# A New Notion of Commutativity for the Algorithmic Lovász Local Lemma 

David G. Harris $\square$<br>University of Maryland, College Park, MD, USA<br>Fotis Iliopoulos $\square$<br>Institute for Advanced Study, Princeton, NJ, USA<br>Vladimir Kolmogorov $\square$<br>Institute of Science and Technology, Klosterneuburg, Austria


#### Abstract

The Lovász Local Lemma (LLL) is a powerful tool in probabilistic combinatorics which can be used to establish the existence of objects that satisfy certain properties. The breakthrough paper of Moser \& Tardos and follow-up works revealed that the LLL has intimate connections with a class of stochastic local search algorithms for finding such desirable objects. In particular, it can be seen as a sufficient condition for this type of algorithms to converge fast.

Besides conditions for convergence, many other natural questions can be asked about algorithms; for instance, "are they parallelizable?", "how many solutions can they output?", "what is the expected "weight" of a solution?". These questions and more have been answered for a class of LLL-inspired algorithms called commutative. In this paper we introduce a new, very natural and more general notion of commutativity (essentially matrix commutativity) which allows us to show a number of new refined properties of LLL-inspired local search algorithms with significantly simpler proofs.


2012 ACM Subject Classification Mathematics of computing $\rightarrow$ Probabilistic algorithms; Mathematics of computing $\rightarrow$ Combinatorics

Keywords and phrases Lovász Local Lemma, Resampling, Moser-Tardos algorithm, latin transversal, commutativity

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2021.31
Category RANDOM
Related Version Full Version: https://arxiv.org/abs/2008.05569
Funding Fotis Iliopoulos: This material is based upon work directly supported by the IAS Fund for Math and indirectly supported by the National Science Foundation Grant No. CCF-1900460. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work is also supported by the National Science Foundation Grant No. CCF-1815328.
Vladimir Kolmogorov: Supported by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 616160.

## 1 Introduction

The Lovász Local Lemma (LLL) is a powerful tool in probabilistic combinatorics which can be used to establish the existence of objects that satisfy certain properties [9]. At a high level, it states that given a collection of bad events in a probability space $\mu$, if each bad-event is not too likely and, further, is independent of most other bad events, then the probability of avoiding all of them is strictly positive.


In its simplest, "symmetric" form, it states that if each bad-event has probability at most $p$ and is dependent with at most $d$ others, where epd $\leq 1$, then with positive probability no bad-events become true. In particular, a configuration avoiding all the bad-events exists. Although the LLL applies to general probability spaces, most constructions in combinatorics use a simpler setting we refer to as the variable version $L L L$. Here, the probability space $\mu$ is a cartesian product with $n$ independent variables, and each bad-event is determined by a subset of the variables. Two bad-events may conflict if they depend on a common variable.

For example, consider a CNF formula with $n$ variables where each clause has $k$ literals and each variable appears in at most $L$ clauses. For each clause $c$ we can define the bad event $B_{c}$ that $c$ is violated in a chosen assignment of the variables. For a uniformly random variable assignment, each bad-event has probability $p=2^{-k}$ and affects at most $d \leq k L$ others. So when $L \leq \frac{2^{k}}{e k}$, the formula is satisfiable; crucially, this criterion does not depend on the number of variables $n$.

A generalization known as the Lopsided LLL (LLLL) allows bad-events to be positively correlated with others; this is as good as independence for the purposes of the LLL. Some notable probability spaces satisfying the LLLL include the uniform distribution on permutations and the variable setting, where two bad-events $B, B^{\prime}$ are dependent only if they disagree on the value of a common variable.

In a seminal work, Moser \& Tardos [25] presented a simple local search algorithm to make the variable-version LLLL constructive. This algorithm can be described as follows:

Algorithm 1 The Moser-Tardos (MT) resampling algorithm.

```
Draw the state \sigma from distribution }
while some bad-event is true on }\sigma\mathrm{ do
        Select, arbitrarily, a bad-event B true on }
        Resample, according to distribution }\mu\mathrm{ , all variables in }\sigma\mathrm{ affecting B
```

Moser \& Tardos showed that if the symmetric LLL criterion (or more general asymmetric criterion) is satisfied, then this algorithm quickly converges. Following this work, a large effort has been devoted to making different variants of the LLL constructive. This research has taken many directions, including further analysis of Algorithm 1 and its connection to different LLL criteria $[6,22,26]$.

One line of research has been to use variants of Algorithm 1 for general probability spaces beyond the variable LLL. These include applications of the LLL to permutations and matchings of the clique $[1,2,17,21,19]$ as well as settings not directly connected to the LLL itself $[3,7,18]$. At a high level of generality, we summarize this in the following framework. There is a discrete state space $\Omega$, with a collection $\mathcal{F}$ of subsets (which we call flaws) of $\Omega$. There is also some problem-specific randomized procedure called the resampling oracle $\Re_{f}$ for each flaw $f$; it takes some random action to attempt to "fix" that flaw, resulting in a new state $\sigma^{\prime} \leftarrow \Re_{f}(\sigma)$. With these ingredients, we define the general local search algorithm as follows:

Algorithm 2 The general local search algorithm.

```
Draw the state \(\sigma\) from some distribution \(\mu\)
while some flaw holds on \(\sigma\) do
    Select a flaw \(f\) of \(\sigma\), according to some rule \(S\).
    Update \(\sigma \leftarrow \mathfrak{R}_{f}(\sigma)\).
```

We refer to Algorithm 2 throughout as the Search Algorithm. The most important question about its behavior is whether it converges to a flawless object. But, there are other important questions to ask; for instance, "is it parallelizable?", "how many solutions can it output?", "what is the expected "weight" of a solution?". These questions and more have been answered for the Moser-Tardos algorithm in a long series of papers $[6,8,11,12,15,16,22,25]$. As a prominent example, the result of Haeupler, Saha and Srinivasan [12], as well as follow-up work of Harris and Srinivasan [14, 16], allows one to argue about the dynamics of Algorithm 1, resulting in several new applications such as estimating the entropy of the output distribution, partially avoiding bad events and dealing with super-polynomially many bad events.

There is one important difference between Algorithm 1 and Algorithm 2: the choice of which flaw to resample, if multiple flaws are simultaneously true. The flaw selection rule $S$ in the Search Algorithm should select a flaw $f \ni \sigma$ at each time $t$; it may depend on the prior states and may be randomized. The original MT algorithm allows nearly complete freedom for this. For general resampling oracles, $S$ is much more constrained; only a few relatively rigid rules are known to guarantee convergence, such as selecting the flaw of least index [19]. However, in [23], Kolmogorov identified a general property called commutativity that allows a free choice for $S$. This free choice, seemingly a minor detail, turns out to play a key role in extending the additional properties of the MT algorithm to the general Search Algorithm. For instance, it leads to parallel algorithms [23] and to bounds on the output distribution [20].

At a high level, our goal is to provide a more conceptual, algebraic explanation for the commutativity properties of resampling oracles and their role in the Search Algorithm. We do this by introducing a notion of commutativity, essentially matrix commutativity, that is both more general and simpler than the definition in [23]. Most of our results had already been shown, in slightly weaker forms, in prior works [23, 20, 14]. However, the proofs were computationally heavy and narrowly targeted to certain probability spaces, with numerous technical side conditions and restrictions.

Before we provide formal definitions, let us give some intuition. For each flaw $f$, consider an $|\Omega| \times|\Omega|$ transition matrix $A_{f}$. Each row of $A_{f}$ describes the probability distribution obtained by resampling $f$ at a given state $\sigma$. We call the resampling oracle commutative if the transition matrices commute for any pair of flaws which are "independent" (in the LLL sense). We show a number of results for such oracles:

1. We obtain bounds on the distribution of the state at the termination of the Search Algorithm. These bounds are comparable to the LLL-distribution, i.e., the distribution induced by conditioning on avoiding all bad events. Similar results, albeit with a number of additional technical restrictions, had been shown in [20] for the original definition of commutativity.
2. For some probability spaces, stronger and specialized distributional bounds are available, beyond the "generic" LLL bounds [14]. Previously, these had been shown with ad-hoc arguments specialized to each probability space. Our construction recovers most of these results automatically.
3. We develop a generic parallel implementation of the Search Algorithm. This extends results of [23, 15], with simpler and more general proofs.
4. In many settings, flaws are formed from smaller "atomic" events [15]. We show that, if the atomic events satisfy the generalized commutativity definition, then so do the larger "composed" events. This natural property did not seem to hold for the original commutativity definition of [23].

