Exponential improvements to the average-case hardness of random circuits

Shaun Datta
Stanford University

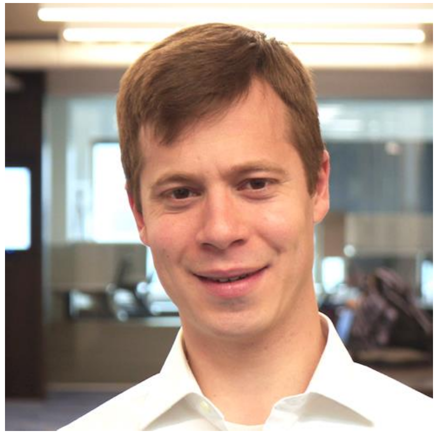
FOCS 2025

To be posted as arXiv:2411.04566 v2 soon!

Exponential improvements to the average-case hardness of random circuits

Shaun Datta
Stanford University

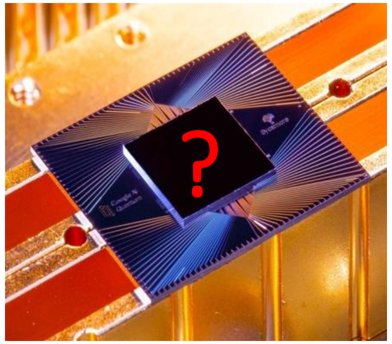
Joint work with Adam Bouland, Bill Fefferman, Felipe Hernández

Sampling from random circuits—why should I care?

To understand the power of near-term quantum experiments

Many random sampling experiments: how hard are they to simulate?



Sampling from random circuits—why should I care?

To understand the power of near-term quantum experiments

Many random sampling experiments: how hard are they to simulate?

To separate classical and quantum computation. Is $BPP \neq BQP$?

We have excellent oracular (blackbox) evidence. What about whitebox?

Dream 1: $BQP \not\subset BPP$ (way beyond current techniques)

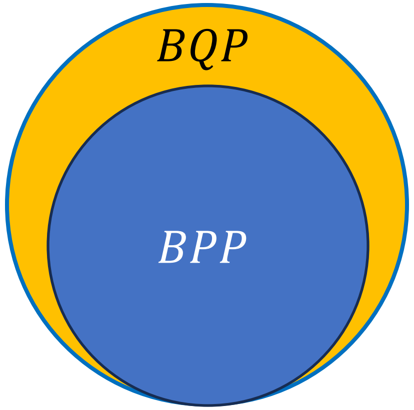
Dream 2: $BQP \subseteq BPP \Rightarrow PH$ collapses (still seems difficult to show)

Dream 3: $sampBQP \subseteq sampBPP \Rightarrow PH$ collapses

*This can be proven!** *[e.g., TD04, BJS10, AA10]*

Caveat: *result is brittle—pertains to worst-case, exact sampling*

How far can we push these separations?



We have yet to realize this dream!

Sampling from random circuits—why should I care?

To understand the power of near-term quantum experiments

Many random sampling experiments: how hard are they to simulate?

To separate classical and quantum computation. Is $BPP \neq BQP$?

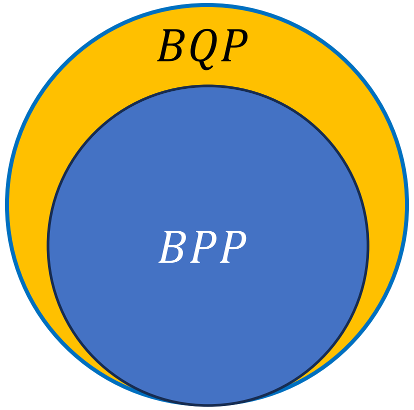
We have excellent oracular (blackbox) evidence. What about whitebox?

Dream 1: $BQP \not\subset BPP$ (way beyond current techniques)

Dream 2: $BQP \subseteq BPP \Rightarrow PH$ collapses (still seems difficult to show)

Dream 3: $sampBQP \subseteq sampBPP \Rightarrow PH$ collapses

*This can be proven!** *[e.g., TD04, BJS10, AA10]*



**Caveat:* *result is brittle—pertains to worst-case, exact sampling*

How far can we push these separations?

Lastly, cryptography, e.g. from [Khurana Tomer '24b]

Sampling from random circuits

Computational task (Random Circuit Sampling, BosonSampling, IQP, …):

