# Parallel Discrete Sampling via Continuous Walks

Nima Anari[1], Yizhi Huang[2], Tianyu Liu[1], Thuy-Duong Vuong[1], Brian Xu[1], and Katherine Yu[1]

[1]Stanford University, {anari,tliu248,tdvuong,brianx,katherineyu}@stanford.edu
[2]Tsinghua University, huangyizhi01@gmail.com

**Abstract**

We develop a framework for sampling from discrete distributions $\mu$ on the hypercube $\{\pm 1\}^n$ by sampling from continuous distributions supported on $\mathbb{R}^n$ obtained by convolution with spherical Gaussians. We show that for well-studied families of discrete distributions $\mu$, convolving $\mu$ with Gaussians yields well-conditioned log-concave distributions, as long as the variance of the Gaussian is above an $O(1)$ threshold. We then reduce the task of sampling from $\mu$ to sampling from Gaussian-convolved distributions. Our reduction is based on a stochastic process widely studied under different names: backward diffusion in diffusion models, and stochastic localization. We discretize this process in a novel way that allows for high accuracy and parallelism.

As our main application, we resolve open questions Anari, Hu, Saberi, and Schild raised on the parallel sampling of distributions that admit parallel counting. We show that determinantal point processes can be sampled via RNC algorithms, that is in time $\log(n)^{O(1)}$ using $n^{O(1)}$ processors. For a wider class of distributions, we show our framework yields Quasi-RNC sampling, i.e., $\log(n)^{O(1)}$ time using $n^{O(\log n)}$ processors. This wider class includes non-symmetric determinantal point processes and random Eulerian tours in digraphs, the latter nearly resolving another open question raised by prior work. Of potentially independent interest, we introduce and study a notion of smoothness for discrete distributions that we call transport stability, which we use to control the propagation of error in our framework. Additionally, we connect transport stability to constructions of optimally mixing local random walks and concentration inequalities.

# 1 Introduction

The conventional wisdom in the field of sampling and counting is that the tasks of approximate sampling and approximate counting are algorithmically equivalent, i.e., an efficient algorithm for one can be translated to an efficient algorithm for the other. This was formalized in a celebrated result of Jerrum, Valiant, and Vazirani [JVV86] for self-reducible problems, but the equivalence holds in many other cases beyond the boundary of self-reducibility [see, e.g., DGJ04]. This equivalence is frequently used to convert sampling algorithms such as those based on Markov chains, into approximate counting algorithms. For example, the breakthrough results of Jerrum, Sinclair, and Vigoda [JSV04] for approximating the permanent and Dyer, Frieze, and Kannan [DFK91] for approximating the volume of convex sets both rely on reductions from counting to sampling.

A somewhat overlooked caveat about the equivalence of sampling and counting is that the standard reductions in both directions only preserve polynomial runtimes. For efficiency criteria stricter than polynomial runtime, say *parallel efficiency*, the reductions break down. This is most palpable for a range of problems that admit determinant-based counting algorithms. For example, Kirchhoff's matrix-tree theorem [see, e.g., CK78] allows us to compute the number of spanning trees in a graph as the determinant of a submatrix of its Laplacian. The classic BEST theorem [AB51; TS41] relates the count of Eulerian tours in directed graphs to arborescences, which can be written as a determinant by a generalization of the matrix-tree theorem. In another example, Temperley and Fisher [TF61] and Kasteleyn [Kas63] showed how to compute the number of perfect matchings in planar graphs as determinants of carefully signed adjacency matrices. Computing determinants is in the class NC [Csa76], which means that the number of spanning trees, directed Eulerian tours, and planar perfect matchings on graphs of size $n$ can all be computed in $\log(n)^{O(1)}$ time using $n^{O(1)}$ processors on a PRAM. But this does not automatically translate to parallel algorithms for sampling uniformly random spanning trees, directed Eulerian tours, or random planar perfect matchings. The main question we address in our work is:

Can parallel counting algorithms yield parallel sampling algorithms?

Noting that the standard reduction from sampling to counting is sequential, Teng [Ten95] and Anari, Hu, Saberi, and Schild [Ana+21] raised questions about designing efficient parallel sampling algorithms. In particular, Anari, Hu, Saberi, and Schild [Ana+21] enumerated a list of problems known to admit parallel determinant-based counting and raised as an open question designing RNC-type sampling algorithms for them, i.e., sampling from these distributions using $n^{O(1)}$ processors in $\log(n)^{O(1)}$ time. The list of problems includes arborescences in directed graphs (generalizing spanning trees), Eulerian tours in digraphs, determinantal point processes (DPP) and variants of them, and planar perfect matchings. So far, only one problem from this list, arborescences, has found an efficient parallel sampler [Ana+21]. In this work, we resolve or nearly resolve all but the last problem in the list.

We consider sampling from a distribution $\mu$ on the hypercube $\{\pm 1\}^n$, which we identify with subsets of $\{1, \ldots, n\}$. All aforementioned applications, with the exception of Eulerian tours in digraphs, have natural descriptions as set families; for reductions from Eulerian tours to the hypercube, see Section 6.2. Our main result reduces sampling from $\mu$ to computing partition functions, a.k.a. counting, for exponential tilts of $\mu$, a.k.a. external fields applied to $\mu$. An exponential tilt is defined by a vector $w \in \mathbb{R}^n$, and we denote it by $\tau_w \mu$:

$$\tau_w \mu(x) \propto \exp(\langle w, x \rangle) \cdot \mu(x).$$

The normalizing factor in this definition, i.e., the partition function, is $\sum_x \exp(\langle w, x \rangle) \mu(x)$. Viewed as

a function of $w$, the partition function is also known as the Laplace transform of $\mu$. Exponential tilts are widely studied in sampling [see, e.g., ES22; Ana+22b; CE22], but they have a special relationship with determinant-based counting because roughly speaking, the tilt $\tau_w$ corresponds to scaling the rows and/or columns of the matrix whose determinant provides the count, hence, these distribution families are closed under tilts. As an example, tilts of the uniform spanning tree distribution are simply weighted spanning tree distributions for which the matrix-tree theorem still provides the partition function.

For our main results, we prove that if tilts of $\mu$ vary in a stable manner with respect to the tilt parameter $w$, for notions of "stability" that will be made precise, then there is an efficient parallel reduction from approximately sampling $\mu$ to computing the Laplacian of $\mu$. We then obtain our main applications, parallel samplers for a range of distributions, by proving the required notions of stability for the distributions of interest.

Our strongest main result is obtained when the transportation distance between tilts $\tau_w\mu$ can be bounded in terms of the distance of their tilt parameters $w$. We call this new concept *transport-stability*. As side results, in Appendix A, we connect notions of transport-stability to concentration inequalities and the existence of fast mixing local Markov chains. For our main result, we use a specific instantiation of transport-stability defined using the Wasserstein metric w.r.t. the Hamming distance for tilts $\tau_w\mu$, and the $\ell_1$ distance between tilt parameters $w$. Denote by $\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu)$ the average Hamming distance traveled, when probability mass is transported optimally from $\tau_w\mu$ to $\tau_{w'}\mu$. Then $\mathcal{W}_1$-$O(1)\ell_1$ transport-stability is the notion that for all $w, w'$ we have

$$\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu) \le O(1) \cdot \|w - w'\|_1.$$

With transport-stability in mind, we now state our main result:

**Theorem 1.** *Suppose that a distribution $\mu$ on $\{\pm 1\}^n$ is $\mathcal{W}_1$-$O(1)\ell_1$ transport-stable and we have an oracle for approximately computing the Laplace transform of $\mu$. Then we can sample from a distribution $\epsilon$-close in total variation distance to $\mu$, in $\log(n/\epsilon)^{O(1)}$ time using $(n/\epsilon)^{O(1)}$ processors.*

We show that (symmetric) determinantal point processes and partition-constrained versions of them satisfy $\mathcal{W}_1$-$O(1)\ell_1$ transport stability. A symmetric determinantal point process (DPP) is defined by an $n \times n$ matrix $L \succeq 0$ and samples a subset $S$ of $\{1, \dots, n\}$ with $\mathbb{P}_\mu[S] \propto \det(L_{S,S})$. DPPs and variants of them, such as partition-constrained DPPs, are widely used in numerical linear algebra and machine learning [KT12; Cel+18; Der21] . As a remark, uniformly random spanning trees are DPPs too [see, e.g., KT12]. For more detailed definitions, see Section 2.

**Corollary 2.** *Suppose that $\mu$ is a DPP or a partition-constrained DPP with $O(1)$ parts on a ground set of size $n$. Then there is an algorithm to sample $\epsilon$-closely from $\mu$ in total variation distance, in $\log(n/\epsilon)^{O(1)}$ time using $(n/\epsilon)^{O(1)}$ processors.*

We next show that a weaker notion of stability is enough for efficient parallel sampling if we allow quasi-polynomially many processors. We use the notion coined *semi-log-concavity* by Eldan and Shamir [ES22], which can be stated equivalently as

$$\|\mathrm{mean}(\tau_w\mu) - \mathrm{mean}(\tau_{w'}\mu)\|_2 \le O(1) \cdot \|w - w'\|_2,$$

where $\mathrm{mean}(\mu) = \mathbb{E}_{x\sim\mu}[x] \in [-1, +1]^n$ is the mean vector.[1]

[1]The use of $\|\cdot\|_2$ instead of $\|\cdot\|_1$, which is closer to transport-stability, only makes this condition weaker.

**Theorem 3.** *Suppose that a distribution $\mu$ on $\{\pm 1\}^n$ is semi-log-concave and we have an oracle for computing the Laplace transform of $\mu$. Then we can sample from a distribution $\epsilon$-close in total variation distance to $\mu$, in $\log(n/\epsilon)^{O(1)}$ time using $(n/\epsilon)^{O(\log n)}$ processors.*

Semi-log-concavity is a relatively weak assumption. It is satisfied by many widely-studied classes of distributions, such as Rayleigh distributions [ES22], distributions spectrally independent under tilts, a.k.a. fractionally log-concave distributions [ES22; Ali+21], and sector-stable distributions [Ali+21]. We use these implications to obtain the following corollary about non-symmetric DPPs and Eulerian tours. A non-symmetric DPP is defined the same way as a (symmetric) DPP, except the matrix $L$ is not required to be symmetric, and is only required to satisfy $L + L^\intercal \succeq 0$ [Gar+19]. For a more detailed definition see Section 2.

**Corollary 4.** *Suppose that $\mu$ is a non-symmetric DPP on a ground set of size $n$ or the distribution of uniformly random Eulerian tours in a digraph of size $n$. Then we can sample from a distribution $\epsilon$-close in total variation distance to $\mu$ in time $\log(n/\epsilon)^{O(1)}$ using $(n/\epsilon)^{O(\log n)}$ processors.*

For all of the applications mentioned in Corollaries 2 and 4, our work is the first to obtain the ultimate goal of polylogarithmic parallel runtime.

### 1.1 Discrete sampling via continuous sampling

We obtain our results through a framework that reduces discrete sampling to continuous sampling, by running the stochastic localization process of Eldan [Eld13] in discrete time steps. The equivalence to stochastic localization will be shown in Section 3 by appealing to an alternative characterization of stochastic localization due to El Alaoui and Montanari [EM22].

We reduce sampling from $\mu$ to the task of sampling from tilts of $\mu$ convolved with spherical Gaussians $\mathcal{N}(0, cI)$ of variance $c \in \mathbb{R}_{>0}$. We denote convolutions by $*$. The convolution $\mu * \nu$ is the distribution of $x + y$ if $x \sim \mu$ and $y \sim \nu$ are independent samples. Our framework is described in Algorithm 1.

**Algorithm 1:** Framework for discrete sampling via continuous sampling

$w_0 \leftarrow 0$
**for** $i = 0, \dots, T-1$ **do**
  $x \leftarrow$ (approximate) sample from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$
  $w_{i+1} \leftarrow w_i + x/c$
**return** $\operatorname{sign}(w_T) \in \{\pm 1\}^n$

This framework is parameterized by the number of steps $T \in \mathbb{N}$, the variance $c \in \mathbb{R}_{>0}$, and quite importantly, the choice of how to implement sampling from $\tau_{w_i}\mu * \mathcal{N}(0, cI)$.

One might wonder at first glance if sampling from the convolved distribution is any easier than sampling from $\mu$ itself. Our key observation is that $\tau_w\mu * \mathcal{N}(0, cI)$ is a continuous well-conditioned log-concave distribution, i.e., the "easiest kind" of continuous distribution for sampling; we show that well-conditioned log-concavity for a constant $c$ is precisely equivalent to semi-log-concavity of $\mu$. Surprisingly, this does not seem to have been observed in prior works [ES22; CE22].