### 1.1 Example application: latin transversals

As a motivating example, let us examine a classic combinatorial problem of latin transversals. Consider an $n \times n$ array $C$, wherein each entry of $C$ has a color. A latin transversal of $C$ is a permutation $\pi$ over $\{1, \ldots, n\}$ such that all the colors $C(i, \pi i)$ are distinct for $i=1, \ldots, n$.

The "lopsided" variant of the LLL was first developed by Erdős \& Spencer [10] for this problem. The underlying probability space is the uniform distribution on permutations. For each pair of cells $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)$ of the same color, there is a corresponding flaw defined by $\pi x_{1}=y_{1} \wedge \pi x_{2}=y_{2}$. They showed that if each color appears at most $\Delta=\frac{n}{4 e}$ times, then the LLL criteria are satisfied and a transversal exists. The cluster-expansion criterion [5] tightens this to $\Delta=\frac{27 n}{256}$, which is the strongest bound currently known.

This construction has been a motivating example for much of the research on the algorithmic LLL, particularly for "exotic" probability spaces (beyond the variable setting). In particular, [17] showed that the Search Algorithm generates a latin transversal $\pi$ under the same conditions as the existential LLLL. One of the main applications in this paper is to show that $\pi$ has nice distributional properties. In particular, we show the following:

- Theorem 1. If each color appears at most $\frac{27}{256} n$ times in the array, then the Search Algorithm generates a latin transversal $\pi$ where, for every pair $(x, y)$, we have $0.53 / n \leq$ $\operatorname{Pr}(\pi x=y) \leq 1.36 / n$.

The upper bound improves quantitatively over a similar bound of [14]; to the best of our knowledge, no non-trivial lower bound of any kind could previously be shown. Intriguingly, such bounds are not known to hold for the LLL-distribution itself.

To better situate Theorem 1, note that Alon, Spencer, \& Tetali [4] showed that there is a (minuscule) universal constant $\beta>0$ with the following property. If each color appears at most $\Delta=\beta n$ times in the array and $n$ is a power of two, then the array can be partitioned into $n$ independent transversals. In this case, if we randomly select one transversal from this list, we would have $\operatorname{Pr}(\pi x=y)=1 / n$. Theorem 1 can be regarded as a simplified fractional analogue of their result, i.e. we fractionally decompose the given array into $O(n)$ transversals, such that $\operatorname{Pr}(\pi x=y)=\Theta(1 / n)$ for all pairs $x, y$. Furthermore, we achieve this guarantee automatically, merely by running the Search Algorithm.

### 1.2 Overview of our approach

Although it will require significant definitions and technical development to state our results formally, let us try to provide a high-level summary here. As a starting point, consider the MT algorithm. Moser \& Tardos [25] used a construction referred to as a witness tree for the analysis: for each resampling of a bad-event $B$ at a given time, there is a corresponding witness tree which records an "explanation" of why $B$ was true at that time. More properly, it provides a history of all the prior resamplings which affected the variables involved in $B$.

The main technical lemma governing the behavior of the MT algorithm is the "Witness Tree Lemma," which states that the probability of producing a given witness tree is at most the product of the probabilities of the corresponding events. The bounds on the algorithm runtime, as well as parallel algorithms and distributional properties, then follow from a union bound over witness trees.

Versions of this Witness Tree Lemma have been shown for some variants of the MT algorithm [13, 18] Iliopoulos [20] further showed that it held for general spaces which satisfy the commutativity property; this, in turn, leads to the nice algorithmic properties such as parallel algorithms.

Our main technical innovation is to generalize the Witness Tree Lemma. Instead of keeping track of a scalar product of probabilities in a witness tree, we instead consider a matrix product. We bound the probability of a given witness tree (or, more properly, a slight generalization known as the witness DAG) in terms of the products of the transition matrices of the corresponding flaws. Commutativity can thus be rephrased and simplified in terms of matrix commutativity for the transition matrices.

At the end, we obtain the scalar form of the Witness Tree Lemma by projecting everything to a one-dimensional space. For this, we take advantage of some methods of [3] for viewing the evolution of the Search Algorithm in terms of spectral bounds.

## 2 Background and Basic Definitions

Throughout the paper we consider implementations of the Search Algorithm. For each flaw $f$, state $\sigma \in f$, and state $\sigma^{\prime} \in \Omega$, we define $A_{f}\left[\sigma, \sigma^{\prime}\right]$ to be the probability that applying the resampling oracle $\mathfrak{R}_{f}$ to $\sigma$ yields state $\sigma^{\prime}$, i.e. $A_{f}\left[\sigma, \sigma^{\prime}\right]=\operatorname{Pr}\left(\mathfrak{R}_{f}(\sigma)=\sigma^{\prime}\right)$. For $\sigma \notin f$, we define $A_{f}\left[\sigma, \sigma^{\prime}\right]=0$. We sometimes write $\sigma \xrightarrow{f} \sigma^{\prime}$ to denote that the algorithm resamples flaw $f$ at $\sigma$ and moves to $\sigma^{\prime}$. Observe that, for any vector $\theta$ over $\Omega$, there holds $\left\|\theta^{\top} A_{f}\right\|_{1}=\sum_{\sigma \in f} \theta[\sigma] \leq\left\|\theta^{\top}\right\|_{1}$. Thus, the matrix $A_{f}$ is substochastic.

We define a trajectory $T$ to be a finite or countably infinite sequence of the states and flaws of the form $\left(\sigma_{0}, f_{1}, \sigma_{1}, f_{2}, \ldots,\right)$, and len $(T)$ is its length (possibly len $\left.(T)=\infty\right)$. We define the shift of $T$ to be $\left(\sigma_{1}, f_{2}, \sigma_{2}, f_{3}, \ldots\right.$, . We define $\hat{T}$ to be the sequence states and flaws resampled during the Search Algorithm, i.e. $\sigma_{i}$ is the state at time $i$ and flaw $f_{i} \in \sigma_{i}$ is the flaw resampled.

For our purposes, we use an undirected notion of dependence. Formally, we suppose that we have fixed some symmetric relation $\sim$ on $\mathcal{F}$, with the property that $f \sim f$ for all $f$ and for every distinct pair of flaws $f \nsim g$, we are guaranteed that resampling flaw $f$ cannot introduce $g$ or vice-versa, i.e. $\mathfrak{R}_{f}$ never maps a state $\Omega-g$ into $g$ and likewise $\mathfrak{R}_{g}$ never maps a state from $\Omega-f$ into $f$. We define $\bar{\Gamma}(f)$ to be the set of flaws $g$ with $f \sim g$, and we also define $\Gamma(f)=\bar{\Gamma}(f) \backslash\{f\}$.

We say that a set $I \subseteq \mathcal{F}$ is stable if $f \nsim g$ for all distinct pairs $f, g \in I$.
For an arbitrary event $E \subseteq \Omega$, we define $e_{E}$ to be the indicator vector for $E$, i.e. $e_{E}[\sigma]=1$ if $\sigma \in E$ and $e_{E} \sigma=0$ otherwise. For a state $\sigma \in \Omega$, we write $e_{\sigma}$ as shorthand for $e_{\{\sigma\}}$, i.e. the basis vector which has a 1 in position $\sigma$ and zero elsewhere.

For vectors or matrices $u, v$ we write $u \preceq v$ if $u[i] \leq v[i]$ for all entries $i$. We write $u \propto v$ if there is some scalar value $c$ with $u=c v$.