1. Initialize a fiducial starting state
2. Evolve by a random circuit
3. Measure to generate a sample

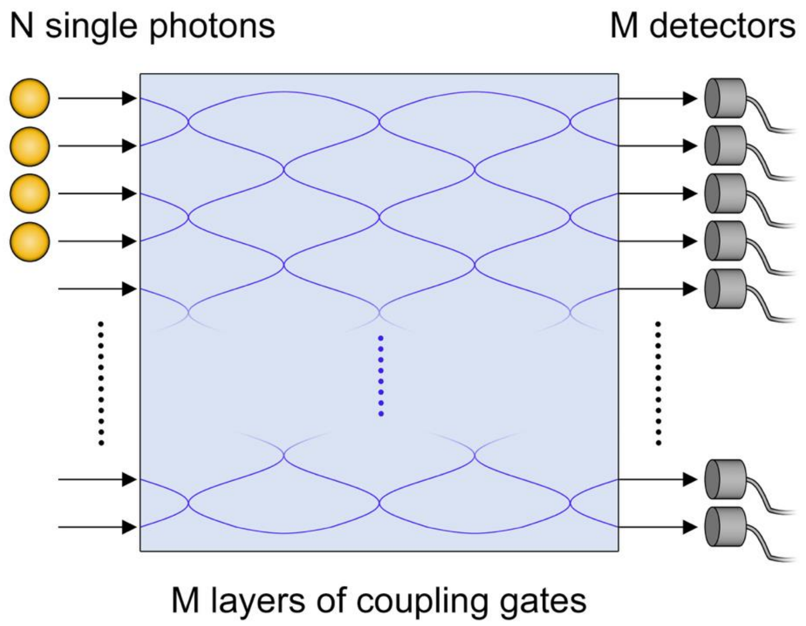
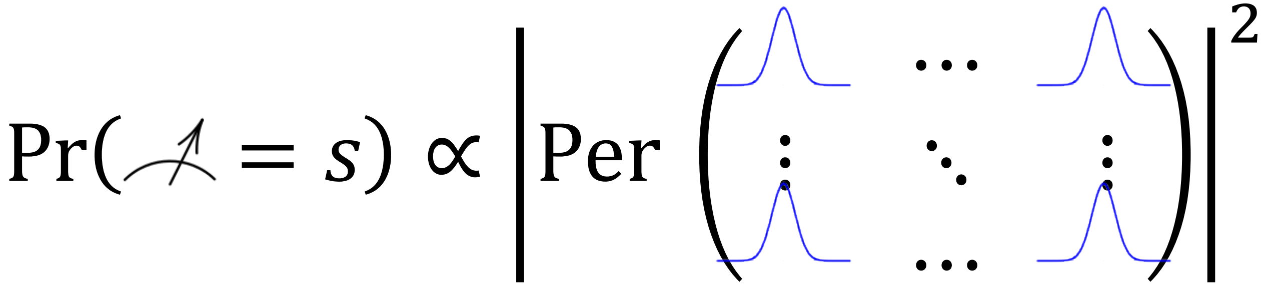
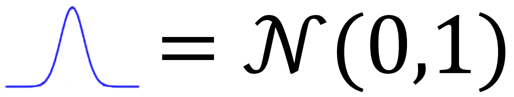


Fig. from GRS '19

For BosonSampling,





From sampling to computing

Prior work [AA10, BFNV19] established that classical computers cannot sample from random circuits…

…if it is $\#P$-hard to **estimate** an output probability to within $\pm\delta$:

Random Circuit Sampling (RCS): $\delta = 2^{-n-O(\log n)}$

BosonSampling: $\delta = \exp(-n\log n - n - O(\log n))$

Permanent-of-Gaussians Conjecture (PGC) [AA10]

From sampling to computing

Prior work [AA10, BFNV19] established that classical computers cannot sample from random circuits…

…if it is $\#P$-hard to **estimate** an output probability to within $\pm\delta$:

Random Circuit Sampling (RCS): $\delta = 2^{-n-O(\log n)}$

BosonSampling: $\delta = \exp(-n\log n - n - O(\log n))$

Central open problem: prove one of these conjectures for any random sampling task!

What's the status of proving these conjectures?

We want to show it is hard to estimate output probabilities to $\pm\,\delta$, but so far we have only proven it is hard to $\pm\delta' \ll \delta$

"Robustness" "Robustness gap"

For example, for BosonSampling:

AA10	$e^{-O(n^4)}$
BFLL21	$e^{-6n\log n - O(n)}$
Kro22	$e^{-4n\log n - O(n)}$
[This work]	$e^{-n\log n - n - O(n^{\varepsilon})}\ \forall\varepsilon > 0$
Goal: PGC	$e^{-n\log n - n - O(\log n)}$

Why has progress been so difficult?

- Classical algorithms have solved related tasks [e.g., EM18, JJL21]
- Prior proofs are limited by barriers:
  - Depth barrier for RCS (Napp, et al. ‘22)
  - Jerrum-Sinclair-Vigoda barrier for BosonSampling
  - Convexity barrier (AA10), Noise (BFLL21), “Born rule” barrier (Kro22), …

In this work, we overcome all the known proof barriers.

Second result: hardness of sampling

[This work] There is no classical sampler that succeeds for

$\geq 1 - 2^{-\tilde{O}\left(\sqrt[3]{N}\right)}$ fraction of instances of size N

Trivial: $\geq 1 - 2^{-\tilde{O}(N)}$

But we want to show: $\geq 1 - 1/\mathrm{poly}(N)$

This is the first nontrivial hardness of average-case sampling result!

Start of the proof sketch

The standard worst-to-average-case reduction

[Lipton91, AA10]

$$\mathrm{Per}\big(R(t)\big) \coloneqq \mathrm{Per}((1-t)R + tW)$$

has three desirable properties:

- Polynomial in t of degree n (For RCS, degree $\approx$ number of gates m)
- $R(t) \approx R$ for small t
- $\mathrm{Per}\big(R(1)\big) = \mathrm{Per}(W)$

$$\mathrm{Per}\left((1-t)\; \boxed{R \sim \mathcal{N}(0,1)^{n\times n}} \;+\; t\; \boxed{W \in \{0,1\}^{n\times n}}\right)$$

The standard worst-to-average-case reduction

[Lipton91, AA10]

$\mathrm{Per}\big(R(t)\big) \coloneqq \mathrm{Per}((1-t)R + tW)$

has three desirable properties:

- Polynomial in t of degree n
- $R(t) \approx R$ for small t
- $\mathrm{Per}\big(R(1)\big) = \mathrm{Per}(W)$

