# Near-Optimal Parallel Approximate Counting via Sampling


Yiyao Zhang

Nanjing University

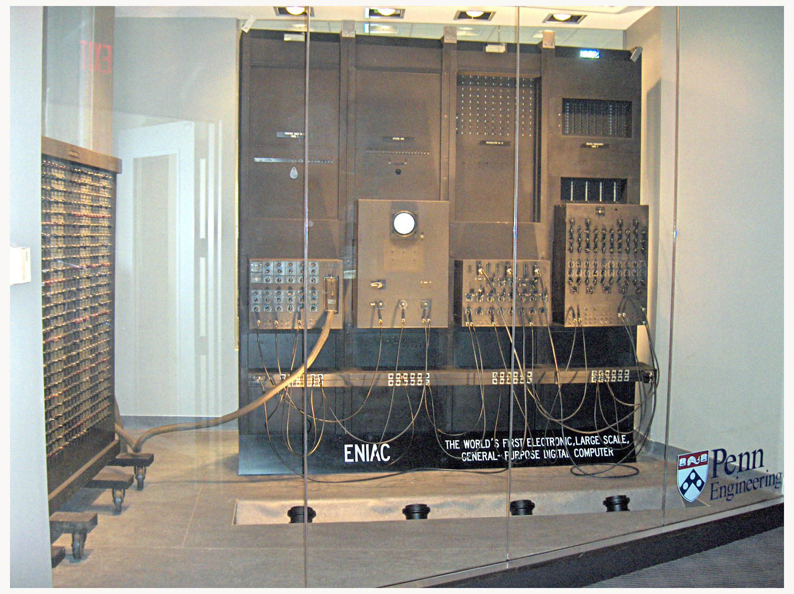
Joint work with:

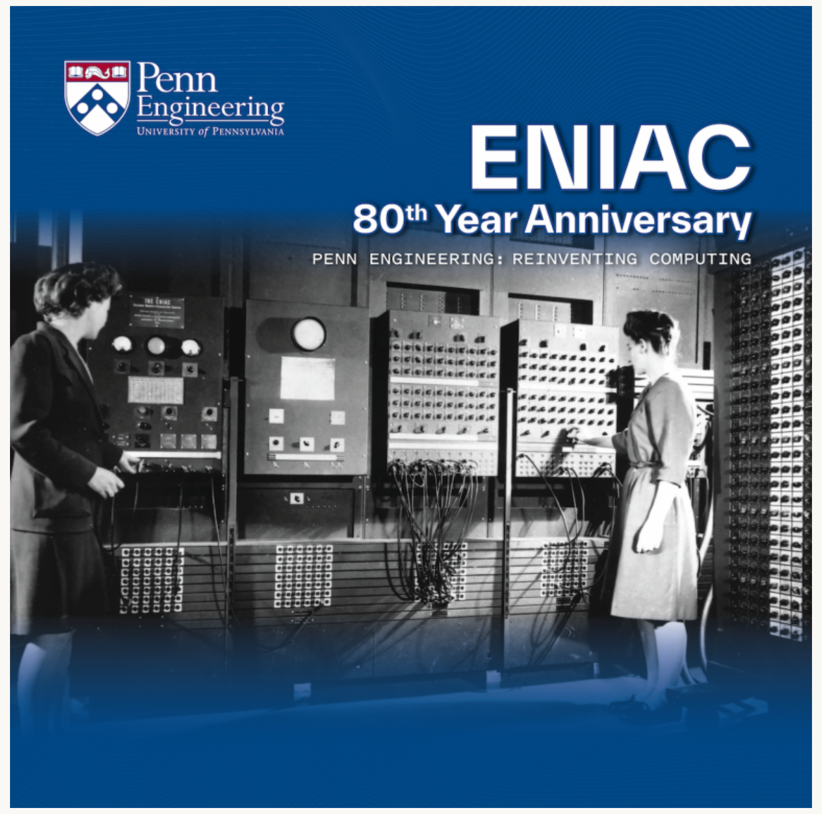
| | |
|---|---|
| David G. Harris | University of Maryland |
| Vladimir Kolmogorov | Institute of Science and Technology Austria |
| Hongyang Liu | Nanjing University |
| Yitong Yin | Nanjing University |