**Lemma 5.** *For any semi-log-concave distribution $\mu$ and any $w \in \mathbb{R}^n$, the distribution $\nu = \tau_w\mu * \mathcal{N}(0, cI)$ is well-conditioned log-concave for $c$ larger than a $O(1)$ threshold. Well-conditioned log-concavity means that $-\alpha I \preceq \nabla^2 \log \nu \preceq -\beta I$ for some $\alpha, \beta > 0$ such that $\alpha/\beta = O(1)$.*

We next make the crucial observation that the density of $\tau_w \mu * \mathcal{N}(0, cI)$ and its derivatives can be computed via the Laplace transform of $\mu$. This enables us to utilize a plethora of off-the-shelf continuous sampling methods, which only need access to the density and low-order derivatives of it. For our applications, we choose the randomized midpoint algorithm of Shen and Lee [SL19], which is importantly parallelizable.

**Lemma 6.** *The density, up to normalization, and the gradient of the log-density for $\tau_w \mu * \mathcal{N}(0, cI)$ can be computed efficiently in parallel, in $\log(n)^{O(1)}$ time using $n^{O(1)}$ processors, given access to an oracle for the Laplace transform of $\mu$.*

We formally show by directly appealing to the characterization of El Alaoui and Montanari [EM22] that Algorithm 1, assuming no approximation error in continuous sampling, is a faithful simulation of stochastic localization at discrete time steps. As a consequence we show that with perfect continuous sampling, $c/T \cdot w_T$ is distributed as $\mu * \mathcal{N}(0, c/T \cdot I)$. For roughly $T \simeq c \log n$ the variance $c/T$ is low enough that $\operatorname{sign}(w_T)$ will be distributed approximately as $\mu$. This means that under perfect continuous sampling, the process needs to be run for only $\simeq c \log n$ steps.

**Lemma 7.** *If continuous samples are exact in Algorithm 1, then for $T = \Omega(c \log(n/\epsilon))$ the output of the algorithm is $\epsilon$-close in total variation distance to $\mu$.*

Finally, we need to deal with the issue of approximation error in the continuous sampling step. This is the most difficult part in our analysis and we tackle it in Section 4. Errors in one step change the tilt parameter $w_i$ of future steps, and a priori, that can significantly change $\tau_{w_i} \mu$. Here, we introduce and use the notion of transport stability to control the propagation of error. If our distribution satisfies $\mathcal{W}_1$-$\ell_1$ transport stability:

$$\mathcal{W}_1(\tau_w \mu, \tau_{w'} \mu) \leq C \cdot \|w - w'\|_1,$$

the errors in continuous sampling grow only exponentially at a rate dictated by $C$. For constant $C$, after $T \simeq O(\log n)$ steps, they would be only polynomially larger, and therefore it is enough to start with an inverse-polynomially small continuous sampling error. We show how to conclude Theorem 1 from this analysis. We then note that transport stability is satisfied with a super-constant parameter $C \simeq n$ for any distribution $\mu$, which results in a quasi-polynomial blowup of error. Therefore, by having a continuous sampling procedure that is accurate within an inverse quasi-polynomially small error, we obtain Theorem 3.

### 1.2 Related work

Our general sampling framework is based on stochastic localization [Eld13], which has been a very successful analysis tool [see, e.g., CE22]. Stochastic localization has also been recently used as an algorithm by El Alaoui, Montanari, and Sellke [EMS22] for the task of sampling from the Sherrington-Kirkpatrick (SK) model. Our framework, Algorithm 1, is a new way of using stochastic localization algorithmically. We briefly describe the differences with the prior algorithmic use. El Alaoui, Montanari, and Sellke [EMS22] use an algorithm obtained by the standard Euler discretization of the stochastic differential equation (SDE) defining stochastic localization. This roughly corresponds to replacing the sample $x \sim \tau_{w_i} \mu * \mathcal{N}(0, cI)$ in Algorithm 1 by $\mathcal{N}(\operatorname{mean}(\tau_{w_i} \mu), cI)$. Viewed as an approximate sample from $\tau_{w_i} \mu * \mathcal{N}(0, cI)$, this has a large approximation error – the entirety of $\tau_{w_i} \mu$ is replaced by its mean. This error can blow up after a constant number of steps. For the SK model, computing the mean itself can be done only approximately, so this large approximation error is somewhat unavoidable; indeed, approximate mean computation occupies the main technical bulk of their result, whereas in our case we have exact access to the Laplace transform of $\mu$, from which

the mean can be computed exactly. The large per-step approximation error is partly the reason that El Alaoui, Montanari, and Sellke [EMS22] only run their process for a constant number of steps and obtain a sample accurate within $o(n)$ Wasserstein distance of the SK model, a guarantee much weaker than total variation accuracy. The control of the approximation error in the work of El Alaoui, Montanari, and Sellke [EMS22] is specific to the SK model, so it is not clear if even the weaker guarantee of $o(n)$ Wasserstein accuracy can be obtained by running their Euler-discretized stochastic localization for $O(1)$ steps on the distributions of interest in this work; nevertheless, we obtain the much stronger total variation accuracy guarantee by using the high-accuracy discretization obtained by combining Algorithm 1 with the randomized midpoint method of Shen and Lee [SL19].

Parallelizing Markov chains, arguably the most widely used sampling tool, has been studied heavily [see LY22, and references within]. Two recent examples include the work of Feng, Hayes, and Yin [FHY21] for parallelizing Metropolis chains and the work of Liu and Yin [LY22] for parallelizing the more general class of single-site Markov chains including Glauber dynamics. The latter work shows how to obtain RNC-type sampling algorithms for distributions that have a bounded Dobrushin influence matrix, assuming the underlying Markov chain is single-site, that it changes one coordinate at a time, and that it mixes in nearly-linear time. These assumptions are unfortunately not satisfied by the distributions of interest in our work. Putting aside the Dobrushin influence matrix, single-site updates are not ergodic for our applications, even in the simplest case of a DPP, because the distribution can be supported on a slice of the hypercube $\{\pm 1\}^n$ [see, e.g., KT12]. Even worse, for other applications, we do not currently have Markov chains mixing in nearly-linear time with local moves, much more general than single-site updates. We remark that as a byproduct of our techniques, we show the existence of local Markov chains, with $O(1)$-site updates, mixing in nearly-linear time for just the case of a partition-constrained DPP. But even then, our construction is not explicit; we merely show the existence, and we cannot implement the moves of the chain by an efficient algorithm.

There has been recent interest in designing parallel sampling algorithms for distributions enjoying determinant-based counting. Anari, Hu, Saberi, and Schild [Ana+21] based on earlier work of Teng [Ten95] designed parallel samplers for spanning trees and more generally arborescences in directed graphs, that is spanning trees with directed edges oriented away from the root. Their work is based on parallelizing the classic algorithm of Broder [Bro89] and Aldous [Ald90] which extracts an arborescence from the trace of a random walk on the graph. Unfortunately, this algorithm is highly specific to arborescences, and there is no clear way of generalizing it to other distributions of interest in our work, all of which were stated as open problems by Anari, Hu, Saberi, and Schild [Ana+21]. Going beyond arborescences, there has been partial progress by Anari, Burgess, Tian, and Vuong [Ana+22a] for DPPs and their variants. Roughly speaking, they show how to sample from these distributions in $n^{1/2+\epsilon}$ time using $n^{O(1)}$ processors for any constant $\epsilon > 0$. While an improvement over the naive reduction from sampling to counting, which takes roughly linear parallel time, this is still far from the goal of $\log(n)^{O(1)}$ runtime. While we achieve this goal, for the specific case of non-symmetric DPPs, we use quasi-polynomially many processors, and so our result in that particular case is not strictly speaking comparable to their work.

The notion of transport-stability we define is somewhat reminiscent of transport-entropy inequalities [see, e.g., ELL17], which bound the transportation distance between two distributions as a function of their relative entropy. However, transport-entropy inequalities on the hypercube have a dependence on the dimension $n$, which is crucial for our applications to avoid. For example, consider the distribution $\mu$ which is uniform over $\{\pm 1\}^n$ and an exponential tilt of it by some standard basis vector. The relative entropy of these distributions is $\Theta(1)$, which using the standard transport-

entropy inequality only implies a Wasserstein distance of $O(\sqrt{n})$. We avoid this through transport stability and obtain an optimal Wasserstein distance of $O(1)$. As a side result, we show in Appendix A that versions of transport-stability are formally stronger than transport-entropy inequalities, by proving the existence of local Markov chains with optimal MLSI constants under transport-stability. For this, we follow almost identically the arguments Hermon and Salez [HS23] who argued the same for the more restricted class of strongly Rayleigh distributions.

### 1.3 Organization

We present the preliminaries in Section 2.

In Section 3 we prove the key lemmas needed to analyze Algorithm 1, namely Lemmas 5 to 7. The notion of transport-stability and how it controls the propagation of error in Algorithm 1 is deferred to Section 4; this is where we formally prove our main results Theorem 1 and Theorem 3. We then prove the transport-stability of DPPs and partition-constrained variants of them in Section 5, obtaining Corollary 2. In Section 6 we reduce the task of sampling Eulerian tours in digraphs to non-symmetric DPPs, proving Corollary 4.

Finally, as bonus results, in Appendix A we expand on the notion of transport-stability, and show by importing results of Hermon and Salez [HS23] that transport stability relates to strong concentration inequalities as well as the existence of fast mixing local Markov chains. Both of these were not known previously for partition-constrained DPPs.

### Acknowledgment

This work was supported by an NSF CAREER Award CCF-2045354 and a Sloan Research Fellowship. Thuy-Duong Vuong was supported by a Microsoft Research Ph.D. Fellowship. Tianyu Liu was supported by the NSF under Grant 2127309 to the Computing Research Association for the CIFellows Project. Yizhi Huang was supported by the Stanford UGVR program. Brian Xu and Katherine Yu were supported by the Stanford CURIS program.

## 2 Preliminaries

We use $[n]$ to denote $\{1, \ldots, n\}$. For a set $S \subseteq [n]$ we use $\mathbb{1}_S$ to denote the $n$-dimensional indicator vector of $S$. We identify $\{\pm 1\}^n$ with the family of subsets of $[n]$ by identifying a set $S$ with $2\mathbb{1}_S - \mathbb{1}_{[n]}$. We thus view distributions on $2^{[n]}$ alternatively as distributions on $\{\pm 1\}^n$. We use $I$ to denote the identity matrix. We use $*$ to denote the convolution of two distributions, i.e., the distribution of the sum of independent samples from them. We use $\tanh(x)$ to denote the hyperbolic tangent $(e^x - e^{-x})/(e^x + e^{-x})$. We use the fact that $\tanh$ is 1-Lipschitz, i.e., $|\tanh(x) - \tanh(y)| \leq |x - y|$.

We use $\|\cdot\|_p$ to denote the standard $\ell_p$ norm on $\mathbb{R}^n$. We let $\|\cdot\|_0$, a.k.a. $\ell_0$ denote the number of nonzero coordinates, which defines a metric.

For a distribution $\mu$ supported on $\mathbb{R}^n$ we use $\mathrm{mean}(\mu)$ to denote the mean $\mathbb{E}_{x\sim\mu}[x]$ and $\mathrm{cov}(\mu)$ to denote the covariance matrix $\mathbb{E}_{x\sim\mu}[xx^\intercal] - \mathbb{E}_{x\sim\mu}[x]\,\mathbb{E}_{x\sim\mu}[x]^\intercal$.

### 2.1 Wasserstein metric

For a metric space equipped with distance $d$, one can define the transport/Wasserstein distance induced by $d$ on the space of probability distributions supported on the metric space.

**Definition 8** (Wasserstein distance)**.** The Wasserstein $p$-distance with respect to $d$ between two distributions $\mu, \nu$ supported on a metric space $(\Omega, d)$ is defined as

$$\mathcal{W}_p(\mu,\nu) \coloneqq \inf\left\{\mathbb{E}_{(x,y)\sim\pi}[d(x,y)^p]^{1/p} \;\middle|\; \pi \text{ is a coupling of } \mu,\nu\right\}.$$

When not specified $\mathcal{W}$, is taken to mean $\mathcal{W}_1$. Note that $\mathcal{W}_p$ is monotonically increasing in $p$.

Note that for finite $\Omega$, by compactness of the space of couplings, all $\mathcal{W}_p$, including $\mathcal{W}_\infty$ are well-defined. In particular $\mathcal{W}_\infty(\mu,\nu)$ is the minimum threshold $t$ such that there is a coupling between $\mu, \nu$ only supported on pairs $(x, y)$ with $d(x, y) \le t$.