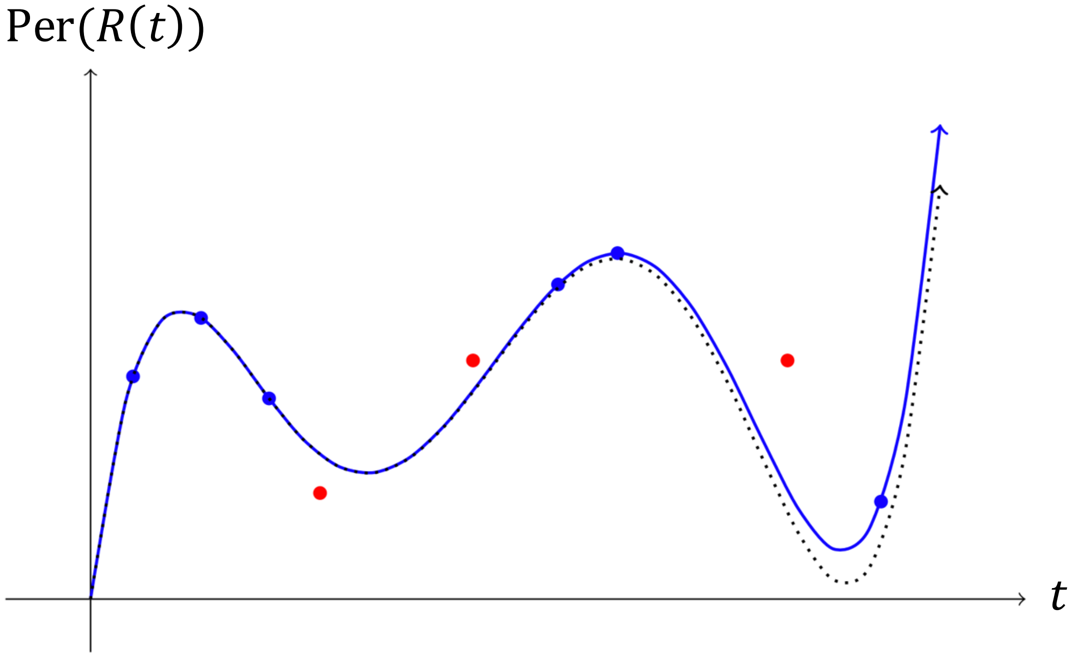


Fig. from Bouland Fefferman Nirkhe Vazirani '19

Key idea: polynomial extrapolation

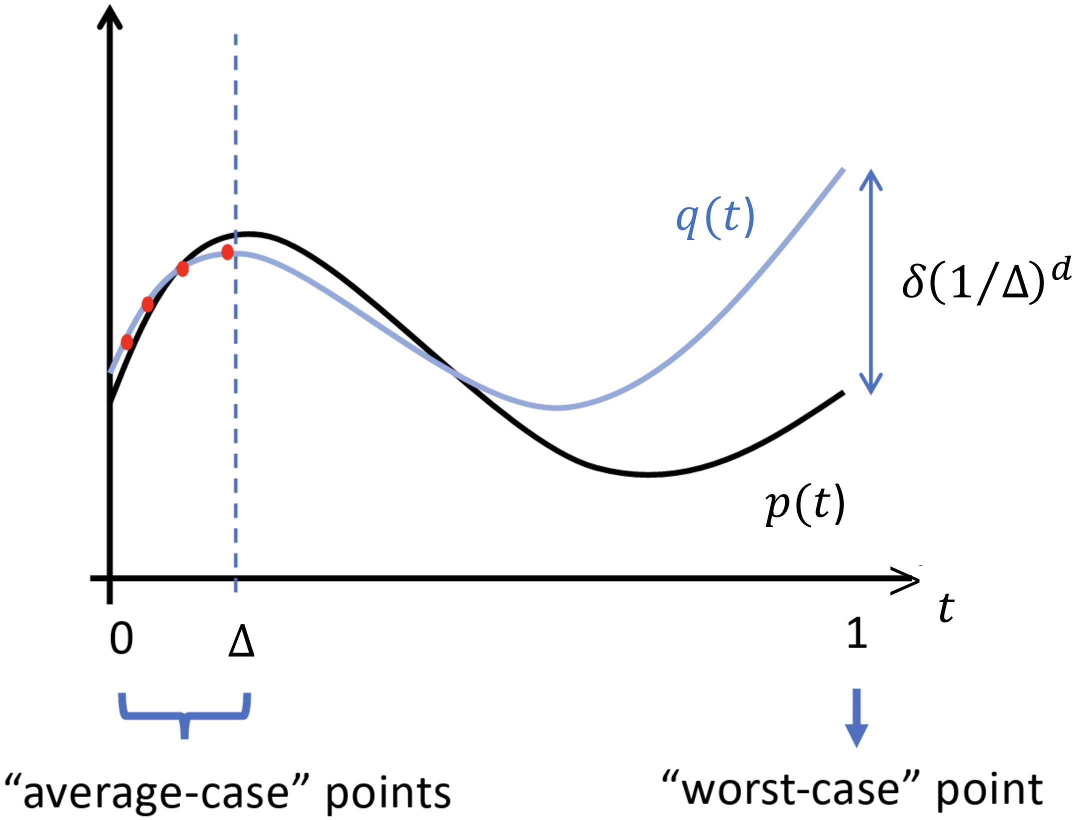
Infer $\mathrm{Per}(W)$ from noisy estimates to $\mathrm{Per}\big(R(t)\big)$ for small values of t

What controls robustness?

Polynomial extrapolation is ill-conditioned

Error blowup given by the Remez inequality: estimating degree d polynomial to error $\leq \delta$ on interval $[0, \Delta]$ incurs $\delta(1/\Delta)^d$ blowup

Moral: to improve robustness, need to decrease extrapolation distance $1/\Delta$ or decrease the polynomial degree d



New techniques to decrease $1/\Delta$ and d

- Dilution
- Coefficient extraction
- The square trick
- Magnification
- Rare events lemmas