Regenerating oracles. The Moser-Tardos algorithm and extensions to other probability spaces can be viewed in terms of regenerating oracles [19], i.e. each resampling action $\mathfrak{R}_{f}$ should convert the distribution of $\mu$ conditioned on $f$ into the unconditional distribution $\mu$. We provide more detail later in Section 4, but, we can summarize this crisply with our matrix notation: the resampling oracle $\Re$ is regenerating if $\mu$ is a left-eigenvector of each $\operatorname{matrix} A_{f}$, with associated eigenvalue $\mu(f)$, i.e.

$$
\begin{equation*}
\forall f \quad \mu^{\top} A_{f}=\mu(f) \cdot \mu^{\top} \tag{1}
\end{equation*}
$$

### 2.1 The new commutativity definition

The original definition of commutativity given by Kolmogorov [23] required that for every $f \nsim g \in \mathcal{F}$, there is an injective mapping from state transitions $\sigma_{1} \xrightarrow{f} \sigma_{2} \xrightarrow{g} \sigma_{3}$ to state transitions $\sigma_{1} \xrightarrow{g} \sigma_{2}^{\prime} \xrightarrow{f} \sigma_{3}$, so that $A_{f}\left[\sigma_{1}, \sigma_{2}\right] A_{g}\left[\sigma_{2}, \sigma_{3}\right]=A_{g}\left[\sigma_{1}, \sigma_{2}^{\prime}\right] A_{f}\left[\sigma_{2}^{\prime}, \sigma_{3}\right]$.

This definition is cumbersome, as well as lacking important symmetry and invariance properties. As one of the major contributions of this paper, we introduce a more natural notion of algorithmic commutativity that is also more general than the notion of [23].

- Definition 2 (Transition matrix commutativity). We say that the resampling oracle is transition matrix commutative with respect to dependence relation $\sim$ if $A_{f} A_{g}=A_{g} A_{f}$, for every $f, g \in \mathcal{F}$ such that $f \nsim g$.
- Observation 3. If the resampling oracle is commutative in the sense of [23], then it is transition matrix commutative.

Proof. Consider $f \nsim g$ and states $\sigma, \sigma^{\prime}$. By symmetry, we need to show that $A_{f} A_{g}\left[\sigma, \sigma^{\prime}\right] \leq$ $A_{g} A_{f}\left[\sigma, \sigma^{\prime}\right]$. Since $f \nsim g$, we can see that both the LHS and RHS are zero unless $\sigma \in f \cap g$.

Let $V$ denote the set of states $\sigma^{\prime \prime}$ with $A_{f}\left[\sigma, \sigma^{\prime \prime}\right] A_{g}\left[\sigma^{\prime \prime}, \sigma^{\prime}\right]>0$. By definition, there is an injective function $F: V \rightarrow \Omega$ such that $A_{f}\left[\sigma, \sigma^{\prime \prime}\right] A_{g}\left[\sigma^{\prime \prime}, \sigma^{\prime}\right]=$ $A_{g}\left[\sigma, F\left(\sigma^{\prime \prime}\right)\right] A_{f}\left[F\left(\sigma^{\prime \prime}\right), \sigma^{\prime}\right]$. Therefore, we have $\left(A_{f} A_{g}\right)\left[\sigma, \sigma^{\prime}\right]=\sum_{\sigma^{\prime \prime} \in V} A_{f}\left[\sigma, \sigma^{\prime \prime}\right] A_{g}\left[\sigma^{\prime \prime}, \sigma^{\prime}\right]=$ $\sum_{\sigma^{\prime \prime} \in V} A_{g}\left[\sigma, F\left(\sigma^{\prime \prime}\right)\right] A_{f}\left[F\left(\sigma^{\prime \prime}\right), \sigma^{\prime}\right]$.

Since $F$ is injective, each term $A_{g}[\sigma, \tau] A_{f}\left[\tau, \sigma^{\prime}\right]$ is counted at most once in this sum with $\tau=F\left(\sigma^{\prime \prime}\right)$. So $\left(A_{f} A_{g}\right)\left[\sigma, \sigma^{\prime}\right] \leq \sum_{\tau \in f} A_{g}[\sigma, \tau] A_{f}\left[\tau, \sigma^{\prime}\right]=\left(A_{g} A_{f}\right)\left[\sigma, \sigma^{\prime}\right]$.

For brevity, we say commutative to mean transition matrix commutative throughout this paper; by contrast, we refer to the previous notion as commutative in the sense of [23].

When this definition applies, we define $A_{I}$ to be the matrix $\prod_{f \in I} A_{f}$ for a stable set $I$; note that this product is well-defined (without specifying ordering of $I$ ) since the matrices $A_{f}$ all commute.

For the remainder of this paper, we assume that the resampling oracle $\mathfrak{R}$ is transition-matrix commutative unless explicitly stated otherwise.

## 3 Witness DAGs and matrix bounds

In this section, we study witness $D A G s$, a key graph structure developed in [11] for analyzing the evolution of commutative resampling oracles. The main result of this section is Lemma 5, which is a generalization of the Witness Tree Lemma described in the introduction. Notably, while our result is more general, its proof is significantly simpler. At a high level, the role of a witness DAG is to give an "explanation" of why a certain flaw appeared during the execution of the algorithm. To bound the probability that flaw $f$ appears during the algorithm execution, we simply add up the probabilities of all the witness DAGs that explain it.

Formally, we define a witness $D A G$ (abbreviated $w d a g$ ) to be a directed acyclic graph $G$, where each vertex $v \in G$ has a label $L(v)$ from $\mathcal{F}$, and such that for all pairs of vertices $v, w \in G$, there is an edge between $v$ and $w$ (in either direction) if and only if $L(v) \sim L(w)$. For a wdag $G$ with sink nodes $v_{1}, \ldots, v_{k}$, note that $L\left(v_{1}\right), \ldots, L\left(v_{k}\right)$ are all distinct and $\left\{L\left(v_{1}\right), \ldots, L\left(v_{k}\right)\right\}$ is a stable set which we denote by $\operatorname{sink}(G)$. We say that a flaw $f$ is unrelated to a wdag $G$ if there is no node $v \in G$ with $L(v) \sim f$.

We define $\mathfrak{H}$ to be the collection of all wdags, and we define $\mathfrak{H}(I)$ to be the collection of all such wdags with $\operatorname{sink}(H)=I$. With some abuse of notation, we also write $\mathfrak{H}(f)$ as shorthand for $\mathfrak{H}(\{f\})$.

There is a key connection between wdags and the transition matrices. For any wdag $H$, we define an associated $|\Omega| \times|\Omega|$ matrix $A_{H}$ inductively as follows. If $H=\emptyset$, then $A_{H}$ is the identity matrix on $\Omega$. Otherwise, we choose an arbitrary source node $v$ of $H$ and set $A_{H}=A_{L(v)} A_{H-v}$.

- Proposition 4. The definition of $A_{H}$ does not depend on the chosen source node $v$. Furthermore, there is an enumeration of the nodes of $H$ as $v_{1}, \ldots, v_{t}$ such that $A_{H}=$ $A_{L\left(v_{1}\right)} \cdots A_{L\left(v_{t}\right)}$.

Proof. Let us show the first property by induction on $|H|$. When $|H|=0$ this is vacuously true. For induction case, suppose $H$ has two source nodes $v_{1}, v_{2}$. We need to show that we get the same value by recursing on $v_{1}$ or $v_{2}$, i.e $A_{L\left(v_{1}\right)} A_{H-v_{1}}=A_{L\left(v_{2}\right)} A_{H-v_{2}}$.

We can apply the induction hypothesis to $H-v_{1}$ and $H-v_{2}$, noting that $v_{2}$ is a source node of $H-v_{1}$ and $v_{1}$ is a source node of $H-v_{2}$. We get $A_{H-v_{1}}=A_{L\left(v_{2}\right)} A_{H-v_{1}-v_{2}}, A_{H-v_{2}}=$ $A_{L\left(v_{1}\right)} A_{H-v_{1}-v_{2}}$. Thus, in order to show $A_{L\left(v_{1}\right)} A_{H-v_{1}}=A_{L\left(v_{2}\right)} A_{H-v_{2}}$, it suffices to show that $A_{L\left(v_{1}\right)} A_{L\left(v_{2}\right)}=A_{L\left(v_{2}\right)} A_{L\left(v_{1}\right)}$. Since $v_{1}, v_{2}$ are both source nodes, we have $L\left(v_{1}\right) \nsim L\left(v_{2}\right)$. Thus, this follows from commutativity.