The metric $d$ used in the definition of $\mathcal{W}_p$ has to be inferred from context. By default, for distributions supported on the hypercube $\{\pm1\}^n$ or more generally $\mathbb{R}^n$ we take $d$ to be the metric induced by the $\ell_1$ norm; in the hypercube case, this is the same as the Hamming distance up to a factor of 2. We use the notation $\mathcal{W}_{p,q}$ to denote the Wasserstein $p$-distance induced by the $\ell_q$ norm, and we specify $q$ when it is not equal 1 to avoid confusion.

We can translate between $\mathcal{W}_{p,q}$ for different $q$ by standard comparisons between $\ell_q$ metrics. We use the following simple fact.

**Proposition 9.** *For any distributions $\mu, \nu$ supported on $\mathbb{R}^n$,*

$$\mathcal{W}_{2,1}(\mu,\nu) \le \sqrt{n}\,\mathcal{W}_{2,2}(\mu,\nu).$$

*Proof.* This follows directly from the fact that for any $x \in \mathbb{R}^n$ we have $\|x\|_1 \le \sqrt{n}\|x\|_2$. □

We note that Wasserstein distance between two distributions also upperbounds the distance between their means by the triangle inequality:

**Proposition 10.** *We always have*

$$\|\mathrm{mean}(\mu) - \mathrm{mean}(\nu)\|_q \le \mathcal{W}_{1,q}(\mu,\nu).$$

## 2.2 Tilts, logarithmic Laplace transform, and semi-log-concavity

We define the notion of an exponential tilt, which is also known as an external field.

**Definition 11.** We define $\tau_w$ to be the operator that maps a distribution $\mu$ on $\{\pm1\}^n$ to the distribution $\tau_w\mu$ defined by

$$\tau_w\mu(x) \propto \exp(\langle w, x\rangle)\mu(x).$$

We call the normalizing factor in this definition $\sum_x \exp(\langle w, x\rangle)\mu(x)$ viewed as a function of $w$ the Laplace transform of $\mu$. The logarithm of the Laplace transform is more convenient for calculations and we follow the lead of Eldan and Shamir [ES22] and use $\mathcal{L}_\mu$ to denote it.

**Definition 12** (Logarithmic Laplace transform)**.** We define the logarithmic Laplace transform of a distribution $\mu$ supported on $\{\pm1\}^n$ as

$$\mathcal{L}_\mu(w) = \log\left(\sum_x \exp(\langle w, x\rangle)\mu(x)\right).$$

We use the following fact about derivatives of $\mathcal{L}_\mu$.

**Proposition 13** ([ES22]). *We have*

$$\nabla \mathcal{L}_\mu(w) = \mathrm{mean}(\tau_w \mu),$$
$$\nabla^2 \mathcal{L}_\mu(w) = \mathrm{cov}(\tau_w \mu).$$

Finally we state the definition of semi-log-concavity [ES22].

**Definition 14** ([ES22]). A distribution $\mu$ is called $\beta$-semi-log-concave when

$$\nabla^2 \mathcal{L}_\mu(w) \preceq \beta I.$$

When $\beta$ is omitted, semi-log-concave means $O(1)$-semi-log-concavity.

Note that $\mathcal{L}_\mu$ is convex, as its Hessian is a covariance matrix which is $\succeq 0$. Consequently, $\beta$-semi-log-concavity is equivalent to $\nabla \mathcal{L}_\mu$ being $\beta$-Lipschitz w.r.t. the $\ell_2$ norm:

$$\|\mathrm{mean}(\tau_w \mu) - \mathrm{mean}(\tau_{w'} \mu)\|_2 = \|\nabla \mathcal{L}_\mu(w) - \nabla \mathcal{L}_\mu(w')\|_2 \le \beta \|w - w'\|_2.$$

Finally, we note that $\beta$-Lipschitzness in any other $\ell_q$ norm is stronger than $\beta$-semi-log-concavity. This is because $\nabla^2 \mathcal{L}_\mu$ is a symmetric PSD matrix, and hence its $\ell_q$ to $\ell_q$ matrix norm is lowerbounded by its maximum eigenvalue, which equals its $\ell_2$ to $\ell_2$ norm.

**Proposition 15.** *If we have*

$$\|\mathrm{mean}(\tau_w \mu) - \mathrm{mean}(\tau'_w \mu)\|_q \le \beta \|w - w'\|_q$$

*for any $q \ge 1$, then $\mu$ is $\beta$-semi-log-concave.*

### 2.3 Determinantal point processes

A determinantal point process (DPP) is a probability distribution over subsets $S \subseteq [n]$. It is parameterized by a matrix $L \in \mathbb{R}^{n \times n}$ with

$$\mathbb{P}[S] \propto \det(L_{S,S}),$$

with $L_{S,S}$ being the principal submatrix whose columns and rows are indexed by $S$. We call $L$ the ensemble matrix. Note that we need $\det(L_{S,S}) \ge 0$ for all $S$ for this definition to work. This is satisfied by symmetric PSD $L$, which yield the traditional (symmetric) DPPs [KT12], and more generally for any $L$ whose symmetrization is PSD, that is $L + L^\intercal \succeq 0$, which are called non-symmetric DPPs [Gar+19].

Given a cardinality $k \in \mathbb{N}$, the $k$-DPP parameterized by $L$ is a distribution over subsets $S$ of size $k$, defined by conditioning the samples from the DPP to have size $k$.

More generally, consider a matrix $L \in \mathbb{R}^{n \times n}$, and a partition $V_1 \sqcup \cdots \sqcup V_r = [n]$ of the ground set, and a tuple $c \in \mathbb{N}^r$ of integers. The partition-constrained DPP, $\mu_{L;V,c} : 2^{[n]} \to \mathbb{R}_{\ge 0}$, is defined by conditioning the DPP formed by $L$ to only the sets $S$ which have $|S \cap V_i| = c_i$ for all $i$.

For any $Y \subseteq [n]$, if we condition the DPP defined by $L$ on the event that items in $Y$ are included in the sample, we still get a DPP; the new ensemble matrix is given by the Schur complement

$L^Y = L_{\tilde{Y}} - L_{\tilde{Y},Y} L_{Y,Y}^{-1} L_{Y,\tilde{Y}}$ where $\tilde{Y} = [n] \setminus Y$. Similarly, excluding items corresponds to deleting the corresponding rows/columns from $L$.

Similar statements hold for partition-constrained DPPs. Conditioning $\mu_{L;V,c}$ on $Y$ being included in the set results in a partition-constrained DPP $\mu_{L^Y;V',c'}$ with ensemble matrix $L^Y$ and partition $V_1' \sqcup \cdots \sqcup V_r' = [n] \setminus Y$ with $V_i' = V_i \setminus Y$, and $c_i' = c_i - |V_i \cap Y|$.

**Proposition 16.** *DPP (respectively partition-constrained DPPs) $\mu$, are closed under exponential tilts.*

*Proof.* Identified with distribution s on $\{\pm 1\}^n$, an exponential tilt $\tau_w \mu$ is the same as the DPP (respectively partition-constrained DPP) defined by the matrix

$$DLD,$$

where $D$ is a diagonal matrix whose $i$-th entry is $\exp(w_i)$. This is because for any set $S$, we have

$$\det((DLD)_{S,S}) = \det(L_{S,S}) \prod_{i \in S} \exp(2w_i) \propto \det(L_{S,S}) \frac{\prod_{i \in S} \exp(w_i)}{\prod_{i \notin S} \exp(w_i)} = \exp(\langle 2\mathbb{1}_S - \mathbb{1}_{[n]}, w \rangle) \det(L_{S,S}).$$

Note that mapping $L$ to $DLD$ preserves being PSD, and/or symmetric. □

**Proposition 17.** *There is an* NC *algorithm for computing the partition function $\sum_S \det(L_{S,S})$ of DPPs and partition-constrained DPPs with $O(1)$ parts.*

*Proof.* For an unconstrained DPP, note that its partition function can be written as

$$\sum_S \det(L_{S,S}) = \det(L + I),$$

hence by the result of Csanky [Csa76], this can be computed in NC.

For partition-constrained DPPs, Celis, Keswani, Straszak, Deshpande, Kathuria, and Vishnoi [Cel+18] show how to compute the partition function by polynomial interpolation. Essentially $r$ variables corresponding to parts are constructed and we compute the determinant of $L + D$, where $D$ is a diagonal matrix with one of the $r$ variables on each entry. Our goal is to compute the coefficient of a certain monomial in this $r$-variate polynomial. This can be done by evaluating $\det(L + D)$ at $\simeq n^r$ places, and using polynomial interpolation to recover the coefficients. All of this can be done in NC assuming $r = O(1)$. □

By combining the previous propositions, we see that computing the Laplace transform of any symmetric/non-symmetric DPP or a partition-constrained variant can be done in NC.

### 2.4 Determinants and stability

We recall the notions of sector stability and fractional log-concavity [Ali+21].

**Definition 18** ([Ali+21])**.** For an open subset $U \subseteq \mathbb{C}^n$, we call a polynomial $g \in \mathbb{C}[z_1, \ldots, z_n]$ $U$-stable iff

$$(z_1, \ldots, z_n) \in U \implies g(z_1, \ldots, z_n) \neq 0.$$

We also call the identically $0$ polynomial $U$-stable. This ensures that limits of $U$-stable polynomials are $U$-stable. For convenience, when $n$ is clear from context, we abbreviate stability w.r.t. regions of the form $U \times U \times \cdots \times U$ where $U \subseteq \mathbb{C}$ simply as $U$-stability.

A set-valued distribution $\mu$ is $U$-stable iff its generating polynomial $g_\mu$ is $U$-stable:

$$g_\mu(z_1, \ldots, z_n) = \sum_S \mu(x) \prod_{i \in S} z_i.$$

A distribution is real-stable or strongly Rayleigh if its generating polynomial $g_\mu$ is $\mathbb{H}$-stable where $\mathbb{H} := \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ is the upper half of the complex plane. Symmetric DPPs and $k$-DPPs are strongly Rayleigh [BBL09].

More generally Alimohammadi, Anari, Shiragur, and Vuong [Ali+21] showed that partition-constrained symmetric DPPs with $O(1)$ parts, non-symmetric DPPs, and non-symmetric $k$-DPPs are stable w.r.t. a sector $\{\exp(r + i\theta) \mid r \in \mathbb{R}, \theta \in [-\alpha\pi, +\alpha\pi]\}$ for some constant $\alpha = \Omega(1)$. This property is called sector-stability. We remark that real-stability, although defined via the upper half-plane, also implies stability w.r.t. the right half-plane, i.e., the stability w.r.t. the sector with $\alpha = 1/2$ [Ali+21].

**Proposition 19.** *Any strongly Rayleigh or sector-stable distribution, including all aforementioned determinantal point processes and their variants, are closed under exponential tilts and are spectrally independent [Ali+21]. It follows by the results of Eldan and Shamir [ES22], that these distributions, viewed on the hypercube, are semi-log-concave.*

This shows that we can directly apply Theorem 3 to all of these distributions. However, for symmetric DPPs and their partition-constrained variants, we show a stronger result, that we can apply Theorem 1 to them. We use the following fact in the proof, which appears in Section 5.