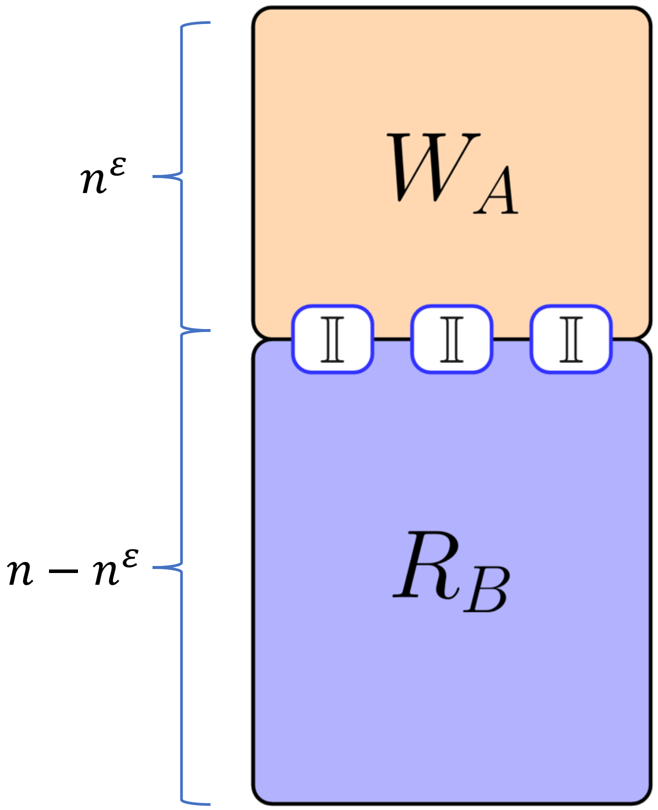
The focus of today's talk

Dilution: technique to decrease $1/\Delta$ and d

Prior work used a worst-case circuit on n qubits

Instead, consider the circuit W_n acting on n qubits:

- Where W_A is worst-case circuit on n^ε qubits, any constant $\varepsilon > 0$
- Where R_B is random but fixed circuit on $n - n^\varepsilon$ qubits



Let $p_y(C)$ be probability to measure y from circuit C

By construction output probability "factorizes"

$$p_{0^n}(W_n) = p_{0^{n^\varepsilon}}(W_A) \cdot p_{0^{n-n^\varepsilon}}(R_B)$$

Observation: $p_{0^{n^\varepsilon}}(W_A)$ is $\#P$-hard to estimate multiplicatively by padding

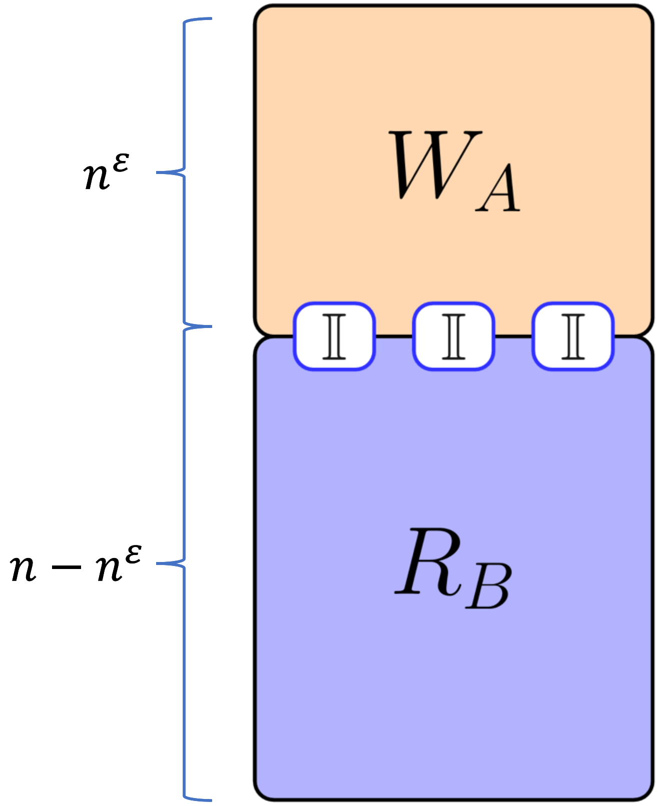
New worst-to-average-case reduction by dilution

Goal: estimate $p_0(W_A) = \frac{p_0(W_n)}{p_0(R_B)}$ ← Need multiplicative estimates

Denominator: can estimate R_B by assumption ✓

Numerator: W_n is a worst-case circuit

- Implement previous worst-to-average-case reduction!
- But only extrapolate over gates in W_A
  - i.e., correlated circuits $C(t_i)$ all share R_B
- Degree of $p(t)$ is $\text{supp}(W_A) = O(n^{\varepsilon})$
- So blow-up is $\frac{1}{(n^{\varepsilon})^{n^{\varepsilon}}} = \frac{1}{2^{n^{\varepsilon} \log(n^{\varepsilon})}}$



New worst-to-average-case reduction by dilution

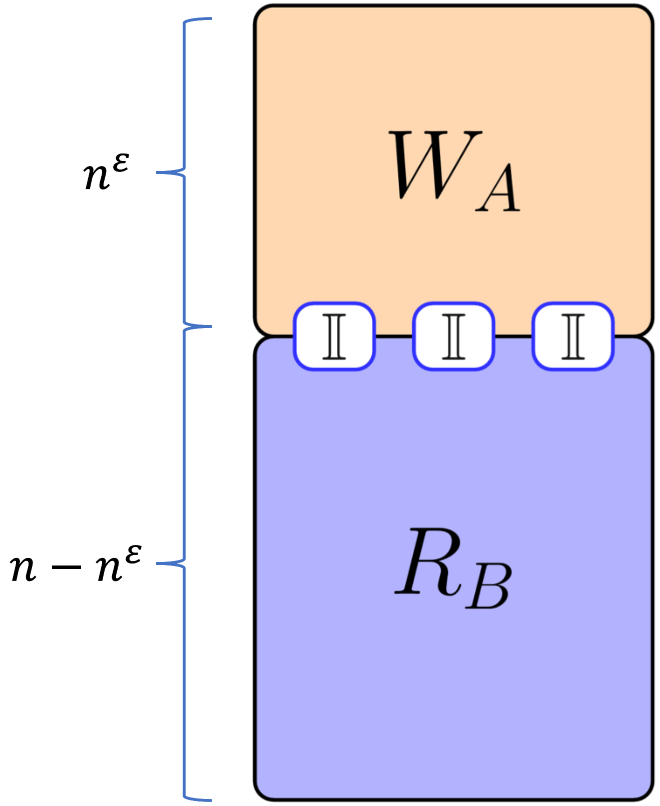
Goal: estimate $p_0(W_A) = \frac{p_0(W_n)}{p_0(R_B)}$ ← Need multiplicative estimates