For the second property, we have $A_{H}=A_{L\left(v_{1}\right)} A_{H-v_{1}}$ for a source node $v$. Recursively peeling away vertices gives $A_{H}=A_{L\left(v_{1}\right)} A_{L\left(v_{2}\right)} \ldots A_{L\left(v_{t}\right)}$.

As a warm-up, we first show how to use wdags to bound the number of resamplings performed in commutative algorithms. This will allow us to show bounds on the expected runtime of the Search Algorithm as well as allowing parallel implementations. The main point here is to demonstrate how the new commutativity definition helps with the crucial task of bounding the probability of appearance of a given wdag.

As in the original proof of Moser \& Tardos [25], we will estimate the expected number of times each flaw $f \in F$ is resampled. Consider an execution of the Search Algorithm with trajectory $T$. For each time $t \leq \operatorname{len}(T)$, we generate a corresponding wdag $G_{t}^{T}$ which provides the history of the $t^{\text {th }}$ resampling. Initially, we set $G_{t}^{T}$ to consist of a singleton node labeled $f_{t}$. Then, for $s=t-1, \ldots, 1$, there are two cases:

1. if wdag $G_{t}^{T}$ has any node with label $g$ where $g \sim f_{s}$, then we add a vertex labeled $f_{s}$, with a sink node to all nodes $w$ such that $L(w) \sim f_{s}$;
2. Otherwise, if $G_{t}^{T}$ is unrelated to $f_{s}$, then we do not modify $G_{t}$.

We define $G_{[s, t]}^{T}$ to be the partial wdag formed only by considering times $t, \ldots, s$; then $G_{t}^{T}=G_{[1, t]}^{T}$ and $G_{[t, t]}^{T}$ is a singleton node labeled $f_{t}$ and $G_{[s, t]}^{T}$ is formed by copying $G_{[s+1, t]}^{T}$ and adding, or not, a node labeled $f_{s}$. We say that a wdag $H$ appears if $H$ is isomorphic to $G_{t}$ for any value $t$; with a slight abuse of notation, we write this is simply as $G_{t}=H$.

To calculate the expected running time of the Search Algorithm, we sum the wdag appearance probabilities. One of the main ingredients in the original proof of Moser \& Tardos is that any wdag $G$ appears with probability at most $\prod_{v \in G} \mu(L(v))$, i.e., the product of probabilities of the flaws that label its vertices. Their proof depends on properties of the variable setting and does not extend to other probability spaces.

Our key message is that commutativity allows us to bound the probability of a wdag appearing by considering the product of transition matrices for flaws that label its vertices. Specifically, we show the following. (Recall that $\mu$ denotes the initial state distribution.)

- Lemma 5. For any wdag $H$, the probability that $H$ appears is at most $\mu^{\top} A_{H} \overrightarrow{1}$.

Proof. We first show that if the Search Algorithm runs for at most $t_{\text {max }}$ steps starting with state $\sigma$, where $t_{\text {max }}$ is an arbitrary integer, then $H$ appears with probability at most $e_{\sigma}^{\top} A_{H} \overrightarrow{1}$. We prove this claim by induction on $t_{\max }$. If $t_{\max }=0$, or $\sigma$ is flawless, the claim can be easily seen to be hold vacuously.

So suppose that $t_{\max } \geq 1$ and $S$ selects a flaw $g$ to resample in $\sigma$, and define $\mathcal{E}_{H}$ to be the event that $H$ appears when running the search algorithm $\mathcal{A}$. By conditioning on the random seed used by the flaw choice strategy $S$ (if any), we may assume that the search strategy $S$ is deterministic.

We can now view the evolution of $\mathcal{A}$ as a two-part process: we first resample $g$, reaching state $\sigma^{\prime}$ with probability $A_{g}\left[\sigma, \sigma^{\prime}\right]$. We then execute a new search algorithm $\mathcal{A}^{\prime}$ starting at state $\sigma^{\prime}$, wherein the flaw selection rule $S^{\prime}$ on history $\left(\sigma^{\prime}, \sigma_{2}, \ldots, \sigma_{t}\right)$ is the same as the choice of $S$ on history $\left(\sigma, \sigma^{\prime}, \sigma_{2}, \ldots, \sigma_{t}\right)$. Let us denote by $G_{s}^{\prime}$ the wdags produced for this new search algorithm $\mathcal{A}^{\prime}$.

Suppose that $H$ appears, so that $G_{s}=H$ for some value $s \leq t_{\max }$. In this case, we claim that one of the two conditions must hold: (i) $H$ has a unique source node $v$ labeled $g$ and $G_{s-1}^{\prime}=H-v$; or (ii) $g$ is unrelated to $H$ and $G_{s-1}^{\prime}=H$. For, suppose that $H$ has another node $w$ with $L(w) \sim L(v)$; in this case, when forming the wdag $G_{s}$, the rule would be to add a new node labeled $g$, which is perforce a source node.

In case (i), then in order for event $\mathcal{E}_{H}$ to occur on the original search algorithm $\mathcal{A}$, we must also have $\mathcal{E}_{H-v}$ hold on $\mathcal{A}^{\prime}$ within $t-1$ timesteps. By induction hypothesis, this has probability at most $e_{\sigma^{\prime}}^{\top} A_{H-v} \overrightarrow{1}$ for a fixed $\sigma^{\prime}$. Summing over $\sigma^{\prime}$ gives a total probability of $\sum_{\sigma^{\prime}} A_{g}\left[\sigma, \sigma^{\prime}\right] e_{\sigma^{\prime}}^{\top} A_{H-v} \overrightarrow{1}=e_{\sigma}^{\top} A_{g} A_{H-v} \overrightarrow{1}=e_{\sigma}^{\top} A_{H} \overrightarrow{1}$ as required.

In case (ii), then in order for event $\mathcal{E}_{H}$ to occur for $\mathcal{A}$, we must also have $\mathcal{E}_{H}$ occur for $\mathcal{A}^{\prime}$ within $t-1$ timesteps. By induction hypothesis, this has probability at most $e_{\sigma^{\prime}}^{\top} A_{H} \overrightarrow{1}$ for a fixed $\sigma^{\prime}$. Summing over $\sigma^{\prime}$ gives a total probability of $\sum_{\sigma^{\prime}} A_{g}\left[\sigma, \sigma^{\prime}\right] e_{\sigma^{\prime}}^{\top} A_{H} \overrightarrow{1}=e_{\sigma}^{\top} A_{g} A_{H} \overrightarrow{1}$. Since $A_{g}$ commutes with $A_{H}$, this is at most $e_{\sigma}^{\top} A_{H} A_{g} \overrightarrow{1}$. Since $A_{g}$ is substochastic, this in turn is at most $e_{\sigma}^{\top} A_{H} \overrightarrow{1}$, which completes the induction.

By countable additivity, we can compute the probability that $H$ ever appears from starting state $\sigma$, as $\operatorname{Pr}\left(\bigvee_{t=1}^{\infty} G_{t}^{\hat{T}}=H\right)=\lim _{t_{\max } \rightarrow \infty} \operatorname{Pr}\left(\bigvee_{t=1}^{t_{\text {max }}} G_{t}^{\hat{T}}=H\right) \leq \lim _{t_{\max } \rightarrow \infty} e_{\sigma}^{\top} A_{H} \overrightarrow{1}=$ $e_{\sigma}^{\top} A_{H} \overrightarrow{1}$.

Finally, integrating over $\tau$, gives $\sum_{\tau} \mu[\tau] e_{\tau}^{\top} A_{H} \overrightarrow{1}=\mu^{\top} A_{H} \overrightarrow{1}$.

This can be used to show a generalization of the key Witness Tree Lemma of Moser \& Tardos:

- Corollary 6. Suppose the resampling oracle is regenerating. Then, for a given wdag $H$, the probability that $H$ appears is at most $\prod_{v \in H} \mu(L(v))$.

