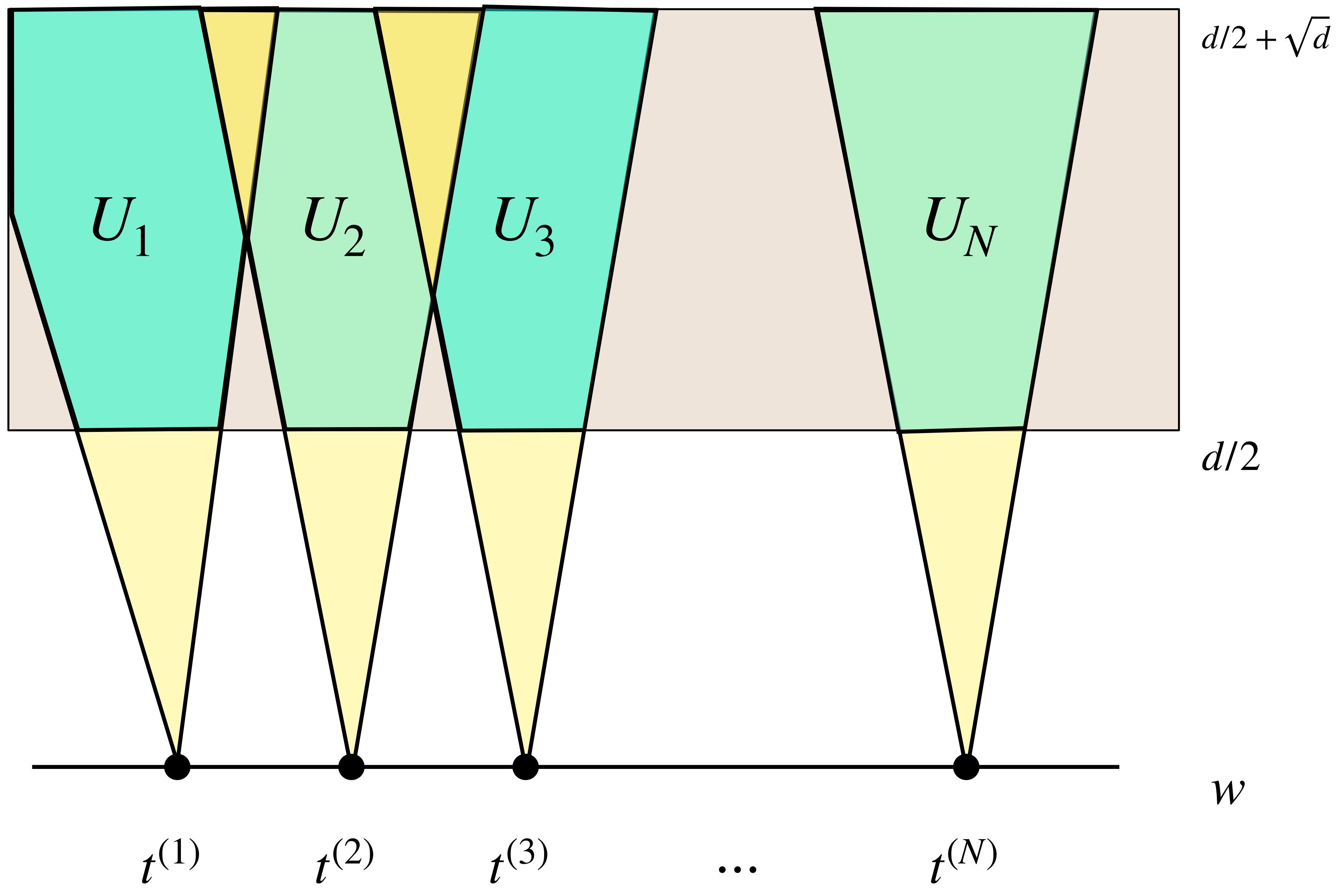
$N = 2^w$

If $|x| = d/2 + \sqrt{d}$...

$$\mathbb{E}_{\boldsymbol{t}}[\#i : t^{(i)} \preceq x]$$
$$= N \cdot (|x|/d)^w$$
$$= N \cdot 2^{-w}(1 + 2/\sqrt{d})^w$$
$$= (1 + 2/\sqrt{d})^w \approx 1$$