Denominator: can estimate R_B by assumption ✓

Numerator: W_n is a worst-case circuit ✓

- Implement previous worst-to-average-case reduction!
- But only extrapolate over gates in W_A
  - i.e., correlated circuits $C(t_i)$ all share R_B
- Degree of $p(t)$ is $\mathrm{supp}(W_A) = O(n^\varepsilon)$
- We get robustness $\delta = 2^{-n-n^\varepsilon \log n^\varepsilon}$

2^{-n} would suffice to show no classical sampler



Feature: our hardness argument circumvents the depth barrier

Random circuits have a phase transition in depth from easy to hard

Reason: **entanglement**

Efficient classical algorithms can exploit shallow depth, e.g., Napp, et al. ’22

By contrast, prior hardness arguments were agnostic to depth

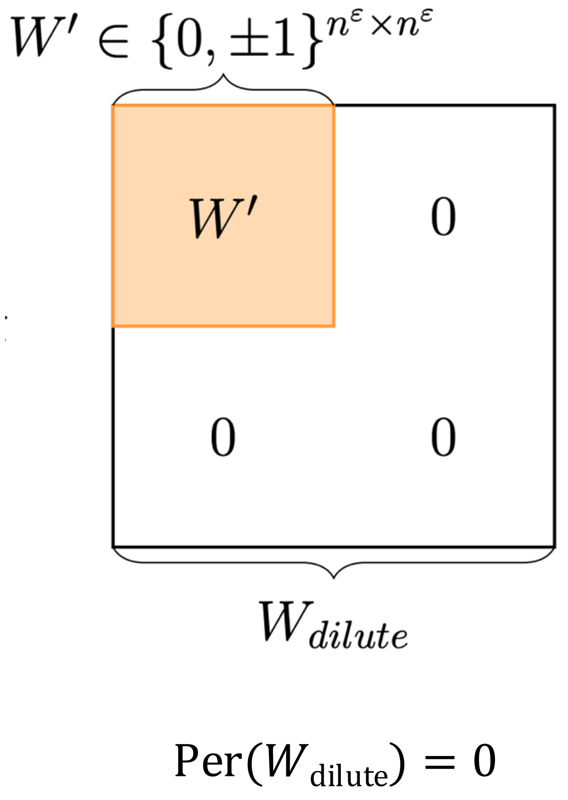
Our argument requires anticoncentration, which requires log depth [Dalzell, et al. ‘22 & Deshpande, et al. ‘22]

Dilution does not trivially extend to BosonSampling!

Simply shrink worst-case instance W' to have size $n^\varepsilon \times n^\varepsilon \ \forall \varepsilon > 0$

$\mathrm{Per}(W')$ is still $\#P$-hard (padding)

Idea: extrapolate $\mathrm{Per}[(1-t)R + t\, W_{\mathrm{dilute}}]$ to $t = 1$



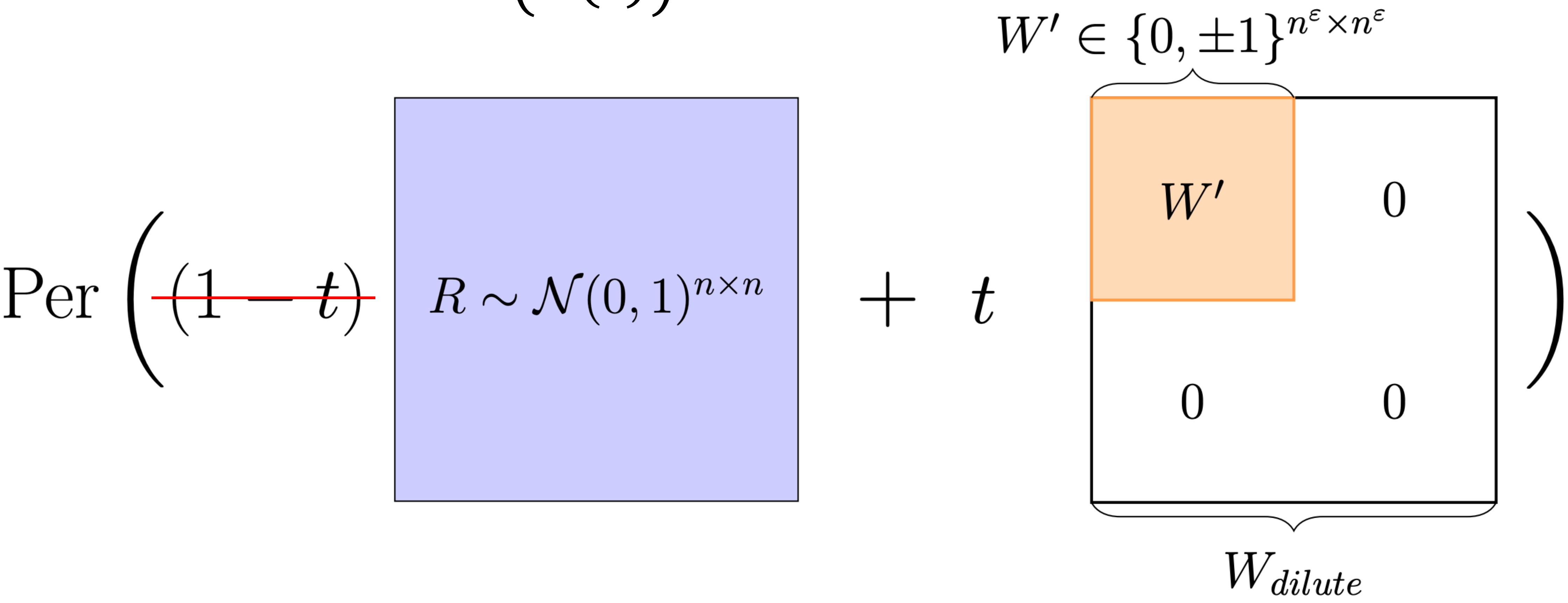
Good news: degree $O(n^\varepsilon)$, extrapolation distance $O(n^\varepsilon)$