**Lemma 20.** *Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be strongly Rayleigh and let $V_1 \sqcup \cdots \sqcup V_r = [n]$ be a partition of $[n]$ and let $c_1, \ldots, c_r \in \mathbb{N}$ satisfy $\sum_i c_i = k$. Let $\mu_{V,c}$ be the distribution defined by conditioning $\mu$ to the sets that have exactly $c_i$ elements of $V_i$.*

*Let $(\mu_{V,c})^{V_{j^*}}$ be the projection of $\mu_{V,c}$ on $V_{j*}$, i.e., for $S_{j^*} \subseteq V_{j^*}$,*

$$(\mu_{V,c})^{V_{j^*}}(S_{j^*}) = \sum_{S : S \cap V_{j^*} = S_{j^*}} \mu_{V,c}(S).$$

*Then $(\mu_{V,c})^{V_i}$ is strongly Rayleigh.*

*Proof.* W.l.o.g., assume $j^* = r$ and $V_r = \{1, \ldots, c_r\}$. Consider the generating polynomial $g_\mu(z_1, \ldots, z_n) = \sum \mu(S) z^S$. We obtain a new real-stable homogeneous polynomial $h(z_1, \ldots, z_{c_r}, y_1, \ldots, y_{r-1})$ by the following substitution: for $j \in [r-1]$, set $z_i = y_j$ for all $i \in V_j$. Taking partial derivatives and setting variables to 0 preserves real-stability [Brä07, Proposition 3.1], thus

$$\frac{\partial^{c_1}}{\partial y_1^{c_1}} \frac{\partial^{c_2}}{\partial y_2^{c_2}} \cdots \frac{\partial^{c_{r-1}}}{\partial y_{r-1}^{c_{r-1}}} h|_{y_2 = \cdots = y_r = 0}$$

is real stable, and this is exactly the generating polynomial of $(\mu_{V,c})^{V_{j^*}}$. □

**Lemma 21.** *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be real stable and $1 \leq v_1 \leq v_2 \leq \cdots \leq v_r \leq n$ and $(c_1, c_2, \ldots, c_r) \in \{0, 1\}$. Let $\mu_{|V,c}$ be the distribution $[n] \setminus \{v_1, v_2, \ldots, v_r\} \to \mathbb{R}_{\geq 0}$ defined by $\mu_{|V,c}(S) \propto \mu(S \cup \{v_i \mid c_i = 1\})$. Then $\mu_{|V,c}$ is strongly Rayleigh.*

*Proof.* Consider the generating polynomial $g_\mu(z_1, \ldots, z_n) = \sum \mu(S) z^S$, which is real-stable. By substituting $z_{v_i} = c_i$ for all $1 \leq i \leq r$, we obtain a real-stable polynomial that is exactly the generating polynomial of $\mu_{|V,c}$. □

## 3 Algorithmic framework

In this section, we prove Lemmas 5 to 7, the key facts behind the correctness of Algorithm 1. The analysis of the approximation error resulting from continuous sampling will be done after we define the notion of transport-stability in Section 4.

*Proof of Lemma 5.* First, note that for a distribution $\mu$ on the hypercube, the density of $\nu = \mu * \mathcal{N}(0, cI)$ at a point $y$ is up to a global multiplier given by

$$\sum_{x\in\{\pm 1\}^n} \mu(x)\exp(-\|y-x\|^2/2c) \propto \exp(-\|y\|^2/2c)\cdot \sum_x \mu(x)\exp(\langle y, x\rangle/c) = \exp\big(\mathcal{L}_\mu(y/c) - \|y\|^2/2c\big),$$

where we used the fact that for $x \in \{\pm 1\}^n$, $\|x\|^2$ is constant and disappears as a constant of proportionality. As a result

$$\nabla^2 \log \nu = \frac{1}{c^2}\cdot \nabla^2 \mathcal{L}_\mu(y/c) - \frac{I}{c} = \frac{\mathrm{cov}(\tau_{y/c}\mu)}{c^2} - \frac{I}{c}.$$

Note that as $y$ ranges over $\mathbb{R}^n$, $y/c$ also ranges over all of $\mathbb{R}^n$. Now if $\mu$ is $\beta$-semi-log-concave, see Definition 14, then

$$-I/c \preceq \nabla^2 \log \nu \preceq (\beta/c^2 - 1/c)I$$

For $\nu$ to be log-concave we simply need to set $c \geq \beta$. For a slightly larger value, say $c = 2\beta$, we get

$$-I/2\beta \preceq \nabla^2 \log \nu \preceq -I/4\beta,$$

which means that $\nabla^2 \log \nu$ is well-conditioned too, with a condition number of 2.

We proved that $\beta$-semi-log-concavity implies that $\nu = \mu * \mathcal{N}(0, cI)$ is well-conditioned log-concave for $c = \Omega(\beta)$. Next, note that even log-concavity of $\nu$ means that $\mathrm{cov}(\tau_w\mu)$ is bounded by $c$ for *all* $w$ because $y/c$ ranges over all of $\mathbb{R}^n$. This proves that semi-log-concavity and well-conditioned log-concavity of convolutions with Gaussians are equivalent.

Finally, note that from Definition 14 and Proposition 13, a distribution $\mu$ is semi-log-concave iff all tilts of it $\tau_w\mu$ are semi-log-concave. This is because tilts are an additive group action: $\tau_w\tau'_w = \tau'_w\tau_w = \tau_{w+w'}$. This shows that semi-log-concavity of $\mu$ implies the convolution of its tilts $\tau_w\mu$ with Gaussians are also well-conditioned log-concave for large enough $c$. □

*Proof of Lemma 6.* First, note that the logarithmic Laplace transform for $\tau_w\mu$ is the same as the logarithmic Laplace transform for $\mu$ shifted by the vector $w$ (up to a global additive constant):

$$\mathcal{L}_{\tau_w\mu}(y) = \mathcal{L}_\mu(y+w) - \mathcal{L}_\mu(w).$$

So having access to an oracle for $\mu$ is equivalent to having access to an oracle for $\tau_w\mu$. Therefore, we only need to prove the density and its gradient can be computed for $\mu$; it automatically translates to tilts $\tau_w\mu$.

In the proof of Lemma 5, we computed the density of $\mu * \mathcal{N}(0, cI)$ up to a global normalizing constant:

$$\exp(\mathcal{L}_\mu(y/c) - \|y\|^2/2c) = \exp(\mathcal{L}_\mu(y/c))\cdot\exp(-\|y\|^2/2c).$$

This first term is the Laplace transform of $\mu$ evaluated at $y/c$. Hence given oracle access to the Laplace transform of $\mu$, we can compute this expression. We remark that if one wishes, the normalized

density can also be computed by carrying through the normalizing constant for a Gaussian, as well as the global factor we dropped in the calculation in the proof of Lemma 5. However, continuous sampling algorithms only need the density up to a normalizing factor.

Next we compute the gradient $\nabla \log \nu$, which by the rules of calculus also gives us $\nabla \nu$ (continuous sampling algorithms actually use the former). This gradient is equal to

$$\frac{1}{c}\nabla \mathcal{L}_\mu(y/c) - \frac{y}{c} = \frac{\mathrm{mean}(\tau_{y/c}\mu)}{c} - \frac{y}{c}$$

So we just need to show how to compute the mean of tilts of $\mu$. Again by the equivalence of Laplace transform oracles for $\mu$ and its tilts, it is enough to show how to compute $\mathrm{mean}(\mu)$. Let $\mathbb{1}_i$ denote the $i$-th element of the standard basis. Then we have

$$\mathcal{L}_\mu(\mathbb{1}_i) - \mathcal{L}_\mu(0) = \mathbb{E}_{x\sim\mu}[\exp\langle x, \mathbb{1}_i\rangle] = p_i e - (1-p_i)e^{-1} = p_i(e - e^{-1}) - e^{-1},$$

where $p_i = \mathbb{P}_{x\sim\mu}[\langle x, \mathbb{1}_i\rangle = 1]$. Therefore $p_i$ can be computed as a simple linear function of the above quantity. Finally note that $\mathrm{mean}(\mu) = (2p_1 - 1, 2p_2 - 1, \ldots, 2p_n - 1)$ and that all of these entries can be computed in parallel. □

It remains to prove Lemma 7. We first prove the following characterization of the distribution of $w_T$ that is an immediate consequence of the characterization of stochastic localization by El Alaoui and Montanari [EM22].

**Lemma 22.** *If convolutions are sampled exactly in Algorithm 1, then*

$$cw_T/T \sim \mu * \mathcal{N}(0, cI/T).$$

*Proof.* While an elementary proof of this fact is easy to derive by induction on $T$, we appeal to known results on stochastic localization, which is how we developed Algorithm 1.

Stochastic localization [see ES22] is a measure-valued stochastic process $\{\mu_t\}_{t=0}^\infty$ defined for a starting measure $\mu$, that satisfies a stochastic differential equation (SDE). Denoting by $W_t$ a standard Brownian motion, and by $F_t$ the density of $\mu_t$ w.r.t. $\mu$ (that is $\mu_t(x) = F_t(x)\cdot\mu(x)$), the SDE is described by

$$dF_t(x) = F_t(x)\cdot\langle x - \mathrm{mean}(\mu_t), dW_t\rangle.$$

This process results in a martingale $F_t(x)$ for every $x$ [ES22], hence the distribution of $\mu_t$ for every time $t$ can be seen as a *decomposition* of $\mu_0$. Note that this SDE is time-homogeneous, that is its evolution rule does not depend on time $t$. So running stochastic localization up to time $t$, and then running an independent stochastic localization for $s$ time, results in a measure identically distributed as if we had run stochastic localization for $s+t$ time from the start.

El Alaoui and Montanari [EM22] proved that $\mu_t$ in stochastic localization is identically distributed to the following random measure: First sample $z \sim \mu$ and $g \sim \mathcal{N}(0, tI)$ and let $y_t = tz + g$. Then form the distribution

$$\widetilde{\mu}_t(x) \propto \exp\left(\langle y_t, x\rangle - \frac{t}{2}\|x\|^2\right)\mu(x),$$

by normalizing the above appropriately. Then $\widetilde{\mu}_t$ is identically distributed as $\mu_t$ (over the randomness of $z$ and $g$).

In our case, since the distribution $\mu$ is supported on $\{\pm 1\}^n$ and for $x$ on the hypercube $\|x\|^2 = n$ is constant, we can define

$$\widetilde{\mu}_t(x) \propto \exp(\langle y_t, x\rangle)\mu(x),$$

which means that $\widetilde{\mu}_t$ is the same as $\tau_{y_t}\mu$.

We claim that for for every $i$, the distribution of $w_i$ is identical to $y_{i/c}$ and thus $\tau_{w_i}\mu$ is identically distributed as $\mu_{i/c}$. If proved, plugging in $i = T$, shows that $cw_T/T$ is distributed as $y_t/t$ for $t = T/c$ which is distributed as $\mu * \mathcal{N}(0, I/t)$, proving the statement of the lemma.

Now we prove the claim by induction. Note that $w_1$ is obtained by scaling a sample from $\mu * \mathcal{N}(0, cI)$ by $1/c$, so its distribution is identical to $y_{1/c}$. Therefore $\tau_{w_1}\mu$ is also distributed as $\mu_{1/c}$. This is what we replace $\mu$ by in the next iteration of the loop in Algorithm 1. So the distribution of $w_2 - w_1$ conditioned on $w_1$ is identical to the tilt applied by stochastic localization to go from $\mu_{1/c}$ to $\mu_{2/c}$, which means $\tau_{w_2}\mu$ is identically distributed as $\mu_{2/c}$, and so on. □

Now that we know the distribution of $w_T$, it is easy to prove Lemma 7.

*Proof of Lemma 7.* Note that $cw_T/T$ is distributed as $x + g$ where $x \sim \mu$ and $g \sim \mathcal{N}(0, cI/T)$. If $\|g\|_\infty < 1$, then because $x \in \{\pm 1\}^n$, the signs of coordinates in $x$ and $x + g$ will be the same. So we can bound the total variation distance between $\operatorname{sign}(w_T)$ and $\mu$ by

$$\mathbb{P}[\|g\|_\infty \geq 1].$$

Since $g$ is a vector of $n$ centered normals, each of variance $c/T$, by the sub-Gaussian tail inequality we have

$$\mathbb{P}[|g_i| \geq 1] \leq O(1) \cdot \exp(-\Omega(1/(c/T))) = O(1) \cdot \exp(-\Omega(T/c)).$$

By a union bound we have

$$\mathbb{P}[\|g\|_\infty \geq 1] \leq O(n) \cdot \exp(-\Omega(T/c)).$$

For sufficiently large $T/c \simeq \Theta(\log(n/\epsilon))$, this is below $\epsilon$. □

## 4 Transport stability

In this section, we define the notion of transport stability and show how to control the propagation of the approximation error resulting from continuous sampling in Algorithm 1.

**Definition 23** (Transport stability)**.** For a choice of transportation metric $d_1$ on probability distributions supported on $\{\pm 1\}^n$, and another metric $d_2$ on $\mathbb{R}^n$, we say that $\mu$ has $d_1$-$d_2$ transport stability, when

$$d_1(\tau_w\mu, \tau_{w'}\mu) \leq d_2(w, w'),$$

for all $w, w'$.

