# Approximating Tensor Network Contractions with Annealed Importance Sampling

Sreevardhan Atyam [1] Anitej Chanda [1] Aniket Deshpande [1] Qizhao Huang [1] Edgar Solomonik [1]

[1]Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign

## Motivation

Tensor networks have emerged as an essential classical framework for efficiently representing quantum states and operators, with applications in quantum simulation, quantum chemistry, and condensed matter physics. However, contracting general tensor networks is a computationally difficult task known to be #P-hard in the general case. Exact methods become expensive as system size increases, requiring scalable and accurate approximation algorithms.

Annealed Importance Sampling (AIS) combined with Markov Chain Monte Carlo (MCMC) techniques offers a promising step forward. By constructing a series intermediate distributions that gradually transition to the target distribution, AIS estimates otherwise intractable contraction values.

In this work, we systematically explore the performance of AIS-based tensor network contraction methods compared to naive Monte Carlo sampling across various tensor network configurations. Our results show clear advantages in computational efficiency, especially in large-scale diagonally-dominant tensor networks relevant to practical quantum circuit simulations.

## Methods

A tensor network (TN) is a pair $\mathcal{T} = (G, \{T_v\}_{v\in\mathcal{V}})$, where

1. $G = (\mathcal{V}, \mathcal{E})$ is an undirected graph. Nodes $v \in \mathcal{V}$ are tensors; edges $e \in \mathcal{E}$ are shared indices.
2. Each tensor $T_v$ is indexed by the edges incident to $v$ and stores values on the discrete domain.

The contraction we wish to approximate is

$$Z = \sum_{\mathbf{x}\in\mathcal{D}} \psi(\mathbf{x}), \qquad \psi(\mathbf{x}) := \prod_{v\in\mathcal{V}} T_v(\mathbf{x}),$$

where $\mathbf{x}$ specifies a value for each edge index, $\mathcal{D} := \prod_{e\in\mathcal{E}}[d_e]$ denotes the domain of all edge assignements and $T_v(\mathbf{x})$ represents the value of tensor node $T_v$ under the assignment $\mathbf{x}$.

**Annealed Importance Sampling (AIS).** Define a continuum of distributions

$$\pi_\beta(\mathbf{x}) \propto \psi(\mathbf{x})^\beta, \qquad \beta \in [0, 1].$$

with a ladder $0 = \beta_0 < \beta_1 < \cdots < \beta_K = 1$ [2]:

- *Initialize.* Sample $\mathbf{x}_0 \sim \pi_{\beta_0}$ (uniform).
- *Glauber step at $\beta_k$.* Perform $N_{\text{mix}}$ steps of single-site updates. At each step, select an edge $e \in \mathcal{E}$ uniformly at random and resample from the conditional:

$$x_e \sim \pi_{\beta_k}(x_e \mid \mathbf{x}_{\setminus e}) \propto \prod_{\substack{v\in\mathcal{V} \\ e \text{ incident to } v}} T_v(\ldots, x_e, \ldots)^{\beta_k}.$$

  Here, $N_{\text{mix}}$ denotes the number of Glauber updates per chain per $\beta_k$, chosen to ensure adequate mixing before estimating expectations.
- *Incremental weight.* For each $\beta_k$, we draw $N$ independent samples $\mathbf{x}_1^{(k-1)}, \ldots, \mathbf{x}_N^{(k-1)} \sim \pi_{\beta_k}$ via Glauber dynamics, and compute:

$$\hat{w}_k = \frac{1}{N}\sum_{i=1}^{N} \psi\left(\mathbf{x}_i^{(k-1)}\right)^{\Delta\beta_k} = \frac{1}{N}\sum_{i=1}^{N}\left[\prod_v T_v\left(\mathbf{x}_i^{(k-1)}\right)\right]^{\Delta\beta_k}.$$

The AIS estimator is

$$\hat{Z} = Z_0 \cdot \prod_{k=1}^{K} \hat{w}_k.$$

## Methods (Continued)

AIS stems from the telescoping identity

$$Z = Z_0 \prod_{k=1}^{K} \frac{Z_{\beta_k}}{Z_{\beta_{k-1}}}, \qquad Z_\beta := \sum_{\mathbf{x}} \psi(\mathbf{x})^\beta,$$

where $Z_0 = \prod_{e\in\mathcal{E}} d_e$ and each ratio admits the expectation form

$$\rho_k := \frac{Z_{\beta_k}}{Z_{\beta_{k-1}}} = \mathbb{E}_{\mathbf{x}\sim\pi_{\beta_k}}\left[\psi(\mathbf{x})^{\Delta\beta_k}\right] \quad \text{where} \quad \pi_{\beta_k}(\mathbf{x}) \propto \psi(\mathbf{x})^{\beta_k}.$$

To quantify estimator accuracy for each ratio, we report the *geometric mean relative error* across multiple trials:

$$\text{GMRE}(\hat{\rho}_k) = \exp\left(\frac{1}{T}\sum_{t=1}^{T} \log\left(\left|\frac{\hat{\rho}_k^{(t)} - \rho_k}{\rho_k}\right|\right)\right),$$

where $\hat{\rho}_k^{(t)}$ is the $t$-th estimate of the ratio $Z_{\beta_k}/Z_{\beta_{k-1}}$.