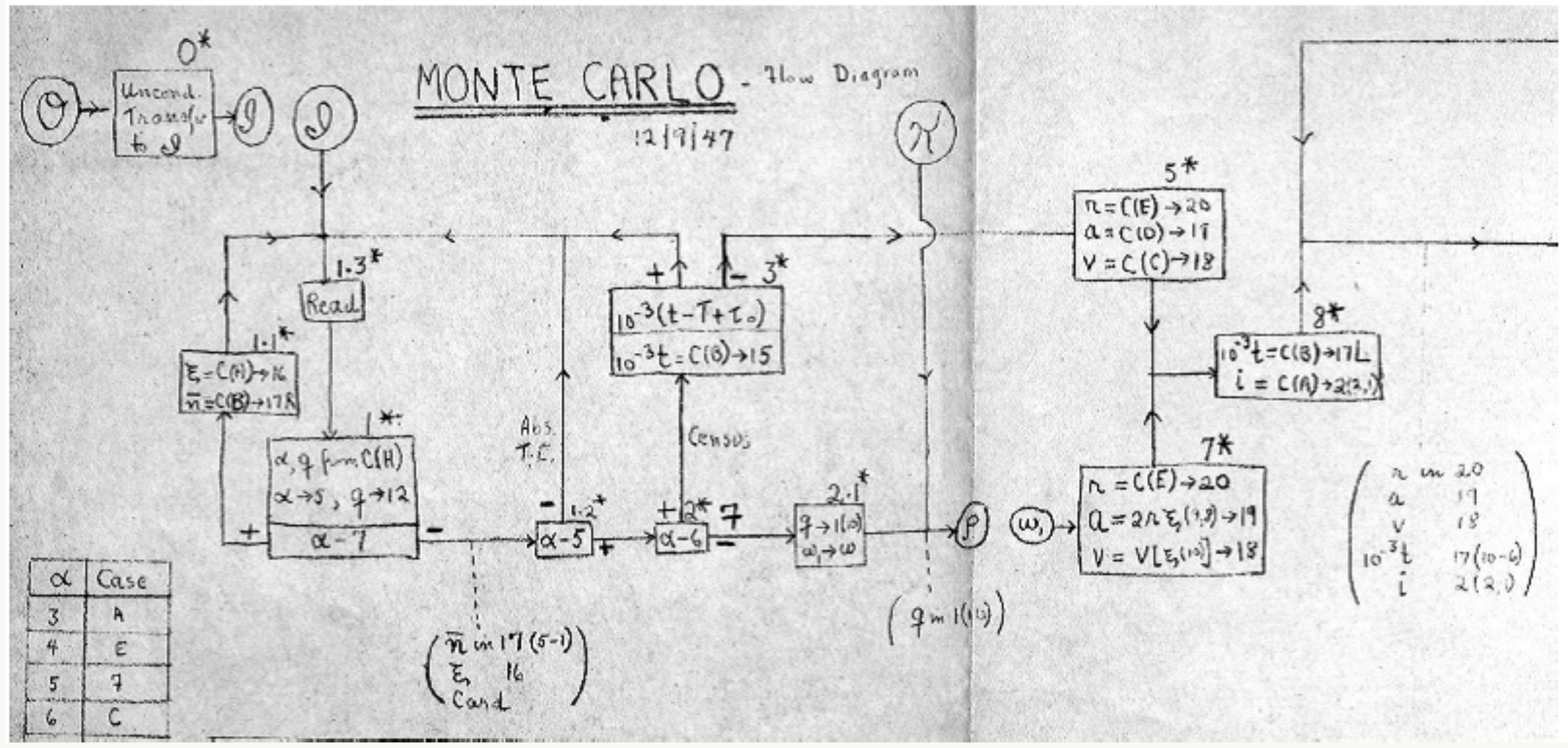


SPAA 2026 London, UK

# Counting and Sampling




Electronic Numerical Integrator and Computer

Monte Carlo Method

Given a finite state set $\Omega$ and unnormalized weight $w : \Omega \to \mathbb{R}_{\geq 0}$

Counting: compute the normalizing factor (partition function) $Z = \sum_{\sigma \in \Omega} w(\sigma)$

Sampling: produce $\sigma \in \Omega$ with $\Pr[\sigma] \propto w(\sigma)$

# Counting and Sampling

[Jerrum, Valiant, Vazirani 1986]

For all self-reducible problems, approximate counting and sampling are computationally equivalent up to polynomial time.

**Sampling** via Counting

**Counting** via Sampling

Encapsulated as an oracle

**RANDOM GENERATION OF COMBINATORIAL STRUCTURES FROM A UNIFORM DISTRIBUTION**

Mark R. JERRUM
*Department of Computer Science, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom*

Leslie G. VALIANT *
*Aiken Computation Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.*

Vijay V. VAZIRANI **
*Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A.*

# Counting and Sampling

**Sampling** via Counting

**Counting** via Sampling

Encapsulated as an oracle

From a parallel perspective:

***how to design reductions with small depth and few oracle calls?***

**Sampling** via Counting

[AHSS21, ABTV22, AHL+23] parallel sampling from determinantal distribution

[AGR24, ABC+26] parallel autoregressive generation

**Counting** via Sampling **Main focus of our paper (Gibbs distribution)**

# Gibbs Distribution

Finite state set $\Omega$ Hamiltonian function $H : \Omega \to \mathbb{R}_{\geq 0}$

Define the Gibbs distribution parameterized by $\beta$:

$$\forall \sigma \in \Omega, \quad \mu_\beta^\Omega(\sigma) \propto \exp(\beta \cdot H(\sigma)).$$