Our default choices for $d_1$ and $d_2$ are the $\mathcal{W}_1$ Wasserstein distance and $O(1) \cdot \ell_1$ respectively, and we call this simply transport stability. For DPPs and partition-constrained DPPs, we prove $\mathcal{W}_\infty$-$O(1) \cdot \ell_0$ transport-stability, which we now show is formally stronger. Note that both $\ell_0$ and $\ell_1$ are summable over coordinates. In other words, if we start from $w, w'$ and define $w_i$ to be identical in the first $i$ coordinates with $w$ and in the remaining coordinates with $w'$, then

$$d(w, w') = d(w_0, w_1) + d(w_1, w_2) + \cdots + d(w_{n-1}, w_n),$$

for $d \in \{\ell_0, \ell_1\}$. If follows that to check transport stability with $\ell_0$ or $\ell_1$ (or scalar multiples of them) as the choice of distance on $\mathbb{R}^n$, it is enough to check the inequality in Definition 23 only for $w, w'$ that differ in one coordinate: $\|w - w'\|_0 = 1$. We prove the $\mathcal{W}_1$-$O(1)\ell_1$ transport stability of DPPs and partition-constrained DPPs by a much stronger $\mathcal{W}_\infty$-$O(1)\ell_0$ inequality.

**Lemma 24.** *If the distribution $\mu$ is $\mathcal{W}_\infty$-$\alpha\ell_0$ transport-stable, then it is also $\mathcal{W}_1$-$O(\alpha)\ell_1$ transport-stable.*

*Proof.* Note that because $\mathcal{W}_1$ is always bounded by $\mathcal{W}_\infty$, from the assumption we can conclude $\mathcal{W}_1$-$\alpha\ell_0$ transport-stability. We now show how to prove $\mathcal{W}_1$-$O(\alpha)\ell_1$ stability. By the preceding arguments it is enough to prove that for $w, w'$ differing in exactly one coordinate, we have

$$\mathcal{W}_1(\tau_w\mu, \tau_{w'}\mu) \le O(\alpha) \cdot \|w - w'\|_1.$$

Let $w' - w = c\mathbb{1}_i$, where $i$ is the differing coordinate and assume w.l.o.g. that $c > 0$, which means $\|w' - w\|_1 = c$. We construct a coupling to bound the $\mathcal{W}_1$ distance between the two. Write $\tau_w\mu$ as $p\nu_1 + (1-p)\nu_2$ where $\nu_1$ is the conditioning on coordinate $i$ being $+1$ and $\nu_2$ is the conditioning on coordinate $i$ being $-1$. Notice that we have $\tau_{w'}\mu = q\nu_1 + (1-q)\nu_2$. This is because to go from $\tau_w\mu$ to $\tau_{w'}\mu$, we only have to apply a tilt in the direction of coordinate $i$, which does not change the two conditional distributions. By the same token, $\nu_1$ can be obtained as the limit of $\tau_{w+\beta\mathbb{1}_i}\mu$ for $\beta \to \infty$ and $\nu_2$ can be obtained as the limit when $\beta \to -\infty$. Because of $\mathcal{W}_1$-$\alpha\ell_0$ transport-stability we can conclude that $\mathcal{W}_1(\nu_1, \nu_2) \le \alpha$.

Next, we claim that $|p - q| \le c/2$. The equation defining the relationship between $p, q$ is that $q, 1-q \propto pe^c, (1-p)e^{-c}$ respectively. Let $\theta_p, \theta_q$ be such that $\tanh(\theta_p) = 2p - 1 \in (-1, 1)$ and $\tanh(\theta_q) = 2q - 1 \in (-1, 1)$. In other words, $p, 1-p \propto e^{\theta_p}, e^{-\theta_p}$ and $q, 1-q \propto e^{\theta_q}, e^{-\theta_q}$. Then we have $\theta_q = \theta_p + c$. It follows that $2|p - q| = |\tanh(p) - \tanh(q)| \le c$.

Now we construct the coupling for $\tau_w\mu$ and $\tau_{w'}\mu$. Let $\pi$ be the coupling between $\nu_1, \nu_2$ realizing $\mathcal{W}_1(\nu_1, \nu_2)$. We sample a pair $(x, y)$ as follows:

- With probability $p$, we sample $z \sim \nu_1$ and output $(z, z)$,
- With probability $1 - q$, we sample $z \sim \nu_2$ and output $(z, z)$,
- With the remaining probability $q - p = |p - q|$, we sample $(x, y) \sim \pi$ and output $(y, x)$.

It is easy to see that this is a coupling for $\tau_w\mu$ and $\tau_{w'}\mu$. Moreover, in all but the last case we output identical points which have distance $0$. Therefore we have

$$\mathbb{E}[\|x - y\|_1] = |p - q| \cdot \mathcal{W}_1(\nu_1, \nu_2) \le O(\alpha c). \qquad \square$$

**Corollary 25.** *Every distribution $\mu$ on $\{\pm 1\}^n$ is $\mathcal{W}_1$-$O(n)\ell_1$ transport-stable.*

*Proof.* Since $\mathcal{W}_\infty$ is uniformly bounded by $O(n)$ and $\ell_0$ is lower bounded by $1$ for distinct points, we automatically have $\mathcal{W}_\infty$-$O(n)\ell_0$ transport-stability. The conclusion follows from Lemma 24. $\square$

We remark that showing $\mathcal{W}_\infty$-$\alpha\ell_0$ transport-stability is basically equivalent to constructing couplings between coordinate-conditioned distributions.

**Proposition 26.** *A distribution $\mu$ has $W_\infty$-$\alpha\ell_0$ transport-stability iff for every $\nu \coloneqq \tau_w\mu$ and every coordinate $i \in [n]$, the distributions $\nu_1, \nu_2$ obtained by conditioning $x \sim \nu$ to have $x_i = 1$ and $x_i = -1$ respectively, satisfy*

$$\mathcal{W}_\infty(\nu_1, \nu_2) \le \alpha.$$

*Proof.* The proof is almost identical to the proof of Lemma 24. Given a coupling between $\nu_1, \nu_2$, this can be extended to a coupling between $\tau_{w+c\mathbb{1}_i}\mu$ and $\tau_{w+c'\mathbb{1}_i}\mu$ for any $c, c' \in \mathbb{R}$; no distance increases, so $\mathcal{W}_\infty$ still remains bounded by $\alpha$. □

We are now ready to prove our main statement about how errors propagate in Algorithm 1.

**Proposition 27.** *Suppose that $\mu$ is $\mathcal{W}_1$-$\alpha\ell_1$ transport-stable. Then if we run Algorithm 1 using an approximate continuous sampling algorithm that guarantees accuracy within $\delta$ in $\mathcal{W}_1$, resulting in a random variable $w_T$, then*

$$\mathcal{W}_1(cw_T/T, \mu * \mathcal{N}(0, cI/T)) \le \delta \cdot (1 + \alpha/c)^T.$$

*Proof.* Suppose we have two executions of Algorithm 1: one using the approximate continuous sampling algorithm resulting in $w_0, \dots, w_T$, and one using exact samples resulting in $w'_0, \dots, w'_T$. We bound the $\mathcal{W}_1$ distance between the random variables $w_i$ and $w'_i$ inductively. For $i = 0$ this distance is 0. We claim that

$$\mathcal{W}_1(w_i, w'_i) \le C_i := \frac{\delta}{c} \cdot (1 + \beta + \cdots + \beta^{i-1}),$$

where $\beta := 1 + \alpha/c$.

Assuming this is true for $i$, we couple $w_i$ and $w'_i$ in a way that realizes this $\mathcal{W}_1$ distance. In other words

$$\mathbb{E}[\|w_i - w'_i\|_1] \le C_i.$$

Now we couple exact samples $\widetilde{x} \sim \tau_{w_i}\mu * \mathcal{N}(0, cI)$ and $x' \sim \tau_{w'_i}\mu * \mathcal{N}(0, cI)$ by using the coupling implied by transport-stability for $\tau_{w_i}\mu$ and $\tau_{w'_i}\mu$ and adding the same independent Gaussian sample from $\mathcal{N}(0, cI)$ to both. It follows that

$$\mathcal{W}_1(\widetilde{x}, x') \le \mathbb{E}[\mathcal{W}_1(\tau_{w_i}\mu, \tau_{w'_i}\mu)] \le \alpha\, \mathbb{E}[\|w_i - w'_i\|_1] \le \alpha C_i.$$

However, note that we do not sample $\widetilde{x}$ exactly in the algorithm, but rather we sample $x$ approximately within $\mathcal{W}_1 \le \delta$ distance of $\widetilde{x}$. By the triangle inequality for $\mathcal{W}_1$ we have

$$\mathcal{W}_1(x, x') \le \delta + \alpha C_i.$$

Finally note that $w_{i+1} = w_i + x/c$ and $w'_{i+1} = w'_i + x'/c$, so we get

$$\mathbb{E}[\|w_{i+1} - w'_{i+1}\|_1] \le \mathbb{E}[\|w_i - w'_i\|_1] + \frac{\mathbb{E}[\|x - x'\|]}{c} \le C_i + (\delta + \alpha C_i)/c = (1 + \alpha/c)C_i + \delta/c = C_{i+1}.$$

This finishes the induction.

Now we simply note that our bound for $i = T$ can be further upper-bounded by $(\delta/c) \cdot T(1 + \alpha/c)^T$. This implies that

$$\mathcal{W}_1(cw_T/T, cw'_T/T) \le \delta(1 + \alpha/c)^T.$$

But we know by Lemma 22 that $cw'_T/T$ is distributed according to $\mu * \mathcal{N}(0, cI/T)$, finishing the proof. □

Next, we state the error guarantee of the approximate continuous sampler we utilize, the randomized midpoint method of Shen and Lee [SL19].

**Proposition 28.** *Suppose that $\mu$ is a semi-log-concave distribution and $c$ is a large constant so that $\nu := \tau_w \mu * \mathcal{N}(0, cI)$ is well-conditioned log-concave, and assume we have oracle access to the Laplace transform of $\mu$. Then we can output a sample that is $\delta$-accurate in $\mathcal{W}_1$ distance from $\nu$ with runtime $\log(n/\delta)^{O(1)}$ using $(n/\delta)^{O(1)}$ processors.*

*Proof.* Since we have oracle access to the Laplace transform of $\mu$, we have oracle access to the log-density and gradients of log-density for $\nu := \tau_w \mu * \mathcal{N}(0, cI)$ by Lemma 6. Note that $\nabla^2 \log \nu$ is sandwiched between $-\alpha I, -\beta I$ for some constants $\alpha, \beta = \Theta(1)$ by the proof of Lemma 5. This means the condition number $\kappa = \alpha/\beta$ is also $O(1)$.

The result of Shen and Lee [SL19] thus allows us to sample from $\nu$ within $\epsilon \cdot D$ error in $\mathcal{W}_{2,2}$ where $D = \sqrt{n/\beta}$ is the effective diameter. This algorithm uses $(\kappa/\epsilon)^{O(1)}$ processors and runs in time $O(\kappa \log(1/\epsilon))$. Note that $\mathcal{W}_1 \le \sqrt{n}\,\mathcal{W}_{1,2} \le \sqrt{n}\,\mathcal{W}_{2,2}$. So by setting $\epsilon = \sqrt{\beta}\delta/n$ we get our desired $\delta$ accuracy in the $\mathcal{W}_1$ metric. With this setting of parameters it is easy to check that the parallel runtime is $O(\log(1/\epsilon)) = O(\log(n/\delta))$ and that the number of processors is $O(1/\epsilon)^{O(1)} = (n/\delta)^{O(1)}$. □

*Remark* 29. Technically, to start the algorithm of Shen and Lee [SL19], one needs an approximate maximizer of the density $\nu$. Since $\log \nu$ is a well-conditioned concave function, gradient ascent converges exponentially fast to the maximizer, so we can start from a point and run gradient ascent for polylogarithmically many steps, and then run the algorithm of Shen and Lee [SL19]. To make sure the number of steps taken is only $\log(n)^{O(1)}$ we have to find an initial point $x$ which has at least $\exp(\exp(-O(\log n^{O(1)})))$ fraction of the maximum density. The point $0$ in our case has this property. Since we are convolving a distribution on $\{\pm 1\}^n$ with a Gaussian $\mathcal{N}(0, cI)$, the density of the result at $0$ compared to the maximum possible density is at least

$$\exp(-\sqrt{n}^2/2c)/\exp(0) = \exp(-n/2c),$$

because the distance of $0$ from all vertices of the hypercube is $\sqrt{n}$.

We can now prove Theorem 1.

*Proof of Theorem 1.* First note that by Propositions 10 and 15, transport-stability implies semi-log-concavity. Combining this with Lemma 5, we get that for $c$ a large enough constant, the distribution $\tau_w \mu * \mathcal{N}(0, cI)$ is always well-conditioned log-concave. Thus, the assumptions of Proposition 28 are satisfied and we can run Algorithm 1 by this approximate continuous sampling scheme. It remains to set the accuracy parameter $\delta$, and the number of steps $T$.