Proof. Let $f_{1}, \ldots, f_{t}$ be the labels of the vertices in $H$, ordered from source nodes to sink nodes. We can write $A_{H}=A_{f_{1}} \cdots A_{f_{t}}$. Since $\mu$ is a left-eigenvector of every transition matrix (see Eq. (1)), we have $\mu^{\top} A_{H} \overrightarrow{1}=\mu^{\top} A_{f_{1}} \cdots A_{f_{t}} \overrightarrow{1}=\mu\left(f_{1}\right) \cdots \mu\left(f_{t}\right) \mu^{\top} \overrightarrow{1}=\mu\left(f_{1}\right) \cdots \mu\left(f_{t}\right)$.

As we have already discussed, this gives the following important corollary:

- Corollary 7. The expected number of steps of the Search Algorithm is at most $\sum_{\substack{f \in \mathcal{F} \\ H \in \mathfrak{H}(f)}} \mu^{\top} A_{H} \overrightarrow{1}$.

For example, if the resampling oracle is regenerating, then, the expected number of steps of the algorithm is at most $\sum_{f \in F} \sum_{H \in \mathfrak{H}(f)} \prod_{v \in H} \mu(L(v))$, i.e. the usual Witness Tree Lemma.

We emphasize that we are not aware of any direct proof of Corollary 6 ; it seems necessary to first show the matrix bound of Lemma 5, and then project down to scalars. As we show in Appendix A, under some natural conditions the matrix commutativity property is necessary to obtain Lemma 5.

## 4 Estimating weights of wdags

The statement of Lemma 5 in terms of matrix products is very general and powerful, but difficult for calculations. To use it effectively, we need to bound the sums of the form

$$
\sum_{H \in \mathfrak{H}} \mu^{\top} A_{H} \overrightarrow{1}
$$

There are two, quite distinct, issues that arise in this calculation. First, for a given fixed wdag $H$, we need to estimate $\mu^{\top} A_{H} \overrightarrow{1}$; second, we need to bound the sum of these quantities over $H$. The second issue is well-studied and is at the heart of the probabilistic and algorithmic conditions for the LLL. The first issue is not as familiar. Following [3], we can bound the matrix product by using a heuristic based on spectral bounds of the matrices $A_{f}$. Let us define a quantity called the charge $\gamma_{f}$ for each flaw $f$ as follows.

$$
\begin{equation*}
\gamma_{f}=\max _{\tau \in \Omega} \sum_{\sigma \in f} \frac{\mu(\sigma)}{\mu(\tau)} A_{f}[\sigma, \tau] \tag{2}
\end{equation*}
$$

The following result of [21] illustrates the connection between this measure and the Lopsided Lovász Local Lemma (LLLL):

- Theorem 8 ([21]). For each set $S \subseteq \mathcal{F}-\Gamma(f)$, there holds $\mu\left(f \mid \bigcap_{g \in S} \bar{g}\right) \leq \gamma_{f}$.

Moreover, as shown in [2], the charge $\gamma_{f}$ captures the compatibility between resampling oracle for $f$ and the measure $\mu$. A resampling oracle $\mathfrak{R}$ with $\gamma_{f}=\mu(f)$ for all $f$, is called a regenerating oracle [19], as it perfectly removes the conditional of the resampled flaw. (This is equivalent to satisfying Eq. (1).)

For a wdag $H$, let us define the scalar value $w(H)=\prod_{v \in H} \gamma_{L(v)}$. We get the following estimate for $\mu^{\top} A_{H} \overrightarrow{1}$ in terms of $w(H)$ :

- Theorem 9. For any event $E \subseteq \Omega$ we have $\mu^{\top} A_{H} e_{E} \leq \mu(E) \cdot w(H)$. In particular, with $E=\Omega$, we have $\mu^{\top} A_{H} \overrightarrow{1} \leq w(H)$.
Proof. From definition of $\gamma_{f}$, it can be observed that $\mu^{\top} A_{f} \preceq \gamma_{f} \mu^{\top}$ for any $f$. In particular, $\mu^{\top} A_{f} \cdot \theta \leq \gamma_{f} \theta$ for any vector $\theta$. Now, by Proposition 4 , we can write $A_{H}=A_{f_{1}} \ldots A_{f_{t}}$ where $f_{1}, \ldots, f_{t}$ are the labels of the nodes of $H$. We thus have:

$$
\mu^{\top} A_{H} e_{E}=\mu^{\top} A_{f_{1}} \ldots A_{f_{t}} e_{E} \leq \mu^{\top} \gamma_{f_{1}} A_{f_{2}} \ldots A_{f_{t}} \leq \cdots \leq \gamma_{f_{1}} \ldots \gamma_{f_{t}} \mu^{\top} e_{E}=w(H) \mu(E)
$$

In light of Theorem 9, we define for any stable set $I$ the key quantity $\Psi(I)=$ $\sum_{H \in \mathfrak{H}(I)} w(H)$. We also define $\Psi(f)=\Psi(\{f\})$ for brevity.

## Corollary 10.

1. Any given wdag $H$ appears with probability at most $w(H)$.
2. The expected number of resamplings of any flaw $f$ is at most $\Psi(f)$.
3. The expected runtime of the Search Algorithm is at most $\sum_{f} \Psi(f)$.

We summarize a few well-known bounds on these quantities.

## - Proposition 11.

1. (Symmetric criterion) Suppose that $\gamma_{f} \leq p$ and $|\bar{\Gamma}(f)| \leq d$ for parameters $p, d$ with $e p d \leq 1$. Then $\Psi(f) \leq e \gamma_{f} \leq e p$ for all $f$.
2. (Asymmetric criterion) Suppose there is some function $x: \mathcal{F} \rightarrow[0,1)$ with the property that $\gamma_{f} \leq x(f) \prod_{g \in \Gamma(f)}(1-x(g))$ for all $f$. Then $\Psi(f) \leq \frac{x(f)}{1-x(f)}$ for all $f$.
3. (Cluster-expansion criterion) Suppose there is some function $\eta: \mathcal{F} \rightarrow[0, \infty)$ with the property that $\eta(f) \geq \gamma_{f} \cdot \sum_{\substack{\subseteq \subseteq \bar{\Gamma}(f) \\ I \text { stable }}} \prod_{g \in I} \eta(g)$ for all $f$. Then $\Psi(f) \leq \eta(f)$ for all $f$.

A related quantity is $\bar{\Psi}(I)=\sum_{J \subseteq I} \Psi(J)$. A useful and standard formula (see e.g., [19, Claim 59]) is that for any set $I$ we have $\Psi(I) \leq \prod_{f \in I} \Psi(f)$ and $\bar{\Psi}(I) \leq \prod_{f \in I}(1+\Psi(f))$. We also write $\Psi_{\mathcal{F}}, \bar{\Psi}_{\mathcal{F}}$ to indicate the role of the flaw set $\mathcal{F}$, if it is relevant.

As an illustration, consider latin transversals. Here, we have a flaw $f$ for each pair of cells $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)$ of the same color, i.e. $\pi x_{1}=y_{1} \wedge \pi x_{2}=y_{2}$. We denote this by flaw [ $\left.\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)\right]$. We define the dependency graph by setting $f \sim f^{\prime}$ if and only if $f$ and $f^{\prime}$ are mutually incompatible, i.e. $f=\left[\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)\right], f^{\prime}=\left[\left(x_{1}^{\prime}, y_{1}^{\prime}\right),\left(x_{2}^{\prime}, y_{2}^{\prime}\right)\right]$ where either $x_{1}=x_{1}^{\prime}, y_{1} \neq y_{1}^{\prime}$ or $x_{1} \neq x_{1}^{\prime}, y_{1}=y_{1}^{\prime}$. We will examine this construction in more detail later in Section 6.

- Proposition 12. Suppose that each color appears at most $\Delta=\frac{27}{256} n$ times in the array. Then the expected number of steps of the Search Algorithm is $O(n)$. Furthermore, $\Psi(f) \leq \frac{256}{81 n^{2}}$ for each $f$.