The corresponding partition function:

$$Z(\beta) = \sum_{\sigma \in \Omega} \exp(\beta \cdot H(\sigma)).$$

Approximate the partition function within error $\varepsilon$:

$$(1-\varepsilon) \cdot Z(\beta) \leq \hat{Z}(\beta) \leq (1+\varepsilon) \cdot Z(\beta).$$

Applications:

- Count combinatorial objects

e.g. graph independent sets [LV97]

spin configurations [JS93]

constraint satisfying solutions [FGYZ21]

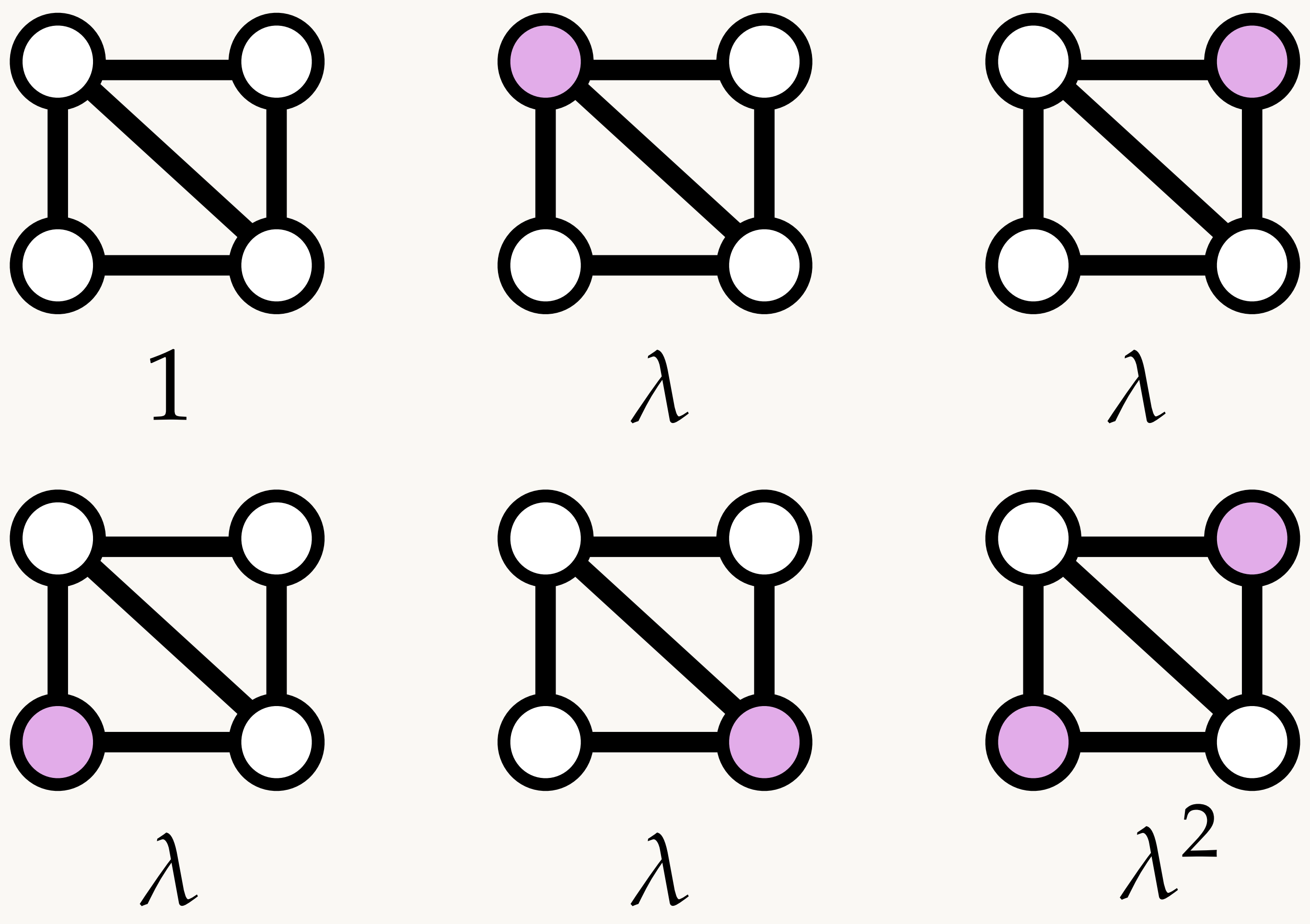
- Approximate permanent [JSV04]
- Estimate volume of convex body [DFK91]