We first set $T = A \log(n/\epsilon)$ for a large enough constant $A$ that depends on $c$. This is to make sure that with high probability a sample from $\mathcal{N}(0, cI/T)$ has $\|\cdot\|_\infty$ bounded by $1/5$, which is possible for large enough $A$ by the same argument as in Lemma 7.

Once $T$ is set, we choose $\delta$ to be small enough so that the final $\mathcal{W}_1$ error of $cw_T/T$ resulting from Proposition 27 is at most $\epsilon/10$. This is possible as we only need to set $1/\delta$ to be $(1 + O(1)/c)^T = (n/\epsilon)^{O(1)}$. Overall this results in a runtime of $\log(n/\epsilon)^{O(1)}$ with $(n/\epsilon)^{O(1)}$ many processors.

It remains to conclude that the output is $\epsilon$-close in total variation distance. We know that $cw_T/T$ is $\epsilon/10$-close in $\mathcal{W}_1$ distance to $\mu * \mathcal{N}(0, cI/T)$. Let $x \sim \mu$ and $g \sim \mathcal{N}(0, cI/T)$ be independent samples, and assume that $cw_T/T$ is coupled optimally (w.r.t. $\mathcal{W}_1$) to $x + g$:

$$\mathbb{E}[\|cw_T/T - x - g\|_1] \le \epsilon/10.$$

It follows from Markov's inequality that

$$\mathbb{P}[\|cw_T/T - x - g\|_1 \geq 1/5] \leq \epsilon/2.$$

Moreover, by a similar argument to the proof of Lemma 7, we have

$$\mathbb{P}[\|g\|_\infty \geq 1/5] \leq \epsilon/2.$$

Therefore with probability at least $1-\epsilon$ we have $\|g\|_\infty < 1/5$ and $\|cw_T/T - x - g\|_1 < 1/5$. Under these two conditions we have $\|cw_T/T - x\|_\infty < 2/5 < 1$, which means that $\operatorname{sign}(w_T)$ is the same as $x$. This proves that the output of Algorithm 1 is $\epsilon$-close in total variation distance to $\mu$. □

Next, we prove Theorem 3. The proof is almost identical, except we have to choose a smaller accuracy $\delta$.

*Proof of Theorem 3.* The only difference from the proof of Theorem 1 is that to guarantee $cw_T/T$ is $\epsilon/10$-close in $\mathcal{W}_1$, we no longer can set $1/\delta$ to be $(n/\epsilon)^{O(1)}$. We have to use the trivial transport-stability guaranteed by Corollary 25. This implies that with a $\delta$-accurate continuous sampler given by Proposition 28, the final $cw_T/T$ will be within $\mathcal{W}_1$ distance $\delta(1+O(n)/c)^T = \delta(n/\epsilon)^{O(\log n)}$ of $\mu * \mathcal{N}(0, cI)$. This means that we need to set $1/\delta$ to be $(n/\epsilon)^{O(\log n)}$ large to get $\epsilon/10$ accuracy.

This value of $\delta$ results in a runtime of $\log(n/\delta)^{O(1)} = \log(n/\epsilon)^{O(1)}$ and the number of processors will be $(n/\delta)^{O(1)} = (n/\epsilon)^{O(\log n)}$. □

## 5 Symmetric DPPs and partition-constrained variants

Here we prove that symmetric DPPs and partition-constrained variants of them with $O(1)$ parts satisfy $\mathcal{W}_\infty$-$O(1)\ell_0$ transport stability. This proves Corollary 2, since the Laplace transforms of these distribution are computable in NC. Our main tool is the fact that symmetric DPPs, a subclass of the so-called strongly Rayleigh distributions, satisfy a condition known as *stochastic covering property* [PP14].

**Lemma 30** ([PP14, Proposition 2.1])**.** *Suppose $\mu : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ is a strongly Rayleigh distribution, including DPPs and k-DPPS, and let $i \in [n]$. Then there exists a coupling $\pi$ between $\mu_-$ and $\mu_+$ obtained from $\mu$ by conditioning coordinate $i$ to be $-1$ or $+1$, such that for any $(p, q) \in \operatorname{supp}(\pi)$, $\|p-q\|_1/2 \leq 2$, i.e., $p$ and $q$ differ in at most one coordinate other than $i$. Moreover, if $j$ is any coordinate other than $i$, then $p_j \leq q_j$ (the opposite of $p_i$ and $q_i$).*

**Corollary 31.** *Any strongly Rayleigh distribution, including DPPs and k-DPPS satisfies $\mathcal{W}_\infty$-$4\ell_0$*

*Proof.* Note that DPPs and $k$-DPPs are closed under exponential tilts. Thus the proof follows by arguments in Section 4, where it was shown that to prove $\mathcal{W}_\infty$-$\alpha\ell_0$ transport-stability it is enough to couple distributions obtained from conditioning only one coordinate. □

**Theorem 32.** *Consider a strongly Rayleigh $\mu$ and partition $V_1 \sqcup \cdots \sqcup V_r = [n]$ and cardinalities $c_1, \ldots, c_r$. Let $\nu$ be the distribution of $\mu$ conditioned on those sets that contain exactly $c_j$ elements of $V_j$ for all $j$. Suppose that $i$ is a coordinate and $\nu_+$ is $\nu$ conditioned on $+1$ for coordinate $i$ and $\nu_-$ is $\nu$ conditioned on $-1$ for coordinate $i$. There exists a coupling $\pi$ between $\nu_+$ and $\nu_-$ where the Hamming distance $\|p-q\|_1/2 \leq 2^r$ for $(p, q) \in \operatorname{supp}(\pi)$.*

*Proof.* We construct a pair of random variables $(p, q)$ which couple $\nu_+$ and $\nu_-$ and satisfy $\|p - q\|_1 \leq 2^{r+1}$ almost surely. We construct $p$ and $q$ piece by piece, one part of the partition at a time. W.l.o.g. assume that $i \in V_1$. First, look at the projection of $\nu$ on $V_1$. By Lemma 20, this is a strongly Rayleigh distribution. So we can sample the coordinates in $V_1$, namely $p_{V_1}$ and $q_{V_2}$, in a coupled way, through the coupling guaranteed by Lemma 30, so that they follow the projections of $\nu_+$ and $\nu_-$ on the $V_1$ coordinates. This will create at most 2 different coordinates (one of them is $i$).

Now look at $V_1 \cup V_2$, and consider the distributions obtained by $\nu_+$ and $\nu_-$ projected onto $V_1 \cup V_2$, conditioned on agreement with the filled-out part (coordinates $V_1$) of $p$ and $q$. We argue that these are also strongly Rayleigh distributions, and moreover they are conditionings of the same strongly Rayleigh distribution. This follows because we can consider the coarsened partition $(V_1 \cup V_2) \sqcup V_3 \sqcup \cdots \sqcup V_r$ where we want $c_1 + c_2$ elements from the first part, $c_3$ from the second, and so on. Then, partition-constraining $\mu$ with this coarsened partition, and then conditioning the coordinates in $V_1$ to be $p_{V_1}$ or $q_{V_2}$ results in the same distributions as if we started with the fully partition-constrained $\nu$ and projected onto $V_1 \cup V_2$ and then conditioned on $V_1$; this is because the coordinates in $V_1$ already satisfy the $c_1$ cardinality constraint, so having $c_1 + c_2$ elements from $V_1 \cup V_2$ conditioned on the $V_1$ part is the same as having $c_1$ elements from $V_1$ and $c_2$ elements from $V_2$. Since the conditionings to $p_{V_1}$ and $q_{V_1}$ differ in at most 2 coordinates, it follows by Corollary 31 that we can sample $p_{V_1 \cup V_2}$ and $q_{V_1 \cup V_2}$ conditioned on the $V_1$ part, in a way that $p_{V_1 \cup V_2}$ and $q_{V_1 \cup V_2}$ will end up having twice as many, at most 4 different coordinates. We continue this argument with $V_1 \cup V_2 \cup V_3$ and so on. Each piece of $p$ and $q$ that we create can have as many different coordinates as the total combined so far; in other words, the number of different coordinates at most doubles every iteration. So in the end $p$ and $q$ will be different in $\leq 2^r$ coordinates. This proves that $\|p - q\|_1/2 \leq 2^r$. □

We get the following corollary.

**Corollary 33.** *A partition-constrained symmetric DPP with $r$ parts, satisfies $\mathcal{W}_\infty$-$O(2^r)\ell_0$ transport-stability.*

This finishes the proof of Corollary 2. In Appendix A, we explore, as side results, further implications of $\mathcal{W}_\infty$-$O(2^r)\ell_0$ transport-stability, by closely following the arguments of Hermon and Salez [HS23].

## 6 Non-symmetric DPPs and Eulerian tours

In this section, we show that sampling Eulerian tours on directed graphs can be reduced to sampling non-symmetric DPPs, via the intermediate problem of sampling weighted Eulerian tours on directed 4-regular graphs, described below.

As a reminder, an *Eulerian tour* is a circuit in a finite graph that visits every edge exactly once (revisiting vertices is allowed). A directed graph (or digraph) has an Eulerian tour if and only if every vertex has equal in-degree and out-degree, and all of its vertices with nonzero degree belong to a single strongly connected component. Such graphs are called *Eulerian digraphs*.

We note that the number of Eulerian tours in digraphs can be computed in polynomial time, since there is a many-to-one direct correspondence between Eulerian tours in an Eulerian digraph and arborescences of the graph (known as the BEST theorem [AB51; TS41]), and the latter can be computed by the directed matrix-tree theorem. This is in sharp contrast to the case of undirected graphs for which the problem is #P-complete [BW05], even for 4-regular graphs [GŠ12].

### 6.1 Eulerian tours on 4-regular Eulerian digraphs

A 4-regular digraph $G = (V, E)$ being Eulerian indicates that every vertex has both in-degree and out-degree two. Given $v \in V$, a *transition* $S_v$ is defined as one of the two possible ways of pairing incoming edges and outgoing edges. We say an Eulerian tour $C$ *traverses* $S_v$ if $C$ enters $v$ from any incoming edge and immediately exits via the corresponding outgoing edge. Denote by $S_v(C)$ the transition at $v$ that is being traversed by $C$, and by $\mathcal{E}(G)$ the set of Eulerian tours on $G$. Given an Eulerian tour $R \in \mathcal{E}(G)$ as reference, we set $F_R(C) = \{v \mid S_v(C) \neq S_v(R)\}$. Then we note that $\mathcal{E}(G)$ can be identified by $\mathcal{F}_R = \{F_R(C) \mid C \in \mathcal{E}(G)\}$, as the mapping $F_R$ from Eulerian tours to subsets of vertices is one-to-one and onto. Bouchet [Bou95] proved that the uniform distribution of Eulerian tours on a 4-regular Eulerian digraph is a special case of a non-symmetric DPP. Given a 4-regular graph, we show how to construct the characterizing matrix $L$ for this non-symmetric DPP in RNC.

**Theorem 34.** *Given a 4-regular Eulerian digraph $G = (V, E)$, there exists an* RNC *algorithm that computes an Eulerian tour $R \in \mathcal{E}(G)$ and a skew-symmetric matrix $L \in \{-1, 0, 1\}^{V \times V}$ (whose rows and columns are indexed by $V$) such that $\det(L_{S,S}) = 1$ if $S \in \mathcal{F}_R$, and $\det(L_{S,S}) = 0$ otherwise.*

*Proof.* We first use [AIS84] to find *one* Eulerian tour $R$ using an NC algorihtm. To construct the skew-symmetric matrix $L$, we compute the row $L_{v,w}$ for vertices $v, w \in V$ in parallel.

First, label the two outgoing edges from $v$ by $e_v^+$ and $e_v^-$ arbitrarily; this can be done in one parallel by having each vertex choose its label. Construct the graph $R'$ with vertices consisting of $v^+$ and $v^-$ for $v \in V$, such that $v^+$ ($v^-$ respectively) has exactly one outgoing edge $e_v^+$ ($e_v^-$ respectively) and exactly one in-edge which is the predecessor of $e_v^+$ ($e_v^-$ respectively) in $R$. Note that $R'$ is a simple cycle on $2|V|$ vertices. Following [Bou95], we say vertices $v$ and $w$ has positive (negative respectively) alternance if the vertices in $R'$ appear in the order $v^+ \dots w^+ \dots v^- \dots w^-$ ( $v^+ \dots w^- \dots v^- \dots w^+$ respectively) and has no alternance otherwise.