Proof. We apply the cluster-expansion criterion with $\eta(f)=\frac{256}{81 n^{2}}$ for each flaw $f$. Consider a flaw $f$ corresponding to cells $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)$, and a stable set $I$ of neighbors of $f$.. There can be one or zero elements $g$ of $I$ of the form $\left[\left(x_{1}, y_{1}^{\prime}\right),\left(x_{2}^{\prime}, y_{2}^{\prime}\right)\right]$. There are $n$ choices for $x_{1}$; given that pair $\left(x_{1}, y_{1}^{\prime}\right)$ is determined, there are at most $\Delta-1$ other cells with the same color. Each such $g$ has $\eta(g)=\frac{256}{81 n^{2}}$. Similar arguments apply to elements in $I$ of the form $\left[\left(x_{1}^{\prime}, y_{1}\right),\left(x_{2}^{\prime}, y_{2}^{\prime}\right)\right]$ etc. Overall, the sum over stable neighbor sets $I$ is at most $\left(1+n(\Delta-1) \frac{256}{81 n^{2}}\right)^{4}$.

So we need to show that

$$
\frac{256}{81 n^{2}} \geq \frac{1}{n^{2}} \cdot\left(1+n(\Delta-1) \frac{256}{81 n^{2}}\right)^{4}
$$

and simple calculations show that this holds for $n \geq 2$. (The case $n=1$ holds trivially).
Also, the total number of flaws is at most $n^{2}(\Delta-1) / 2=O\left(n^{3}\right)$. Thus, the expected number of steps is at most $|\mathcal{F}| \cdot \frac{256}{81 n^{2}} \leq O(n)$.

## 5 Parallel algorithms

Moser \& Tardos [25] described a simple parallel version of their resampling algorithm. A variety of parallel resampling algorithms have also been developed for other probability spaces $[17,13]$. One main benefit of the commutativity property is that it enables much more general parallel implementations of the Search Algorithm. As a starting point, [23] discussed a generic framework for parallelization which we summarize as follows:

Algorithm 3 Generic parallel resampling framework.

```
Draw state }\sigma\mathrm{ from distribution }
while some flaw holds on }\sigma\mathrm{ do
    Set V\not=\emptyset to be the set of flaws currently holding on \sigma
    while }V\not=\emptyset\mathrm{ do
            Select, arbitrarily, a flaw }f\inV\mathrm{ .
            Update }\sigma\leftarrow\mp@subsup{\mathfrak{R}}{\sigma}{}(\sigma)\mathrm{ .
            Remove from V all flaws g such (i) }\sigma\not\ing\mathrm{ ; or (ii) f }~
```

Each iteration of the main loop (lines $2-7$ ) is called a round. We emphasize this is a sequential algorithm, which can be viewed as a version of the Search Algorithm with an unusual flaw-selection choice. Most known parallel local search algorithms, including the original parallel algorithm of Moser \& Tardos, fall into this framework. One main result of [23] is that, when the resampling oracle is commutative (in the sense of [23]), then the total number of rounds in Algorithm 3 is polylogarithmic with high probability.

Harris [15] further showed a general method for simulating each round in parallel, for resampling oracles which satisfy a property called obliviousness (see Section 7 for a formal definition). These two results combine to give an overall RNC search algorithm. We will now extend these results to our commutative resampling oracles, via bounding the weights of certain classes of wdags.

We define $V_{k}$ to be the set of flaws $V$ in round $k$, and we define the stable set $I_{k}$ to be the set of flaws which are actually resampled at round $k$ (i.e. a flaw $f$ selected at some iteration of line 5). Let $b_{k}=\sum_{i<k}\left|I_{i}\right|$ be the total number of resamplings made before round $k$; thus $b_{1}=0$, and when "serialize" Algorithm 3 and view it as an instance of the Search Algorithm, the resamplings in round $k$ of Algorithm 3 correspond to the resamplings at iterations $b_{k}+1, \ldots, b_{k+1}$ of the Search Algorithm.

- Proposition 13. For all $f \in V_{k}$ there exists $g \in I_{k-1}$ with $f \sim g$.
- Proposition 14. Consider running Algorithm 3 obtaining trajectory $\hat{T}$. Then, for each $t$ in the range $b_{k}+1, \ldots, b_{k}$ the wdag $G_{t}^{\hat{T}}$ has depth precisely $k$.

The proof of Propositions 13 and 14 are nearly identical to results for the variable LLLL shown in [15]; we omit then here.

```
- Proposition 15. For any \(f \in \mathcal{F}\) and index \(k \geq 1\), we have \(\operatorname{Pr}\left(f \in V_{k}\right) \leq\)
\(\sum_{\substack{H \in \mathfrak{H}(f) \\ \operatorname{depth}(H)=k}} \mu^{\top} A_{H} \overrightarrow{1}\).
```

Proof. As we have discussed, Algorithm 3 can be viewed as an instantiation of the Search Algorithm with a flaw selection rule $S$. For a fixed $f$, let us define a new flaw selection rule $S_{f}$ as follows: it agrees with $S$ up to round $k$; it then selects $f$ to resample at round $k$ if it is true. The behavior for $S$ and $S_{f}$ is identical up through the first $b_{k-1}$ resamplings. Furthermore, Algorithm 3 has $f \in V_{k}$ if and only if the Search Algorithm selects $f$ for resampling at iteration $b_{k}+1$.

Consider the resulting wdag $G_{t}^{\hat{T}}$; by Proposition 14 it has depth $k$. Furthermore, it has a sink node labeled $f$. Thus, if $f \in V_{k}$, then there is some $H \in \mathfrak{H}(f)$ with $\operatorname{depth}(H)=k$ which appears. To bound the probability of $f \in V_{k}$, we take a union bound over all such $H$ and apply Lemma 5.

The usual strategy to bound the sum over wdags $H$ with $\operatorname{depth}(H) \geq t$ is to use an "inflated" weight function defined as $w_{\epsilon}(H)=w(H)(1+\epsilon)^{|H|}$, and corresponding sum $W_{\epsilon}=\sum_{H \in \mathfrak{H}} w_{\epsilon}(H)$ for some $\epsilon>0$. Using standard calculations as well as the bounds of Propositions $13,14,15$, one can show the following results:

- Proposition 16. With probability at least $1-\delta$, Algorithm 3 terminates in $O\left(\frac{\log \left(1 / \delta+\epsilon W_{\epsilon}\right)}{\epsilon}\right)$ rounds and has $\sum_{k}\left|V_{k}\right| \leq O\left(W_{\epsilon} / \delta\right)$. Furthermore, if the resampling oracle is regenerating and satisfies the computational requirements given in [15] for input length $n$, then with probability $1-1 / \operatorname{poly}(n)$ the algorithm of [15] terminates in $O\left(\frac{\log ^{4}\left(n+\epsilon W_{\epsilon}\right)}{\epsilon}\right)$ time on an EREW PRAM.

Bounding $W_{\epsilon}$ is very similar to bounding $\sum_{H} w(H)=W_{0}$, except with a small "slack" in the charges. For example, using standard estimates (see [11, 23, 3]) we get the following bounds:

## - Proposition 17.

1. Suppose that the resampling oracle is regenerating and that the vector of probabilities $p(1+\epsilon)$ still satisfies the LLLL criterion. Then $W_{\epsilon / 2} \leq O(m / \epsilon)$. In particular, Algorithm 3 terminates after $O\left(\frac{\log (m / \delta)}{\epsilon}\right)$ rounds with probability $1-\delta$.
2. Suppose that $\gamma_{f} \leq p$ and $|\bar{\Gamma}(f)| \leq d$ such that epd $(1+\epsilon) \leq 1$. Then $W_{\epsilon / 2} \leq O(m / \epsilon)$. Algorithm 3 terminates after $O\left(\frac{\log (m / \delta)}{\epsilon}\right)$ rounds with probability at least $1-\delta$.

## 6 Distributional properties

The most important consequence of commutativity is that it leads to good bounds on the distribution of objects generated by the Search Algorithm. Consider an event $E$ in $\Omega$, and define $P(E)$ to be the probability that $E$ holds in the output of the Search Algorithm. Also define $N(E)$ to be the expected number of times $t$ such that $E$ is caused to be true at time $t$; this includes both the original sampling at time $t=0$, or if resampling flaw $f$ at time $t$ moved the state from $\bar{E}$ to $E$. Clearly, there holds $P(E) \leq N(E)$. We also write $P_{\mathcal{F}}(E)$ and $N_{\mathcal{F}}(E)$ to emphasize the dependence on flaw set $\mathcal{F}$.