... if $w \approx \sqrt{d}$

Construction works when $N \approx 2^{\sqrt{d}}$

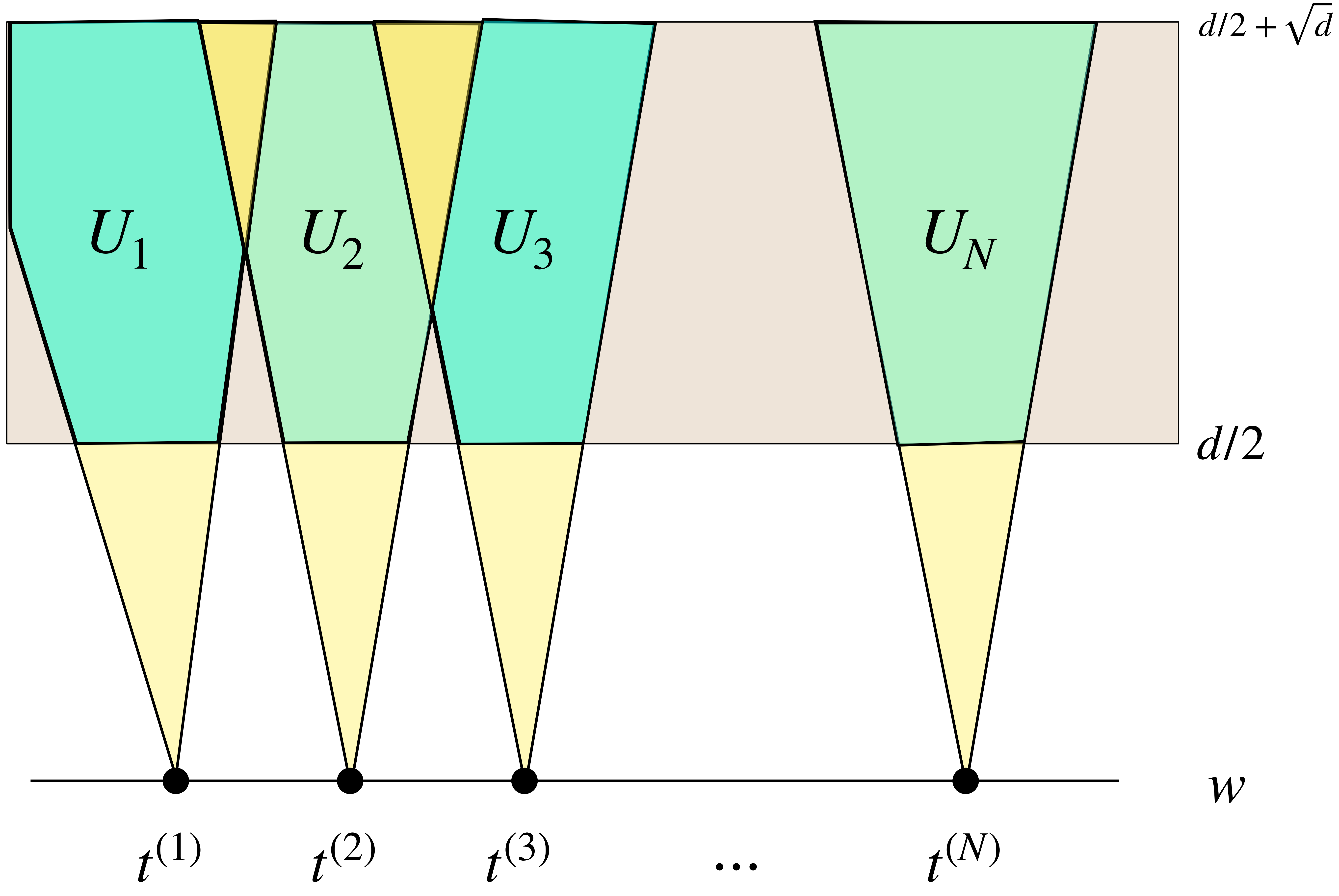



$\implies 2^{\Omega(\sqrt{d})}$ lower bound ☐

# On the parameters $N$ and $w$

- We need $|U| = \Omega(2^d)$
- Number of terms that works is $N = 2^w$
- We need $\Pr[x \in U] = \Omega(1)$ for all $x$ in a window of $\sqrt{d}$ possible Hamming weights

$\implies$ This forces $w \leq \sqrt{d}$



# Generalizing to $k$-monotonicity

- Decompose upper middle layers into $k$ blocks $B_1, \ldots, B_k$
- In each $B_i$ put a DNF