Bad news: if W_{dilute} has small support e.g. $O(n^{2\varepsilon})$ nonzero entries, then $\mathrm{Per}(W_{\mathrm{dilute}}) = 0$

So polynomial extrapolation does not encode information about $\mathrm{Per}(W')$ ☹

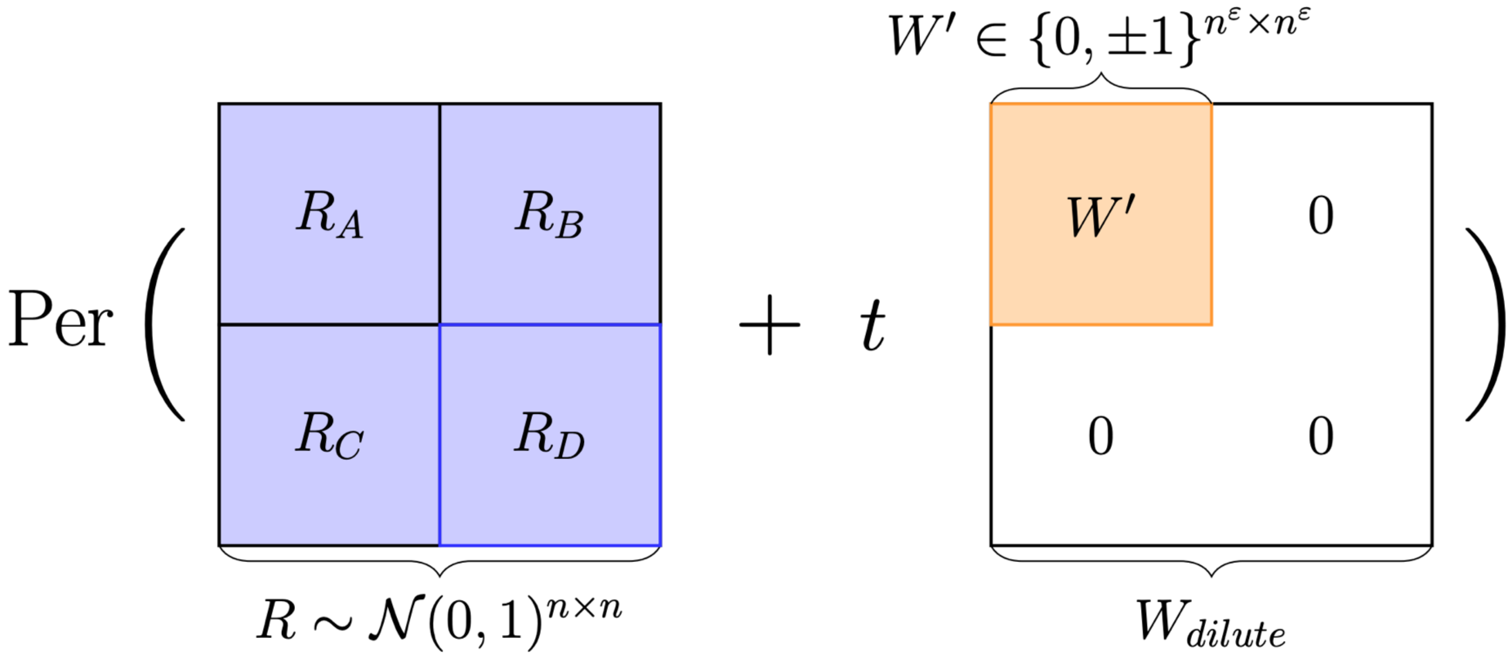
Key idea: coefficient extraction

Consider instead $\text{Per}(R(t)) :=$



Key idea: coefficient extraction

Consider:



Want: Per W'

- $\text{Per}(R(1)) = \text{Per}(R + W_{\text{dilute}})$ is uninteresting
- However, $\text{Per}(R(t))$ still encodes $\text{Per}(W')$:

$$\text{Per}(R(t)) = t^{n^\varepsilon}(\text{Per } W')(\text{Per } R_D) + \sum_{l=0}^{n^\varepsilon - 1} c_l t^l$$

Idea: estimate Per R_D with a recursive call to average-case algo

Feature: our hardness argument circumvents the JSV barrier

Jerrum-Sinclair-Vigoda (JSV) '04: BPP algorithm to approximate the permanent of a **nonnegative** matrix to small relative error

But prior proof techniques were **insensitive** to the difference between nonnegative and mixed sign matrices

By contrast, our proof is **sensitive** to mixed signs

- Our reduction obtains the worst-case permanent to small relative error
- For this to be hard, it needs to have both positive and negative entries

What are the implications for hardness of sampling?