**Key idea.** Glauber dynamics rapidly mixes each chain towards $\pi_{\beta_k}$, while AIS converts local samples into a global estimate through the product of expected ratios.

## Test Structures

We test our algorithm on various tensor networks by evaluating its relative error and scaling behavior under several diagnostic settings.


This tensor network diagram corresponds to the contraction $\sum_{ijkl} A_{ij}B_{jk}C_{kl}D_{li} = \text{Tr}(ABCD)$. We extend this trace of products to a lattice of nine tensors, arranged in a $3 \times 3$ grid, where each interior tensor contracts with its four neighbors (north, south, east, west) and each edge tensor contracts with 2–3 others, forming a closed square grid with periodic or open boundary conditions. The total contraction sums over all internal indices and represents a natural testbed for scaling behavior and high-dimensional complexity.

$$Z = \sum_{\{i_{ab}\}} \prod_{v\in\mathcal{V}} T_v\left(i_{v,1}, i_{v,2}, i_{v,3}, i_{v,4}\right)$$

We construct *diagonally-dominant* tensors in our tests. To make a tensor $T_{ij}$ diagonally-dominant, the elements when $i = j$ have higher values than when $i \neq j$. We do this by defining diagonal and noise components.

$$D(i_1, i_2, \ldots, i_n) = \begin{cases} 1, & \text{if } i_1 = i_2 = \cdots = i_n \\ 0, & \text{otherwise} \end{cases}, \qquad N(i_1, i_2, \ldots, i_n) = 1.$$

With a noise level $\epsilon$, the final $d$-dimensional, order-$n$ tensor is:

$$T_v = (1 - \epsilon)D + \left(\frac{\epsilon}{d^n}\right)N.$$

This structure is well-suited for testing AIS algorithms due to their structured yet non-trivial nature. The majority of the total probability mass is concentrated on a small subset of configurations, making Glauber updates difficult in the high-$\beta$ regime. Additionally, diagonally-dominant tensors are integral to quantum circuit simulation, as they naturally represent $T$, $CZ$, and phase gates, which remain prevalent in variational quantum algorithms and stabilizer codes.

## Test Results

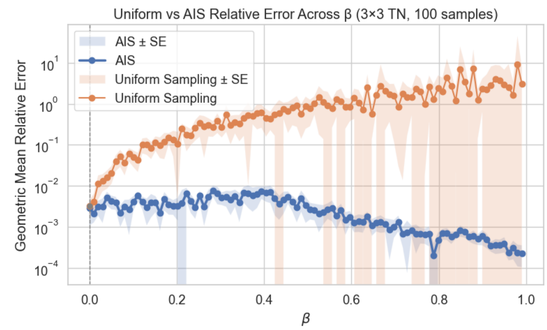


Figure 1. *

**Uniform vs AIS Sampling Error Across $\beta$.** Geometric mean relative error of the single-step ratio estimate as a function of annealing step $\beta_k$ with shaded standard error.

Figure 1 compares uniform sampling and AIS under matched sampling budgets across $\beta$. At $\beta \approx 0$, the target distribution is nearly uniform, so both estimators (sampling from or mixing towards the same distribution) achieve comparable accuracy. As $\beta$ increases and the distribution sharpens, uniform sampling becomes increasingly biased, while AIS leverages local MCMC Glauber updates to maintain accuracy.

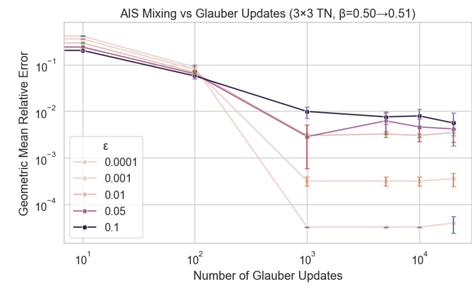


Figure 2. *

**AIS Mixing vs Glauber Updates.** Geometric mean relative error of the single-step AIS ratio estimate after $N = 20$ chains and 30 trials per setting, with 95% confidence intervals.

Figure 2 examines how MCMC mixing affects AIS accuracy across noise levels $\epsilon$. For small $\epsilon$ (structured tensors), even few Glauber updates yield accurate estimates. For larger $\epsilon$ (near-uniform tensors), additional mixing has little effect. This reflects a signal-to-noise tradeoff: as $\epsilon \to 1$, the true ratio $Z(\beta_k)/Z(\beta_{k-1}) \to 1$, reducing sensitivity to configuration quality and limiting the gains from extra mixing.

## References

[1] Elizabeth Crosson and Aram W. Harrow. Rapid mixing of path integral monte carlo for 1d stoquastic hamiltonians. *Quantum*, 5:395, 2021. doi: 10.22331/q-2021-02-11-395.

[2] Radford M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001. doi: 10.1023/A:1008923215028.