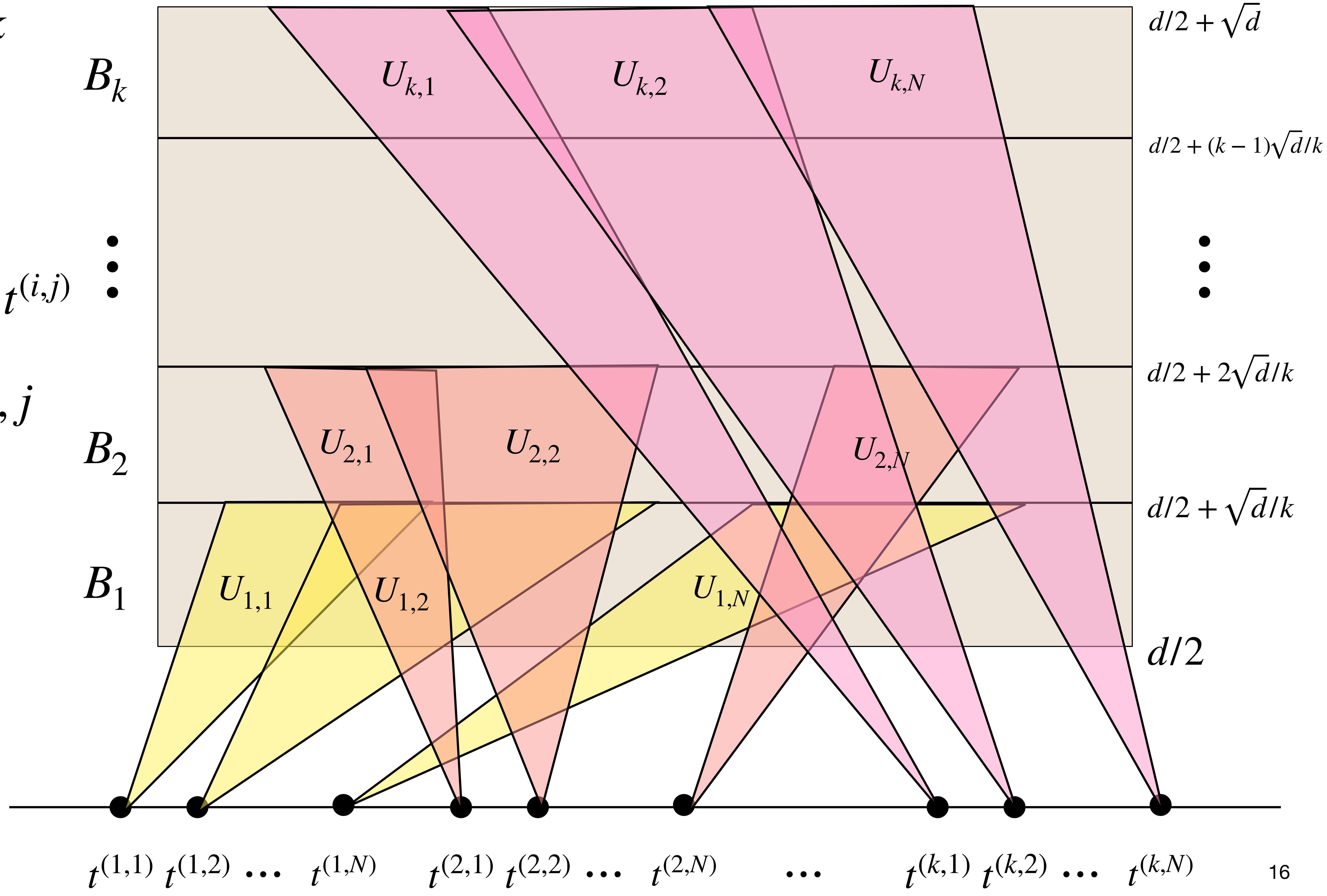
  Terms $\boldsymbol{t}^{(i)} = (t^{(i,1)}, \ldots, t^{(i,N)})$

  $U_{i,j} =$ all $x \in B_i$ satisfied uniquely by $t^{(i,j)}$

- $\mathcal{D}_{yes}$: $f(U_{i,j})$ is a **random constant** $\forall i, j$
- $\mathcal{D}_{no}$: $f(U_{i,j})$ is **random** $\forall i, j$
- Can set $w \approx k\sqrt{d}$ and $N \approx 2^{k\sqrt{d}}$

$\implies$ $\exp(\Omega(k\sqrt{d}))$ lower bound

Similar trick gives $\exp(\Omega(rk\sqrt{d}))$ for functions with range $[r]$



# Summary

**Our lower bound:**

$(k)$-Monotonicity testing of $f\colon \{0,1\}^d \to [r]$ requires $\exp(\Omega(rk\sqrt{d}/\varepsilon))$ samples

- **Testing is not easier than learning** for any $r, k, \varepsilon$
    - (Up to $\log d$ factor in the exponent)
- Was not known even for $r = 2, k = 1$

**Upper bound for learning over product spaces:**

Can learn $(k)$-monotone $f\colon \mathbb{R}^d \to [r]$ under product distributions with $\exp(\widetilde{O}(rk\sqrt{d}/\varepsilon))$ samples

- Improves on $\exp(\widetilde{O}(k\sqrt{d}/\varepsilon^2))$ for $r = 2$ by [Harms-Yoshida 22]