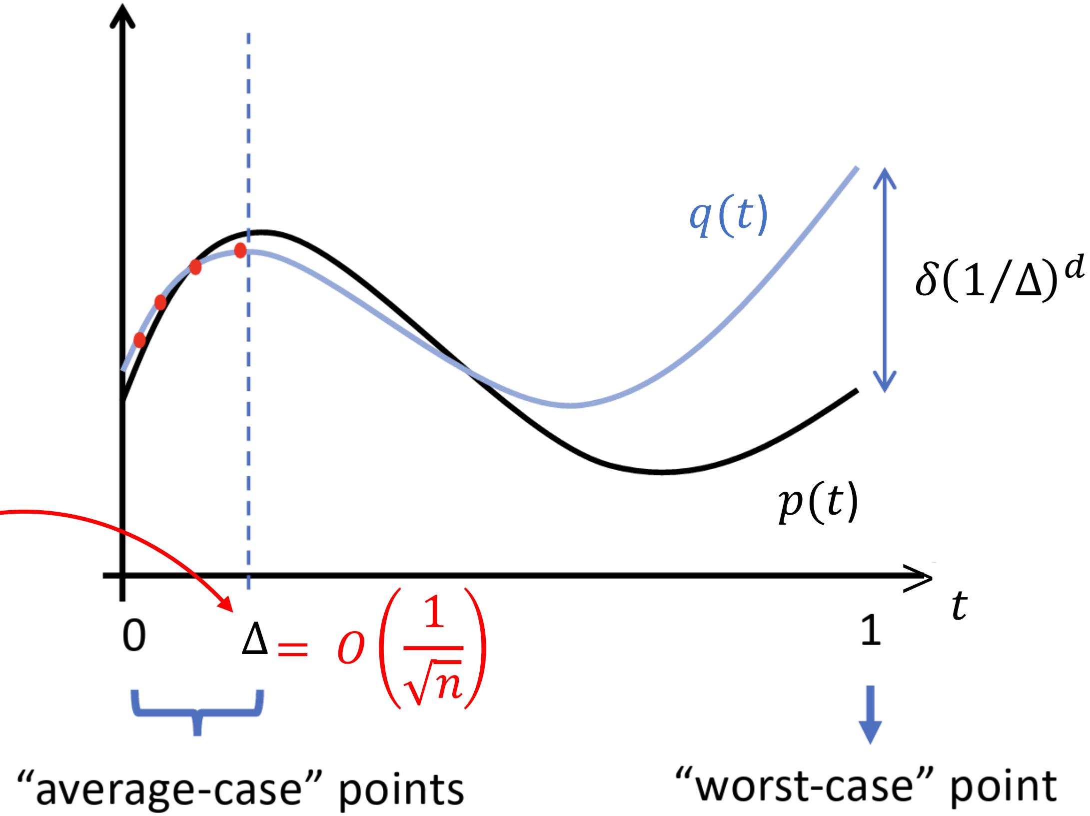
Recall the "moral":

To improve robustness, need to decrease extrapolation distance $1/\Delta$ or decrease the polynomial degree d

Claim: estimating $\mathrm{Per}\big(R(t)\big)$ at $t = O\left(\frac{1}{\sqrt{n}}\right)$

$\Longrightarrow$ hardness of sampling!

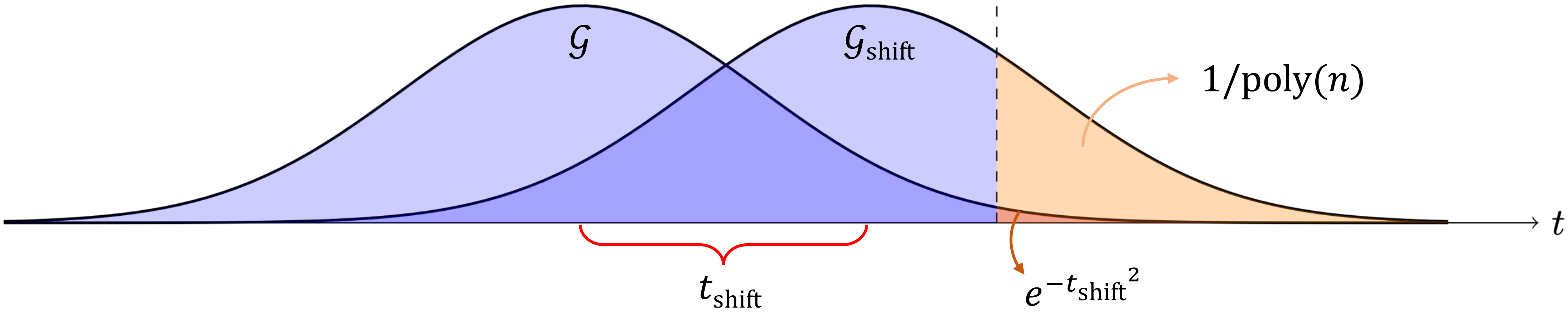
Problem: such $R(t)$ are very *far* from iid Gaussian in TVD—no guarantee algorithm works out-of-distribution



We will show that we *can* estimate these quantities if our average-case algorithm works with sufficiently high probability!

Going out of distribution: rare events lemma

Observe: $R + t_{\text{shift}}W$ is also Gaussian, with shifted mean t_{shift}



We prove:

tail event w.p. $\leq e^{-t_{\text{shift}}^2}$ under $\mathcal{G}$ has prob. $\leq 1/\text{poly}(n)$ under $\mathcal{G}_{\text{shift}}$

If tail event $\equiv$ average-case algorithm fails, then for $\mathcal{G}_{\text{shift}}$,
algorithm fails w.p. $\leq 1/\text{poly}(n)$ if it fails w.p. $\leq e^{-O(n)}$ for $\mathcal{G}$

Hardness of sampling

Combined with a second rare events lemma, we show that this implies:

[This work] There is no classical sampler that succeeds for

$\geq 1 - 2^{-\tilde{O}\left(\sqrt[3]{N}\right)}$ fraction of instances of size N

Trivial: $\geq 1 - 2^{-\tilde{O}(N)}$

But we want to show: $\geq 1 - 1/\mathrm{poly}(N)$

Caveat: because we estimate $R(t)$ far out of distribution, we require a slight generalization of permanent anticoncentration.

With no proof barriers in the way,
can we at last prove PGC?

Thank you! Questions?

http://bit.ly/401GEzy

Anticoncentration conjecture for shifted Gaussian permanents

- Theorem 2 (hardness of sampling) assumes $\mathrm{Per}\big(R(t)\big)$ to anticoncentrate
- This is not implied by standard PACC, as the matrices are out of distribution!