# Gibbs Distribution: Example

Graph independent set (hardcore model)

weight (fugacity) $\lambda \geq 0$ $\quad w(I) = \lambda^{|I|}$

$$G = (V, E) \quad |V| = n \quad |E| = m$$




occupied

unoccupied

$$Z = 1 + 4\lambda + \lambda^2$$

$\lambda = 1$ number of independent sets

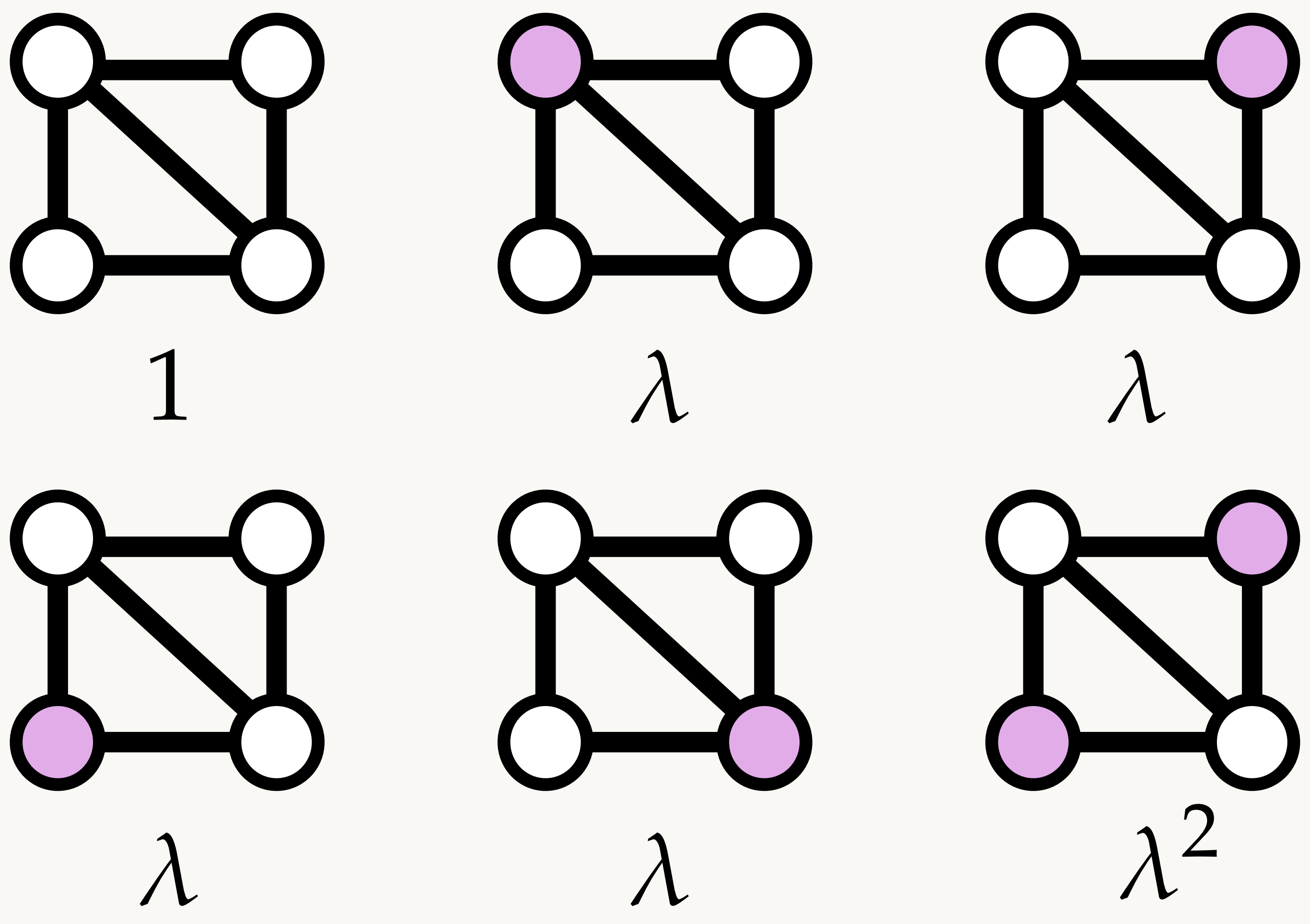
$$\mu(I) = w(I)/Z$$

# Gibbs Distribution: Example

Graph independent set (hardcore model)

weight (fugacity) $\lambda \geq 0 \quad w(I) = \lambda^{|I|}$

$$G = (V, E) \quad |V| = n \quad |E| = m$$




occupied

unoccupied

$$Z = 1 + 4\lambda + \lambda^2$$

$\lambda = 1 \quad$ number of independent sets

$$\mu(I) = w(I)/Z$$

$$\boxed{\begin{gathered}\forall \sigma \in \Omega, \quad \mu_\beta^\Omega(\sigma) \propto \exp(\beta \cdot H(\sigma)). \\ Z(\beta) = \sum_{\sigma \in \Omega} \exp(\beta \cdot H(\sigma)).\end{gathered}}$$