If $v, w$ has positive (negative respectively) alternance we set $L_{v,w} = +1$ ($L_{v,w} = -1$ respectively) and if they have no alternance we set $L_{v,w} = 0$. To figure out the alternance of any pair of vertices, we simply need to know the index of the copies of $v, w$ appearing on the cycle. This can be done in NC by traversing the cycle $R'$ in parallel using the standard doubling trick. □

In order to reduce the sampling of Eulerian tours on Eulerian digraphs with higher vertex degrees to that on 4-regular Eulerian digraphs, we introduce weights on the transitions. Let $w(\cdot)$ be a positive weight function on the transitions. The distribution of weighted Eulerian tours $\mu_{G,w} : \mathcal{E}(G) \to \mathbb{R}_{>0}$ is defined by

$$\mu_{G,w}(C) \propto \prod_{v \in V} w(S_v(C)).$$

Moreover, since every Eulerian tour traverses exactly one transition at every vertex, given any reference $R \in \mathcal{E}(G)$, we have

$$\mu_{G,w}(C) \propto \prod_{v \in V} w(S_v(R)) \prod_{v \in V} \frac{w(S_v(C))}{w(S_v(R))} \propto \prod_{v \in V} \frac{w(S_v(C))}{w(S_v(R))}.$$

Then it is not hard to see that $\mu_{G,w}$ is identical to the non-symmetric DPP instance given by $L$ in Theorem 34 under the $\lambda$-scaling with $\lambda_v = \frac{w(S_v^*(R))}{w(S_v(R))}$, where $S_v^*(R)$ denotes the transition at $v$ that $R$ does not traverse.

## 6.2 Eulerian tours on general Eulerian digraphs

Finally we show how to sample (unweighted) Eulerian tours on general Eulerian digraphs, given an algorithm for sampling weighted Eulerian tours on 4-regular Eulerian digraphs. Denote the in-degree (or out-degree) of a vertex $v$ by $d(v)$ in a Eulerian digraph. In 4-regular Eulerian digraphs, $d(v) = 2$ for any $v$. We will use (acyclically connected) gadgets made of weighted vertices with $d = 2$ to simulate any unweighted vertex with $d > 2$. Note that unweighted vertices with $d = 1$ can be simply replaced by an edge, for sampling purposes.

To sample unweighted Eulerian tours, the local constraint at a vertex is that the incoming $d$ edges and outgoing $d$ edges should be able to "pair up" freely. Viewing it as a permutation problem, our goal is to generate uniformly random permutations on $d$ elements using (potentially biased) probabilistic pairwise swaps. Indeed, if we connect incoming edges $x_i, x_j$ and outgoing edges $y_i, y_j$ (with the same subscripts, respectively) to a weighted degree-4 vertex $v$ with two transitions $S = \{\{x_i, y_i\}, \{x_j, y_j\}\}$ and $S' = \{\{x_i, y_j\}, \{x_i, y_j\}\}$, then locally $v$ favors the swap (of subscripts $i, j$) with probability $\frac{w(S')}{w(S)+w(S')}$ while favors the identity with probability $\frac{w(S)}{w(S)+w(S')}$.

The following lemma states that for any vertex of in-degree $d$, the construction of a corresponding gadget can be done using $O(d^2)$ degree-4 vertices.

**Lemma 35** (Folklore)**.** *A uniformly random permutation on $d$ elements can be generated using $\frac{d(d-1)}{2}$ biased probabilistic pairwise swaps.*

*Proof.* The construction is in $d$ rounds. We denote the operation of swapping two elements at position $i$ and $j$ by $\mathrm{swap}(i, j)$. In the first round, we do $\mathrm{swap}(1, 2)$ with probability $\frac{1}{2}$, then $\mathrm{swap}(1, 3)$ with probability $\frac{1}{3}$, etc., and eventually $\mathrm{swap}(1, d)$ with probability $\frac{1}{d}$. This will put a uniformly random element in position $1$, which takes $d$ probabilistic swaps. Next we can keep the element in position $1$ fixed and recursively work on the remaining $d - 1$ positions. For example, in the second round a uniformly random element amongst the remaining $d - 1$ elements will be put in positions $2$, using $d - 1$ probabilistic swaps. Continuing this process for $d$ rounds yields a perfectly random permutation. Since the round with $i$ elements remaining takes $i$ probabilistic swaps, the total number of swaps is $\frac{d(d-1)}{2}$. □

Therefore, given any Eulerian digraph $G$, the reduction works by replacing every vertex with $d > 2$ by a gadget[2] of $O(d^2)$ vertices with in-degree $d = 2$, and every vertex with $d = 1$ by a directed edge, resulting in a 4-regular Eulerian digraph $G'$ whose size is blown up at most quadratically. Any sample of an Eulerian tour on $G'$ can be converted to a sample of an Eulerian tour on $G$ by virtually viewing each vertex gadget constructed in $G'$ as a vertex in $G$, and letting the edges in $G$ follow their ordering in $G'$ to form an Eulerian tour. Furthermore, according to Lemma 35, the desired distribution will be respected at the aggregate level.

We remark that the idea of simulating higher degree vertices by degree-4 vertices was previously employed in [GŠ12] where they studied Eulerian tours problems on undirected graphs, although their construction only uses unweighted degree-4 vertices and their equivalence result holds in the approximate sense.

[2]The gadget construction can easily be done in NC time

## References


[AB51] T. van Aardenne-Ehrenfest and N. G. de Bruijn. "Circuits and trees in oriented linear graphs". In: *Simon Stevin* 28 (1951), pp. 203–217. ISSN: 0037-5454.

[AIS84] B. Awerbuch, A. Israeli, and Y. Shiloach. "Finding Euler Circuits in Logarithmic Parallel Time". In: *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*. STOC '84. New York, NY, USA: Association for Computing Machinery, 1984, pp. 249–257. ISBN: 0897911334. DOI: 10.1145/800057.808688.

[Ald90] David J. Aldous. "The random walk construction of uniform spanning trees and uniform labelled trees". In: *SIAM J. Discrete Math.* 3.4 (1990), pp. 450–465. ISSN: 0895-4801. DOI: 10.1137/0403039.

[Ali+21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. "Fractionally log-concave and sector-stable polynomials: counting planar matchings and more". In: *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, [2021] ©2021, pp. 433–446. DOI: 10.1145/3406325.3451123.

[Ana+21] Nima Anari, Nathan Hu, Amin Saberi, and Aaron Schild. "Sampling arborescences in parallel". In: *12th Innovations in Theoretical Computer Science Conference*. Vol. 185. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021, Art. No. 83, 18. DOI: 10.4230/LIPIcs.ITCS.2021.83.

[Ana+22a] Nima Anari, Callum Burgess, Kevin Tian, and Thuy-Duong Vuong. *Quadratic Speedups in Parallel Sampling from Determinantal Distributions*. 2022. arXiv: 2203.11190 [cs.DS].

[Ana+22b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic independence: optimal mixing of down-up random walks". In: *STOC '22—Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, [2022] ©2022, pp. 1418–1430. DOI: 10.1145/3519935.3520048.

[BBL09] Julius Borcea, Petter Brändén, and Thomas M. Liggett. "Negative dependence and the geometry of polynomials". In: *J. Amer. Math. Soc.* 22.2 (2009), pp. 521–567. ISSN: 0894-0347. DOI: 10.1090/S0894-0347-08-00618-8.

[Bou95] André Bouchet. "Coverings and delta-coverings". In: *Integer programming and combinatorial optimization (Copenhagen, 1995)*. Vol. 920. Lecture Notes in Comput. Sci. Springer, Berlin, 1995, pp. 228–243. DOI: 10.1007/3-540-59408-6_54.

[Brä07] Petter Brändén. "Polynomials with the half-plane property and matroid theory". In: *Adv. Math.* 216.1 (2007), pp. 302–320. ISSN: 0001-8708. DOI: 10.1016/j.aim.2007.05.011.

[Bro89] A. Broder. "Generating Random Spanning Trees". In: *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*. SFCS '89. USA: IEEE Computer Society, 1989, pp. 442–447. ISBN: 0818619821. DOI: 10.1109/SFCS.1989.63516.

[BW05] Graham R. Brightwell and Peter Winkler. "Counting Eulerian Circuits is #P-Complete". In: *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics, ALENEX /ANALCO 2005, Vancouver, BC, Canada, 22 January 2005*. Ed. by Camil Demetrescu, Robert Sedgewick, and Roberto Tamassia. SIAM, 2005, pp. 259–262.

[CE22] Yuansi Chen and Ronen Eldan. "Localization schemes: a framework for proving mixing bounds for Markov chains". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*. IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022, pp. 110–122. DOI: 10.1109/FOCS54457.2022.00018.

[Cel+18] Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, and Nisheeth Vishnoi. "Fair and Diverse DPP-Based Data Summarization". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 716–725.

[CK78] S. Chaiken and D. J. Kleitman. "Matrix tree theorems". In: *J. Combinatorial Theory Ser. A* 24.3 (1978), pp. 377–381. ISSN: 0097-3165. DOI: 10.1016/0097-3165(78)90067-5.

[Csa76] L. Csanky. "Fast parallel matrix inversion algorithms". In: *SIAM J. Comput.* 5.4 (1976), pp. 618–623. ISSN: 0097-5397. DOI: 10.1137/0205040.

[Der21] Michael W. Dereziński Michałand Mahoney. "Determinantal point processes in randomized numerical linear algebra". In: *Notices Amer. Math. Soc.* 68.1 (2021), pp. 34–45. ISSN: 0002-9920. DOI: 10.1090/noti2202.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. "A random polynomial-time algorithm for approximating the volume of convex bodies". In: *J. Assoc. Comput. Mach.* 38.1 (1991), pp. 1–17. ISSN: 0004-5411. DOI: 10.1145/102782.102783.

[DGJ04] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. "Counting and sampling $H$-colourings". In: *Inform. and Comput.* 189.1 (2004), pp. 1–16. ISSN: 0890-5401. DOI: 10.1016/j.ic.2003.09.001.

[Eld13] Ronen Eldan. "Thin shell implies spectral gap up to polylog via a stochastic localization scheme". In: *Geom. Funct. Anal.* 23.2 (2013), pp. 532–569. ISSN: 1016-443X. DOI: 10.1007/s00039-013-0214-y.

[ELL17] Ronen Eldan, James R. Lee, and Joseph Lehec. "Transport-entropy inequalities and curvature in discrete-space Markov chains". In: *A journey through discrete mathematics*. Springer, Cham, 2017, pp. 391–406. DOI: 10.1007/978-3-319-44479-6_16.

[EM22] Ahmed El Alaoui and Andrea Montanari. "An information-theoretic view of stochastic localization". In: *IEEE Trans. Inform. Theory* 68.11 (2022), pp. 7423–7426. ISSN: 0018-9448. DOI: 10.1109/TIT.2022.3180298.

[EMS22] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. "Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*. IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022, pp. 323–334. DOI: 10.1109/FOCS54457.2022.00038.

[ES22] Ronen Eldan and Omer Shamir. "Log concavity and concentration of Lipschitz functions on the Boolean hypercube". In: *J. Funct. Anal.* 282.8 (2022), Paper No. 109392, 22. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2022.109392.

[FHY21] Weiming Feng, Thomas P. Hayes, and Yitong Yin. "Distributed metropolis sampler with optimal parallelism". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. [Society for Industrial and Applied Mathematics (SIAM)], Philadelphia, PA, 2021, pp. 2121–2140. DOI: 10.1137/1.9781611976465.127.

[Gar+19] Mike Gartrell, Victor-Emmanuel Brunel, Elvis Dohmatob, and Syrine Krichene. "Learning Nonsymmetric Determinantal Point Processes". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019.

[GŠ12] Qi Ge and Daniel Štefankovič. "The complexity of counting Eulerian tours in 4-regular graphs". In: *Algorithmica* 63.3 (2012), pp. 588–601. ISSN: 0178-4617. DOI: 10.1007/s00453-010-9463-4.

[HS23] Jonathan Hermon and Justin Salez. "Modified log-Sobolev inequalities for strong-Rayleigh measures". In: *Ann. Appl. Probab.* 33.2 (2023), pp. 1301–1314. ISSN: 1050-5164. DOI: 10.1214/22-aap1847.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *J. ACM* 51.4 (2004), pp. 671–697. ISSN: 0004-5411. DOI: 10.1145/1008731.1008738.

[JVV86] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoret. Comput. Sci.* 43.2-3 (1986), pp. 169–188. ISSN: 0304-3975. DOI: 10.1016/0304-3975(86)90174-X.

[Kas63] P. W. Kasteleyn. "Dimer statistics and phase transitions". In: *J. Mathematical Phys.* 4 (1963), pp. 287–293. ISSN: 0022-2488. DOI: 10.1063/1.1703953.