To obtain the tightest bounds on $N(E)$, and thereby $P(E)$, we will use a more refined construction of wdags. For this we need several definitions.

We say that a stable set $I \subseteq \mathcal{F}$ is orderable for $E$ if there is an enumeration $I=\left\{g_{1}, \ldots, g_{r}\right\}$ such that

$$
\begin{equation*}
\forall i=1, \ldots, r \quad A_{g_{i}} A_{g_{i+1}} \ldots A_{g_{r}} e_{E} \npreceq A_{g_{i+1}} \ldots A_{g_{r}} e_{E} \tag{3}
\end{equation*}
$$

We define $\mathfrak{O}(E)$ to be the collection of stable sets orderable for $E$. Also, we define $\tilde{\Gamma}(E)$ to be the set of flaws $f \in \mathcal{F}$ which can cause $E$ to occur, i.e. $f$ maps some state $\sigma^{\prime} \notin E$ to $\sigma \in E$.

- Observation 18. If $I \in \mathfrak{O}(E)$, then $I \subseteq \tilde{\Gamma}(E)$.

With this notation, our main distributional bound will be to show that

$$
N(E) \leq \mu(E) \sum_{I \in \mathfrak{O}(E)} \Psi(I)
$$

For a flaw $f$ and wdag $H$, we say that $f$ is dominated by a wdag $H$ for $E$ if $A_{f} A_{H} \overrightarrow{1} \preceq A_{H} \overrightarrow{1}$. Consider a trajectory $T=\left(\sigma_{0}, f_{1}, \ldots,\right)$. For each time $t$ where $E$ holds (including possibly $t=0$ ), we will generate a corresponding wdag $J_{t}^{T}$, however, the rule for adding nodes is slightly more restrictive. See Algorithm 4 for the precise construction.

```
Algorithm 4 Forming \(J_{t}^{T}\).
    Initialize \(J_{t}^{T}=\emptyset\)
    for \(s=t, \ldots, 1\) do
        if \(f_{s}\) is not dominated by \(J_{t}^{T}\) or if \(J_{t}^{T}\) has a source node labeled \(f_{s}\) then
            Add to \(J_{t}^{T}\) a node \(v_{s}\) labeled \(f_{s}\), with an edge from \(v_{s}\) to each \(v_{j}\) such that \(f_{j} \sim f_{s}\)
```

We write $J_{[s, t]}^{T}$ for the wdag $J_{t}^{T}$ after iteration $s$, so that $J_{[s, t]}^{T}$ is derived from $J_{[s+1, t]}^{T}$ by adding (or not) a vertex labeled $f_{s}$. We have $J_{t}^{T}=J_{[1, t]}^{T}$ and $J_{[t+1, t]}^{T}=\emptyset$. Also, if $E$ is not true at time $t$, we define $J_{t}^{T}=\perp$. We define $\mathfrak{H}^{\prime}$ to be the collection of all wdags that can be produced in this process.

- Proposition 19. Consider a wdag $G \in \mathfrak{H}^{\prime}$. If $J_{t}^{T}=G$ for a trajectory $T$ with $t \geq 1$ and $f_{1}$ resampled at time 1, then the wdag $G^{\prime}=J_{t-1}^{T^{\prime}}$ for the shifted trajectory $T^{\prime}$ is uniquely determined according to the following rule:
- If $G$ contains a unique source node $v$ labeled $f_{1}$, then $G^{\prime}=G-v$
- Otherwise, $G^{\prime}=G$ and $f_{1}$ is dominated by $G_{t}^{T}$

Proof. Since $t \geq 1$, Algorithm 4 obtains $J_{t}^{T}$ by possibly adding a node $v_{1}$ labeled $f_{1}$ to $J_{t-1}^{T^{\prime}}$. If Algorithm 4 adds node $v_{1}$ to $G^{\prime}$, then $f_{1}$ is the label of a source node $v$ of $J_{t}^{T}=G$, and $G^{\prime}=J_{t}^{T}-v$. If Algorithm 4 does not add such node, then $J_{t}^{T}=G^{\prime}$. By the rule for adding nodes, it must be that $G^{\prime}$ does not have a source node labeled $f_{1}$, and also $f_{1}$ must be dominated by $J_{t-1}^{T^{\prime}}$. Since $G^{\prime}=J_{t}^{T}$, these imply that $f_{1}$ is dominated by $J_{t}^{T}=G$ as well.

Our main result for this construction will be the following:

- Lemma 20. For any wdag $H \in \mathfrak{H}^{\prime}$, there holds $\operatorname{Pr}\left(\bigvee_{t=0}^{\infty} J_{t}=H\right) \leq \mu^{\top} A_{H} e_{E}$

Proof. Define $\mathcal{E}_{H, t_{\max }}$ to be the event that $J_{t}^{\hat{T}}=H$ holds for some $t \leq t_{\text {max }}$ during execution of the Search Algorithm $\mathcal{A}$. By a limiting arugment, it suffices to show that $\operatorname{Pr}\left(\mathcal{E}_{H, t_{\text {max }}}\right) \leq$ $\mu^{\top} A_{H} e_{E}$ for any integer $t_{\max } \geq 1$. We will prove by induction on $t_{\max }$ that, if we start at any state $\sigma$, then $\operatorname{Pr}\left(\mathcal{E}_{H, t_{\text {max }}}\right) \leq e_{\sigma}^{\top} A_{H} e_{E}$; the Lemma then follows by integrating over starting state $\sigma$.

If $t_{\max }=0$ or $\sigma$ is flawless, then $\mathcal{E}_{H, t_{\max }}$ is impossible and the desired bound. So suppose that $t_{\max } \geq 1$, and that $S$ selects a flaw $g$ to resample in $\sigma$. We can now view the evolution of $\mathcal{A}$ as a two-part process: we first resample $g$, reaching state $\sigma^{\prime}$ with probability $A_{g}\left[\sigma, \sigma^{\prime}\right]$. We then execute a new search algorithm $\mathcal{A}^{\prime}$ starting at state $\sigma^{\prime}$, wherein the flaw selection rule $S^{\prime}$ on history $\left(\sigma^{\prime}, \sigma_{2}, \ldots, \sigma_{r}\right)$ is the same as the choice of $S$ on history $\left(\sigma, \sigma^{\prime}, \sigma_{2}, \ldots, \sigma_{r}\right)$.

Suppose now that $\mathcal{E}_{H, t_{\max }}$ holds for $\mathcal{A}$, i.e. $J_{t}^{\hat{T}}=H$ for some $t \leq t_{\max }$. Note that the actual trajectory $\hat{T}^{\prime}$ for $\mathcal{A}^{\prime}$ is given by $\hat{T}^{\prime}$ which is the shift of $\hat{T}$. Thus, by Proposition 19, one of the two conditions must hold: (i) either $H$ has a unique source node $v$ labeled $g$ and $J_{t-1}^{\hat{T}^{\prime}}=H-v$; or (ii) $H$ has no such node and $J_{t-1}^{\hat{T}^{\prime}}=H$ and $g$ is dominated by $H$.

In the first case, there must also hold $\mathcal{E}_{H-v, t_{\max }-1}$ for $\mathcal{A}^{\prime}$. By induction hypothesis, this has probability at most $e_{\sigma^{\prime}}^{\top} A_{H} e_{E}$ conditional on a fixed $\sigma^{\prime}$. Summing over $\sigma^{\prime}$, we get a total probability of $\sum_{\sigma^{\prime}} A_{g}\left[\sigma, \sigma^{\prime}\right] e_{\sigma^{\prime}}^{\top} A_{H-v} e_{E}=e_{\sigma}^{\top} A_{g} A_{H-v} e_{E}=e_{\sigma}^{\top} A_{H} e_{E}$.