$$\Omega = \{I \subseteq V : I \text{ is an independent set}\}$$

$$H(I) = |I| \qquad \beta = \log \lambda$$

$$\forall I \in \Omega, \quad \mu_\beta^\Omega(I) \propto \lambda^{|I|}$$

$Z(0) =$ number of independent sets

$$Z(-\infty) = 1$$

# Simulated Annealing

$Z(-\infty) = 1$

$Z(0) =$ number of independent sets

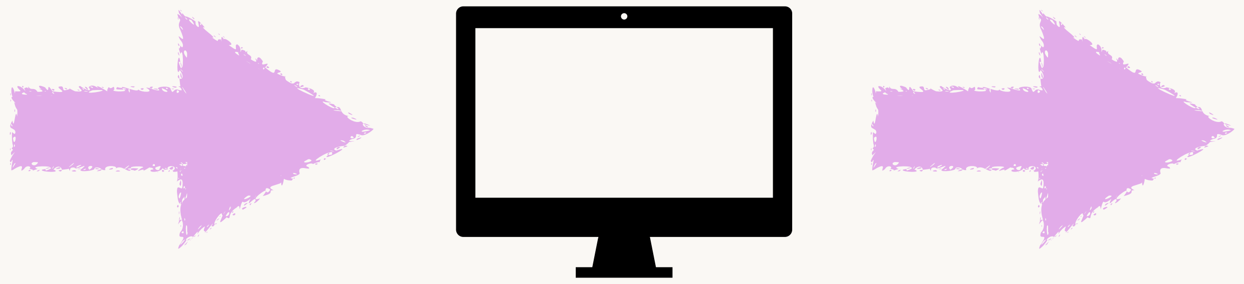

To estimate $Z(0)$, it suffices to estimate the ratio $Q = \dfrac{Z(0)}{Z(-\infty)}$.

# Simulated Annealing

$Z(-\infty) = 1$ $\qquad$ $Z(0) =$ number of independent sets

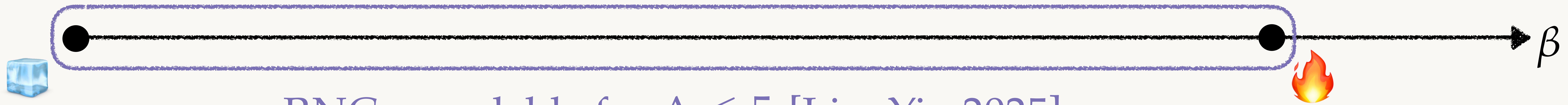

RNC samplable for $\Delta \leq 5$ [Liu, Yin 2025]

To estimate $Z(0)$, it suffices to estimate the ratio $Q = \dfrac{Z(0)}{Z(-\infty)}$ .

Given sampling oracle: Input: $\beta \in [-\infty, 0]$  Output: $X \sim \mu_\beta^\Omega$

# Simulated Annealing

$Z(-\infty) = 1$ $\qquad$ $Z(0) =$ number of independent sets




RNC samplable for $\Delta \leq 5$ [Liu, Yin 2025]

To estimate $Z(0)$, it suffices to estimate the ratio $Q = \dfrac{Z(0)}{Z(-\infty)}$.

Given sampling oracle: $\quad$ Input: $\beta \in [-\infty, 0]$ $\quad$ Output: $X \sim \mu_\beta^\Omega$

***How many computational depth and oracle calls are needed?***

Two parameters $\quad$ $q$ : upper bound for $\log Q$ $\quad$ $h$ : upper bound for $H(\cdot)$

For hardcore model: $\quad$ $q = h = n$ $\qquad$ $G = (V, E)$ $\quad |V| = n$ $\quad |E| = m$

# Our Results: Main Framework

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

Goal: approximate $Q$ within error $\varepsilon$

| Results | Sampling Rounds | Orale Calls |
|---|---|---|
| [Dyer, Frieze, Kannan 1991]<br>[Bezáková, Štefankovič, Vazinari, Vigoda 2008] | 1 | $O(q^2\varepsilon^{-2}\log h)$ |
| [Štefankovič, Vempala, Vigoda 2006] | $\tilde{O}(q^{1/2})$ | $O(q\varepsilon^{-2}\text{polylog}(q,h))$ |
| [Huber 2015], [Kolmogorov 2018],<br>[Harris, Kolmogorov 2024] | $\tilde{O}(q)$ | $O(q\varepsilon^{-2}\log h)$ |
| Ours | 1 | $O(q\varepsilon^{-2}\log^2 h)$ |
| | 3 | $O(q\varepsilon^{-2}\log h)$ |

[Harris, Kolmogorov24]: $\Omega(q\varepsilon^{-2})$ lower bound for number of oracle calls