Conjecture. There exists a polynomial f such that for all n and $\epsilon > 0$,

$$\Pr_{R\sim\mathcal{N}(0,1)^{n\times n}}\left[|\mathrm{Per}(R+tW)| < \frac{\sqrt{n!}}{f(n,1/\epsilon)}\right] < \epsilon,$$

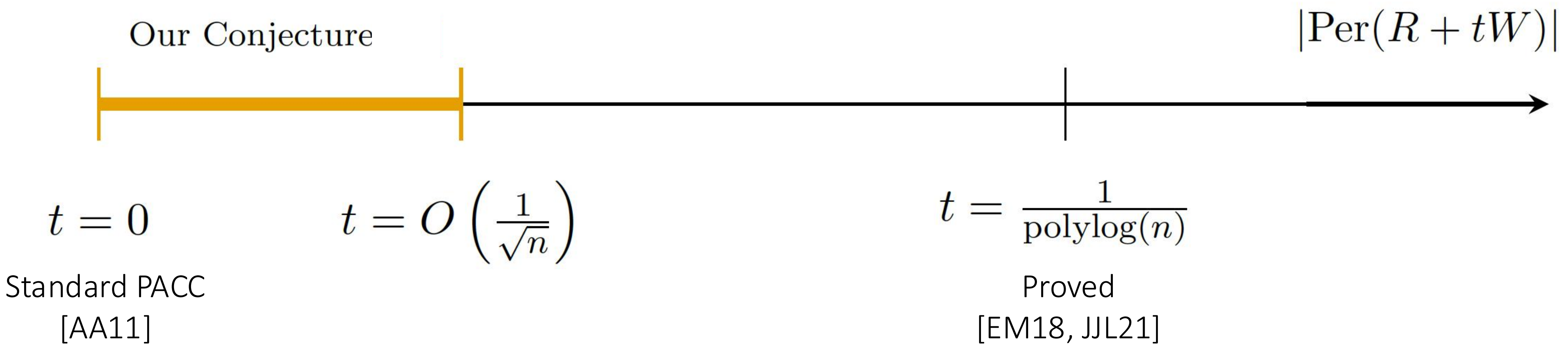
for arbitrary matrix $|\mathrm{Per}W|$ with entries bounded by 1 and $t = O(\frac{1}{\sqrt{n}})$.

Anticoncentration conjecture for shifted Gaussian permanents

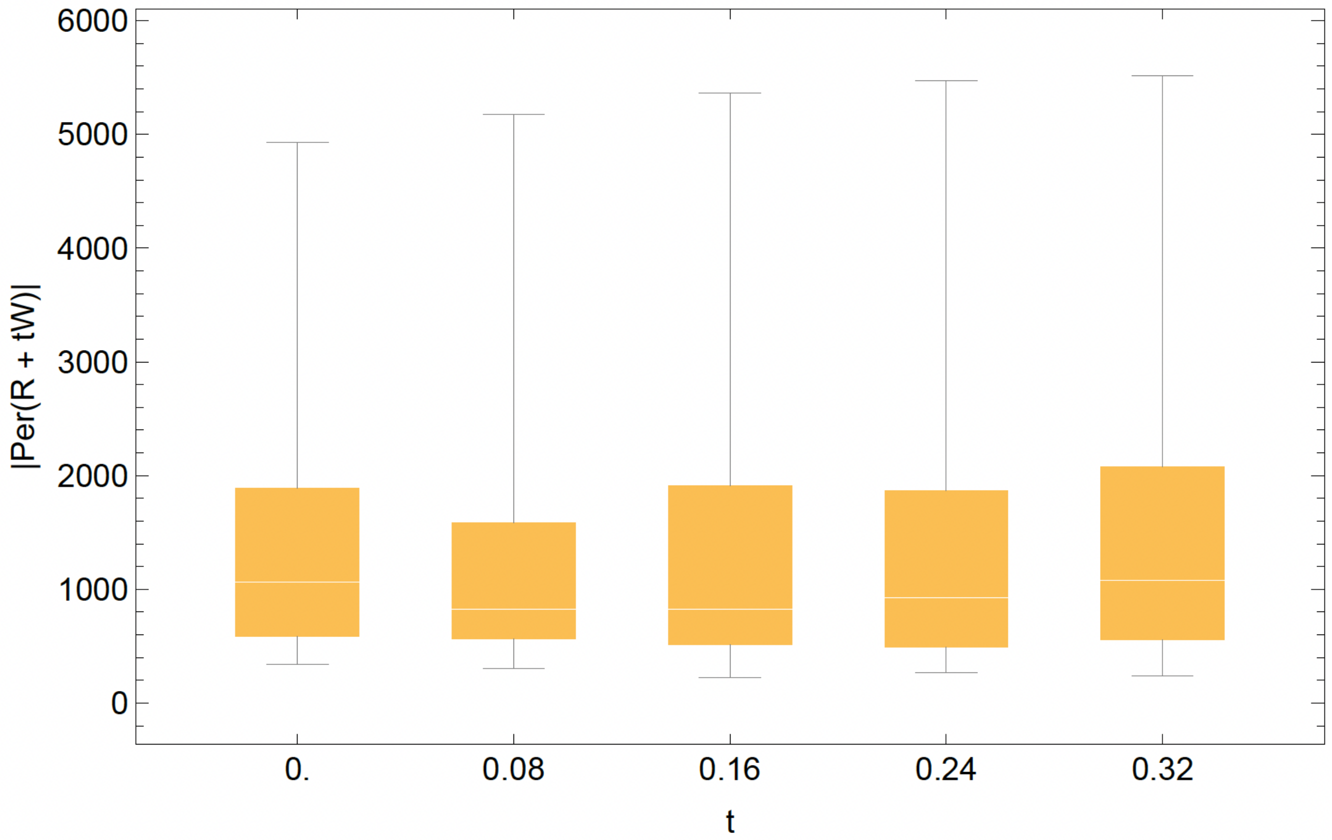
Conjecture. There exists a polynomial f such that for all n and $\epsilon > 0$,

$$\Pr_{R\sim\mathcal{N}(0,1)^{n\times n}}\left[|\mathrm{Per}(R+tW)| < \frac{\sqrt{n!}}{f(n,1/\epsilon)}\right] < \epsilon,$$

for arbitrary matrix $|\mathrm{Per}W|$ with entries bounded by 1 and $t = O(\frac{1}{\sqrt{n}})$.



Numerical evidence for anticoncentration conjecture



Box plots for the distribution of $|\mathrm{Per}(R+tW)|$ for $n = 10$ and $n^{\varepsilon} = 5$. For five equally spaced values of $t \in [0, \frac{1}{\sqrt{n}}]$, we generate 30 such R and W.

Note: Very little variation for increasing t, as conjectured