[KT12] Alex Kulesza and Ben Taskar. "Determinantal Point Processes for Machine Learning". In: *Foundations and Trends® in Machine Learning* 5.2–3 (2012), pp. 123–286. ISSN: 1935-8237. DOI: 10.1561/2200000044.

[LY22] Hongyang Liu and Yitong Yin. "Simple parallel algorithms for single-site dynamics". In: *STOC '22—Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, [2022] ©2022, pp. 1431–1444. DOI: 10.1145/3519935.3519999.

[PP14] Robin Pemantle and Yuval Peres. "Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures". In: *Combin. Probab. Comput.* 23.1 (2014), pp. 140–160. ISSN: 0963-5483. DOI: 10.1017/S0963548313000345.

[SL19] Ruoqi Shen and Yin Tat Lee. "The Randomized Midpoint Method for Log-Concave Sampling". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019.

[Ten95] Shang-Hua Teng. "Independent sets versus perfect matchings". In: *Theoret. Comput. Sci.* 145.1-2 (1995), pp. 381–390. ISSN: 0304-3975. DOI: 10.1016/0304-3975(94)00289-U.

[TF61] H. N. V. Temperley and Michael E. Fisher. "Dimer problem in statistical mechanics—an exact result". In: *Philos. Mag. (8)* 6 (1961), pp. 1061–1063. ISSN: 0031-8086. DOI: 10.1080/14786436108243366.

[TS41] W. T. Tutte and C. A. B. Smith. "On Unicursal Paths in a Network of Degree 4". In: *Amer. Math. Monthly* 48.4 (1941), pp. 233–237. ISSN: 0002-9890. DOI: 10.2307/2302716.

# A Concentration and faster mixing walk from coupling

In this section we connect $\mathcal{W}_\infty$-$O(1)\ell_0$ transport stability, see Definition 23, to concentration inequalities and the existence of fast mixing local Markov chains. We follow almost identically the arguments of Hermon and Salez [HS23] who proved the same for the more restricted class of strongly Rayleigh distributions. Our proof of $\mathcal{W}_\infty$-$O(1)\ell_0$ transport stability from Section 5 extends these results to partition-constrained strongly Rayleigh distributions and DPPs.

We show the existence of a local Markov chain with inverse linear modified log Sobolev constant, which in turn implies sub-Gaussian concentration of Lipschitz function via a standard Herbst argument [HS23].

**Theorem 36.** *Suppose $\mu : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ satisfies $\mathcal{W}_\infty$-$\rho\ell_0$ transport-stability for some parameter $\rho = O(1)$. Then, there exists a reversible Markov chain $P$ that has stationary distribution $\mu$ and modified log-Sobolev constant $\geq \Omega(\frac{1}{n})$, and furthermore $P(x,y) \neq 0$ only if $\|x - y\|_1 \leq \rho$. If we assume additionally that $\mu$ is $k$-homogeneous, then the modified log-Sobolev constant can be lower-bounded by $\Omega(1/k)$.*

*Proof.* Let $\Omega = \{\pm 1\}^n$. Fix a coordinate $\ell \in [n]$. Let $\Omega_- = \{x \in \Omega \mid x_\ell = -1\}$ and $\Omega_+ = \{x \in \Omega \mid x_\ell = +1\}$.

Following the notation in [HS23], it is easier to construct a reversible Markov generator $Q$ with transition rate bounded by $O(n)$ (or $O(k)$ in the homogeneous case) and MLSI constant $\Omega(1)$ and then convert that to a Markov chain, by lazifying $Q$ at a rate of $\simeq 1/n$ (or $\simeq 1/k$ for the homogeneous case). For a Markov generator $Q$, we let $\alpha(Q)$ be its MLSI constant. Let $\Delta(Q) = \max_{x\in\Omega}\{-Q(x,x)\}$ be the maximum rate of change of $Q$.

Consider the projections and restrictions of $\mu$ defined by the partition $\Omega_-$ and $\Omega_+$ denoted by $\hat{\mu}, \mu_-, \mu_+$. Suppose $\mu$ satisfies $\mathcal{W}_\infty$-$\rho\ell_0$, which means there is a coupling $\kappa$ between $\mu_-$ and $\mu_+$ that has distance $\mathcal{W}_\infty(\mu_-, \mu_+) \le \rho$.

The induction hypothesis provides a reversible Markov generator $Q_-$ ($Q_+$ respectively) such that $\alpha(Q_\pm) \ge 1$ and $\Delta(Q_\pm) \le n-1$ always. Additionally if $\mu$ is $k$-homogeneous then $\Delta(Q_-) \le 2k$ and $\Delta(Q_+) \le 2k-2$.

Recall that the projection $\hat{\mu}$ of $\mu$ on $\{\Omega_-, \Omega_+\}$ is defined by $\hat{\mu}(+) = \sum_{x\in\Omega_+}\mu(x)$ and $\hat{\mu}(-) = \sum_{x\in\Omega_-}\mu(x)$. We define $Q$ for $\mu$ as follows. For $x \ne y \in \Omega$

$$Q(x,y) = \begin{cases} Q_-(x,y) & \text{if } (x,y)\in\Omega_-\times\Omega_-, \\ Q_+(x,y) & \text{if } (x,y)\in\Omega_+\times\Omega_+, \\ \frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(x,y)}{\mu(x)} & \text{if } (x,y)\in\Omega_-\times\Omega_+, \\ \frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(y,x)}{\mu(x)} & \text{if } (x,y)\in\Omega_+\times\Omega_-, \end{cases}$$

where the diagonal is defined so that $\sum_y Q(x,y) = 0$. First, we check that $Q$ is reversible. Since $Q_-$ and $Q_+$ are reversible, for $(x,y)\in\Omega_-\times\Omega_-$,

$$\mu(x)Q(x,y) \quad = \quad \hat{\mu}(-)\mu_-(x)Q_-(x,y) \quad = \quad \hat{\mu}(-)\mu_-(y)Q_-(y,x) \quad = \quad \mu(y)Q_-(y,x).$$

The case $(x,y)\in\Omega_+\times\Omega_+$ is similar. For $(x,y)\in\Omega_-\times\Omega_+$,

$$\mu(x)Q(x,y) = \mu(x)\frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(x,y)}{\mu(x)} = \hat{\mu}(+)\hat{\mu}(-)\kappa(x,y) = \mu(y)\frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(x,y)}{\mu(y)} = \mu(y)Q(y,x).$$

$Q$ induces a projection chain with Markov generator $\hat{Q}$ on state space $\{+,-\}$ where

$$\begin{aligned}\hat{\mu}(-)\hat{Q}(-,+) &= \sum_{(x,y)\in\Omega_-\times\Omega_+}\mu(x)Q(x,y) = \sum_{(x,y)\in\Omega_-\times\Omega_+}\mu(x)\frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(x,y)}{\mu(x)} \\ &= \hat{\mu}(+)\hat{\mu}(-)\sum_{(x,y)\in\Omega_-\times\Omega_+}\kappa(x,y) = \hat{\mu}(+)\hat{\mu}(-),\end{aligned}$$

and

$$\begin{aligned}\hat{\mu}(+)\hat{Q}(+,-) &= \sum_{(x,y)\in\Omega_+\times\Omega_-}\mu(x)Q(x,y) = \sum_{(x,y)\in\Omega_+\times\Omega_-}\mu(x)\frac{\hat{\mu}(+)\hat{\mu}(-)\kappa(y,x)}{\mu(x)} \\ &= \hat{\mu}(+)\hat{\mu}(-)\sum_{(y,x)\in\Omega_-\times\Omega_+}\kappa(y,x) = \hat{\mu}(+)\hat{\mu}(-).\end{aligned}$$

For any $(x,y)\in\operatorname{supp}(\kappa)\subseteq\Omega_-\times\Omega_+$, by construction the following equals 1:

$$\frac{\mu(x)Q(x,y)}{\hat{\mu}(-)\hat{Q}(-,+)\kappa(x,y)} = \frac{\hat{\mu}(+)}{\hat{Q}(-,+)} = \frac{\hat{\mu}(-)}{\hat{Q}(+,-)} = \frac{\mu(y)Q(y,x)}{\hat{\mu}(+)\hat{Q}(+,-)\kappa(y,x)}.$$

By [HS23, Lemma 3], the projection Markov generator $\hat{Q}$ satisfies $\alpha(\hat{Q}) \geq \hat{Q}(-,+) + \hat{Q}(+,-)$. Thus following [HS23], letting

$$\chi = \min_{(x,y)\in\Omega_-\times\Omega_+} \left\{ \frac{\mu(x)Q(x,y)}{\hat{\mu}(-)\hat{Q}(-,+)\kappa(x,y)} \right\},$$ [3]

then

$$\chi = \frac{\hat{\mu}(+)}{\hat{Q}(-,+)} = \frac{\hat{\mu}(-)}{\hat{Q}(+,-)},$$

and thus

$$\chi\hat{\alpha}(Q) \geq \chi(\hat{Q}(-,+) + \hat{Q}(+,-)) = \hat{\mu}(+) + \hat{\mu}(-) = 1.$$

Applying [HS23, Lemma 1], we have

$$\alpha(Q) \geq \min\{\chi\alpha(\hat{Q}), \min\alpha(Q_\pm)\} \geq 1.$$

For $(x,y) \in \Omega_- \times \Omega_+$, we say $x \sim y$ iff $\|x-y\|_1 \leq \rho$. Obviously $Q(x,y) \neq 0$ iff $x \sim y$.

Next, we check that $\Delta(Q) \leq n$. For $x \in \Omega_-$,

$$\begin{aligned}
-Q(x,x) &= \sum_{y\in\Omega_-\setminus\{x\}} Q_-(x,y) + \frac{\hat{\mu}(-)\hat{\mu}(+)}{\mu(x)} \sum_{y\in\Omega_+} \kappa(x,y) \\
&\leq \Delta(Q_-) + \frac{\hat{\mu}(-)\hat{\mu}(+)}{\mu(x)} \cdot \mu(x) \\
&= \Delta(Q_-) + \hat{\mu}(-)\hat{\mu}(+) \\
&\leq \Delta(Q_-) + \hat{\mu}(+).
\end{aligned}$$

Similarly, for $x \in \Omega_+$,

$$-Q(x,x) \leq \Delta(Q_+) + \hat{\mu}(-).$$

So we conclude that $\Delta(Q) \leq 1 + \max\{\Delta(Q_-), \Delta(Q_+)\} \leq n$.

Now assume that we have a $k$-homogeneous distribution. In this case, $Q$, as defined, will not necessarily have the desired property that $\Delta(Q) \leq 2k$. But we can get a Markov generator with a similarly large MLSI constant and the desired rate by averaging over the choice of the coordinate $\ell$.

First, notice that for $x \in \Omega_-$, we have

$$-Q(x,x) \leq 2k + \mathbb{E}_{y\sim\mu}[(y_\ell + 1)/2],$$

and for $x \in \Omega_+$, we have

$$-Q(x,x) \leq 2k - 2 + 1 = 2k - 1.$$

In either case we have

$$-Q(x,x) \leq 2k + \mathbb{E}_{y\sim\mu}[y_\ell/2] - x_\ell/2.$$

We rewrite $Q$ as $Q^{(\ell)}$ to make explicit the dependence on the element $\ell$. Take the linear combination $Q^* = \frac{1}{n}\sum_\ell Q^{(\ell)}$. This is again a Markov generator that only transitions betweens states that are $\rho$ apart in Hamming distance, i.e., $Q^*(x,y) \neq 0$ only if $\|x-y\|_1 \leq \rho$. The quantity $\alpha$ is preserved by convex combination, thus $\alpha(Q^*) \geq 1$. And finally,

$$-Q^*(x,x) = \frac{1}{n}\sum -Q^{(\ell)}(x,x) \leq 2k.$$

□

[3] This follow from [HS23, Eq.(48)], with the additional observation that $\kappa(x,y) \neq 0$ iff $x \in \Omega_-$ and $y \in \Omega_+$.

We get the concentration of Lipschitz functionals as a corollary, almost identically to [HS23].

**Corollary 37.** *Suppose $\mu$ satisfies $\mathcal{W}_\infty$-$\rho\ell_0$ transport-stability. Let $f : \binom{[n]}{k} \to \mathbb{R}$ be a $c$-Lipschitz functional. Then*

$$\mathbb{P}_{S\sim\mu}\left[f(S) \geq \mathbb{E}_\mu[f(S)] + a\right] \leq \exp\left(-\Omega\left(\frac{a^2}{k(c\rho)^2}\right)\right).$$