# Our Results: Main Framework

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

Goal: approximate $Q$ within error $\varepsilon$

| Results | Sampling Rounds | Orale Calls |
|---|---|---|
| [Dyer, Frieze, Kannan 1991]<br>[Bezáková, Štefankovič, Vazinari, Vigoda 2008] | RNC $\boxed{1}$ | $O(q^2\varepsilon^{-2}\log h)$ |
| [Štefankovič, Vempala, Vigoda 2006] | $\tilde{O}(q^{1/2})$ | $O(q\varepsilon^{-2}\text{polylog}(q,h))$ |
| [Huber 2015], [Kolmogorov 2018],<br>[Harris, Kolmogorov 2024] | $\tilde{O}(q)$ | $O(q\varepsilon^{-2}\log h)$ |
| Ours | RNC 1 | $O(q\varepsilon^{-2}\log^2 h)$ |
| | 3 | $O(q\varepsilon^{-2}\log h)$ |

[Harris, Kolmogorov24]: $\Omega(q\varepsilon^{-2})$ lower bound for number of oracle calls

# Our Results: Main Framework

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

Goal: approximate $Q$ within error $\varepsilon$

| Results | Sampling Rounds | Orale Calls |
|---|---|---|
| [Dyer, Frieze, Kannan 1991]<br>[Bezáková, Štefankovič, Vazinari, Vigoda 2008] | RNC $\boxed{1}$ | $O(q^2\varepsilon^{-2}\log h)$ |
| [Štefankovič, Vempala, Vigoda 2006] | $\tilde{O}(q^{1/2})$ | $O(q\varepsilon^{-2}\text{polylog}(q,h))$ |
| [Huber 2015], [Kolmogorov 2018],<br>[Harris, Kolmogorov 2024] | $\tilde{O}(q)$ | $O(q\varepsilon^{-2}\log h)$ |
| Ours | RNC 1 | $O(q\varepsilon^{-2}\log^2 h)$ |
| | 3 | $O(q\varepsilon^{-2}\log h)$ |

Work Efficient

Work Efficient

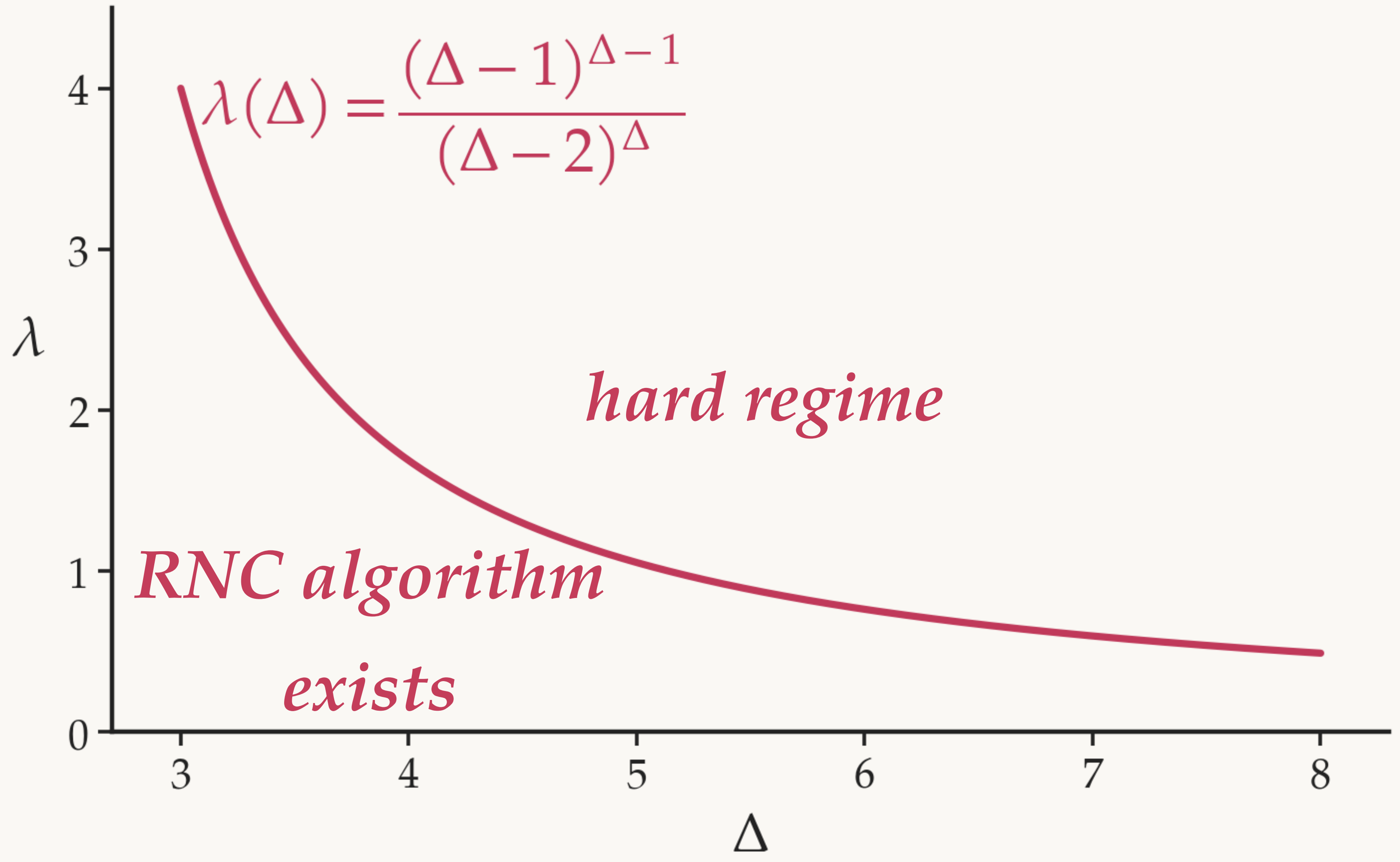
[Harris, Kolmogorov24]: $\Omega(q\varepsilon^{-2})$ lower bound for number of oracle calls

# Our Results: Application

$G = (V, E) \quad |V| = n \quad |E| = m \quad$ maximum degree $\Delta$

$$\Omega = \{I \subseteq V : I \text{ is an independent set}\}$$

weight (fugacity) $\lambda \geq 0 \quad \mu_\lambda(I) \propto \lambda^{|I|} \quad Z(\lambda) = \sum_{I \in \Omega} \lambda^{|I|}$




For $\lambda$ within the uniqueness regime, there exists an algorithm that estimates $Z(\lambda)$ within error $\varepsilon$ in $O(\log^2 n \cdot \log(1/\varepsilon))$ depth using $\tilde{O}(mn\varepsilon^{-2})$ processors.

RNC Sampling [Liu, Yin 2025]

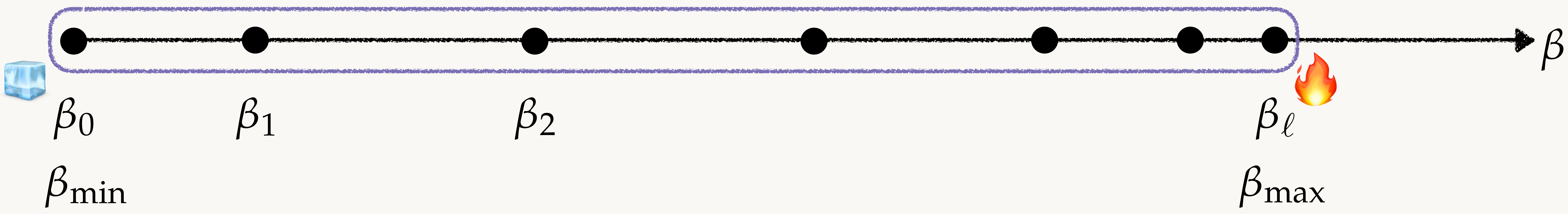
&

Our Counting Framework

*More applications on Monomer-Dimer model, ferromagnetic Ising model*

# Cooling Schedule

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore




Step 1

A cooling schedule
$(\beta_0, \beta_1, \ldots, \beta_\ell)$
with certain conditions

[Huber 2015]
Paired Product Estimator

Step 2 as a black box

A good approximation of $Q$

***one round sampling***

Oracle calls are in parallel
$\tilde{O}(\ell)$ samples are needed

# Cooling Schedule

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

***Main problem: control the length*** $\ell$

Oracle calls are in parallel
$\tilde{O}(\ell)$ samples are needed

# Cooling Schedule

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

***Main problem: control the length*** $\ell$

Oracle calls are in parallel
$\tilde{O}(\ell)$ samples are needed

---

We focus on step 1: construct by calling samples

[Hub15, Kol18, HK24]: construct adaptively (one sample one $\beta$) $\quad \ell = O(q \log h)$

# Cooling Schedule

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

***Main problem: control the length $\ell$***

Oracle calls are in parallel
$\tilde{O}(\ell)$ samples are needed

---

We focus on step 1: construct by calling samples

[Hub15, Kol18, HK24]: construct adaptively (one sample one $\beta$) $\ell = O(q \log h)$

construct directly without any samples $\ell = O(q \log^2 h)$

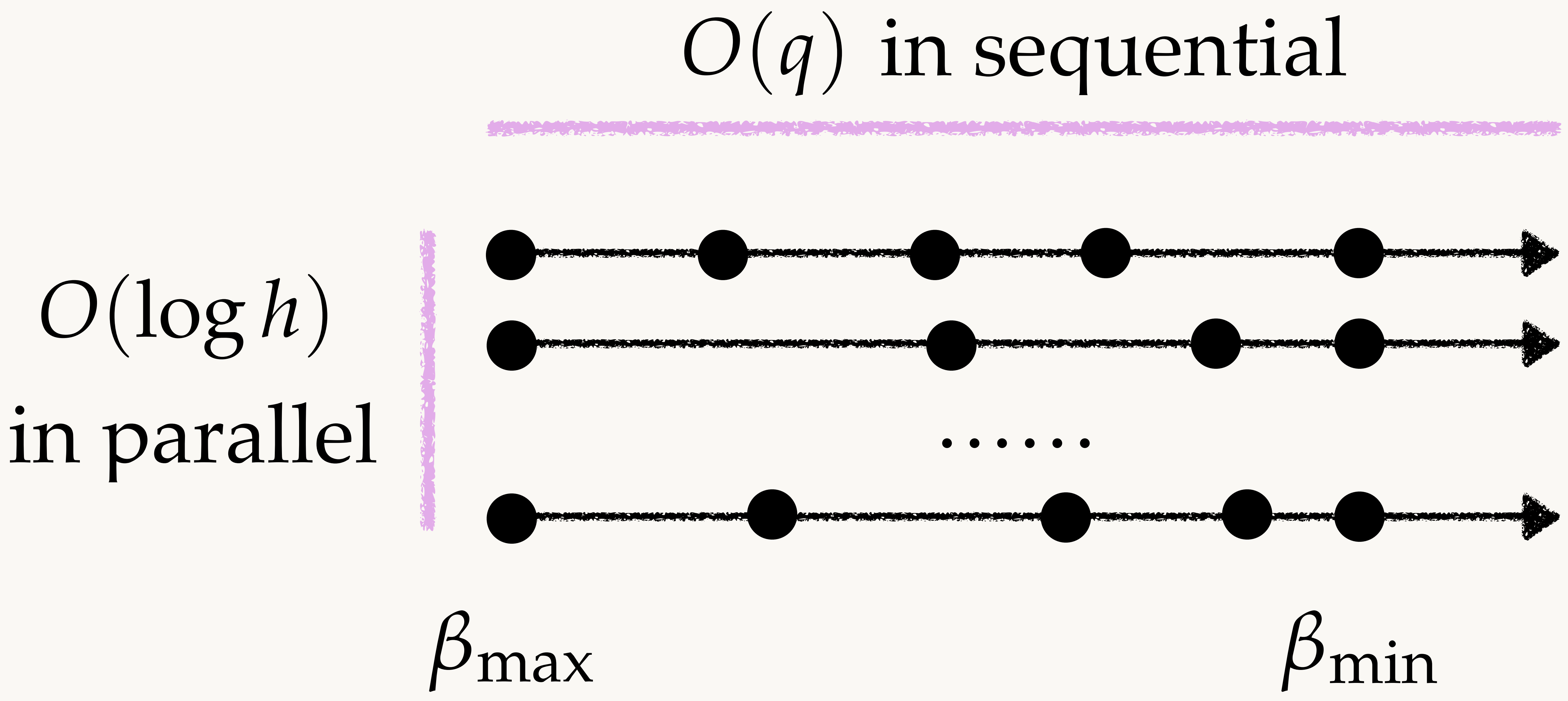
Our work: done by a careful interpolation

construct by calling samples in two rounds $\ell = O(q \log h)$

# Parallel Idea

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

[Harris, Kolmogorov 2024]: TPA process




Merge

$\ell = O(q \log h)$

Poisson Point Process
in the $\log Z(\cdot)$-coordinate
with rate $\Theta(\log h)$

$$Z(\beta) = \sum_{\sigma \in \Omega} \exp(\beta \cdot H(\sigma)).$$

# Parallel Idea

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

Our work

Start from a non-adaptive schedule

$\ell = O(q \log h)$

*one round sampling*

Sparsify by random deleting

rate $\Theta(1)$ $\qquad$ $\ell = O(q)$

*one round sampling*

Repeat one-step TPA process
from points in the sparse schedule

rate $\Theta(\log h)$ $\qquad$ $\ell = O(q \log h)$

# Summary

Finite state set $\Omega$ Hamiltonian function $H : \Omega \to \mathbb{R}_{\geq 0}$

Gibbs distribution $\forall \sigma \in \Omega, \quad \mu_\beta^\Omega(\sigma) \propto \exp(\beta \cdot H(\sigma))$.

The partition function $Z(\beta) = \sum_{\sigma \in \Omega} \exp(\beta \cdot H(\sigma))$.

Goal: estimate the ratio $Q = Z(\beta_{\max}) / Z(\beta_{\min})$ within error $\varepsilon$

$q$ : upper bound for $\log Q$
$h$ : upper bound for $H(\cdot)$
$q = h = n$ for hardcore

---

Given sampling oracle for $\mu_\beta^\Omega$ and $\beta \in [\beta_{\min}, \beta_{\max}]$,

there exists an algorithm using ***one round*** of sampling using $O(q\varepsilon^{-2} \log^2 h)$ samples;

there exists an algorithm using ***three rounds*** of sampling using $O(q\varepsilon^{-2} \log h)$ samples.

**Open Problem: Is one / two round of sampling + $O(q\varepsilon^{-2} \log h)$ samples possible?**

# Thank you!

arXiv: 2604.01263