In the second case, there must also hold $\mathcal{E}_{H, t_{\max }-1}$ for $\mathcal{A}^{\prime}$. By induction hypothesis, this has probability at most $e_{\sigma^{\prime}}^{\top} A_{H} e_{E}$ conditional on a fixed $\sigma^{\prime}$. Summing over $\sigma^{\prime}$, we get a total probability of $\sum_{\sigma^{\prime}} A_{g}\left[\sigma, \sigma^{\prime}\right] e_{\sigma^{\prime}}^{\top} A_{H} e_{E}=e_{\sigma}^{\top} A_{g} A_{H} e_{E}$. Since $g$ is dominated by $H$ for $E$, this is at most $e_{\sigma}^{\top} A_{H} e_{E}$, again completing the induction.

- Proposition 21. Suppose that event $E$ is true at times $s$ and $t$ with $s<t$, but false at some intermediate time. Then $J_{s}^{T} \neq J_{t}^{T}$.

Proof. We prove this by induction on $s$. For the base case $s=0$, we have $J_{s}^{T}=\emptyset$. Suppose for contradiction that $J_{t}^{T}=\emptyset$ as well. Since $E$ is false at an intermediate time but true at time $t$, it must become true due to resampling some $g$ at time $t^{\prime}<t$. Clearly also $J_{\left[t^{\prime}+1, t\right]}^{T}=\emptyset$. Since $g$ makes $E$ be true where it was false, we have $g \in \tilde{\Gamma}(E)$. As a result, $g$ is not dominated for the empty wdag. So the rule for forming $J_{t}^{T}$ would add a node to $J_{\left[t^{\prime}, t\right]}^{T}$, contradicting that $J_{t}^{T}=\emptyset$.

For the induction step, suppose $s>0$ and $J_{t}^{T}=J_{s}^{T}$. Let $T^{\prime}$ be the shift of $T$. By Proposition 19, both $J_{s-1}^{T^{\prime}}$ and $J_{t-1}^{T^{\prime}}$ are updated in the same manner depending on the flaw $f_{1}$. Thus, $J_{s-1}^{T^{\prime}}=J_{t-1}^{T^{\prime}}$. But this contradicts the induction hypothesis.

- Proposition 22. $N(E) \leq \sum_{H \in \mathfrak{H}^{\prime}} \mu^{\top} A_{H} e_{E}$.

Proof. By Proposition 21, for each time $t$ that $E$ switches from false to true, the corresponding wdag $J_{t}^{T}$ must be distinct. Thus, the total number of times that $E$ becomes true is at most $\sum_{H \in \mathfrak{H}^{\prime}}\left[\left[\bigvee_{t \geq 0} J_{t}^{\hat{t}}=H\right]\right]$. Now take expectations and apply Lemma 20.

- Proposition 23. For a wdag $H \in \mathfrak{H}^{\prime}$, there holds $\operatorname{sink}(H) \in \mathfrak{O}(E)$.

Proof. Let us fix some value $t$ where $E$ holds; for $s=1, \ldots, t$ let $H_{s}=J_{[s, t]}^{T}$ and $I_{s}=\operatorname{sink}\left(H_{s}\right)$, and so that $H_{1}=H$. We claim by induction on $s$ that each wdag $H_{s}$ has the stated property. The base case is $s=t$; this holds since $I_{t}=\emptyset \in \mathfrak{O}(E)$.

Now consider some $s<t$, where $H_{s}$ is formed from $H_{s+1}$ by possibly adding a new node labeled $g$. The induction step is obvious if $I_{s}=I_{s+1}$. Thus, we may assume that $g$ is unrelated to $H_{s+1}$ (else it would not form a new sink node.) So the only relevant case is if $g$ is not dominated by wdag $G_{[s, t]}^{T}$.

By induction hypothesis, $I_{s+1} \in \mathfrak{O}(E)$. Then it can be enumerated as $I_{s+1}=\left\{g_{1}, \ldots, g_{r}\right\}$ to satisfy Eq. (3). Suppose, for contradiction, that $I_{s} \notin \mathfrak{O}(E)$. If we enumerate $I_{s}=\left\{g, g_{1}, \ldots, g_{r}\right\}$, then there would hold $\theta^{\top} A_{g} A_{g_{1}} \ldots A_{g_{r}} e_{E} \leq \theta^{\top} A_{g_{1}} \ldots A_{g_{r}} e_{E}$ for all distributions $\theta$ over the states.

Letting $V$ denote the sink nodes of $H_{s+1}$, we have $A_{H_{s+1}}=A_{H_{s+1}-V} A_{V}$. Then, for any state $\sigma$, we have $e_{\sigma}^{\top} A_{g} A_{H_{s+1}} e_{E}=e_{\sigma}^{\top} A_{g} A_{H_{s+1}-V} A_{V} e_{E}=e_{\sigma} A_{H_{s+1}-V} A_{g} A_{g_{1}} \ldots A_{g_{r}} e_{E}$, where in the last inequality we use the facts that $g$ is unrelated to $H_{s+1}$ and that $A_{V}=A_{g_{1}} \ldots A_{g_{r}}$. By our bound with $\theta=e_{\sigma}^{\top} A_{H_{s+1}-V}$, we see that this is at most $e_{\sigma} A_{H_{s+1}-V} A_{g_{1}} \ldots A_{g_{r}} e_{E}=$ $e_{\sigma} A_{H_{s+1}} e_{E}$. Hence $g$ is dominated by $G_{s+1}$ and we would not add the new node to $H_{s}$, a contradiction.

We now get our desired distributional bounds:

- Corollary 24. $N(E) \leq \mu(E) \sum_{I \in \mathfrak{O}(E)} \Psi(I)$.

Proof. We have shown $N(E) \leq \sum_{H \in \mathfrak{H}^{\prime}} \mu^{\top} A_{H} e_{E}$. Since each $H \in \mathfrak{H}^{\prime}$ has $\operatorname{sink}(H) \in$ $\mathfrak{O}(E)$, this is at most $\sum_{I \in \mathfrak{O}(E)} \sum_{H \in \mathfrak{H}(I)} \mu^{\top} A_{H} e_{E}$. By Theorem 9, this is at most $\sum_{I \in \mathfrak{O}(E), H \in \mathfrak{H}(I)} w(H) \mu(E)=\mu(E) \sum_{I \in \mathfrak{O}(E)} \Psi(I)$.

- Corollary 25. $P(E) \leq \mu(E) \sum_{I \in \mathfrak{O}(E)} \Psi_{\mathcal{G}}(I)$ where $\mathcal{G}=\{f \in \mathcal{F}: f \nsubseteq E\}$.

Proof. Consider the first time that $E$ becomes true, if any. Then, only flaws in $\mathcal{G}$ can be resampled up to that point; if some other flaw with $f \subseteq E$ was resampled, then necessarily $E$ was true earlier. Up to this time, the behavior of the Search Algorithm is identical to if had restricted to the flaw set $\mathcal{G}$. So $P(E) \leq N_{\mathcal{G}}(E)$; by Corollary 24 this is at most $\sum_{I \in \mathfrak{Q}(E)} \Psi_{\mathcal{G}}(I)$.

Since any set $I \in \mathfrak{O}(E)$ is also a subset of $\tilde{\Gamma}(E)$, we have the following crisp corollary:

- Corollary 26. $P(E) \leq \mu(E) \bar{\Psi}(\tilde{\Gamma}(E))$.

We note that Iliopoulos [20] had previously shown a bound similar to Corollary 26, but it had three additional technical restrictions: (i) it only worked for commutative resampling oracles in the sense of [23]; (ii) it additionally required the construction of a commutative resampling oracle for the event $E$ itself; and (iii) if the resampling oracle is not regenerating, it gives a strictly worse bound.

The following result shows how to apply these bounds with common LLL criteria.

- Proposition 27. Under the criteria of Proposition 11, we have the following estimates for $P(E)$ :

1. If the symmetric criterion holds, then $P(E) \leq \mu(E) \cdot e^{e|\tilde{\Gamma}(E)| p}$.
2. If function $x$ satisfies the asymmetric criterion, then $P(E) \leq \mu(E) \cdot \prod_{f \in \tilde{\Gamma}(E)} \frac{1}{1-x(f)}$.
3. If function $\eta$ satisfies the cluster-expansion criterion, then $P(E) \leq \mu(E)$. $\sum_{I \in \mathfrak{O}(E)} \prod_{g \in I} \eta(g)$.

One weakness of distributional bounds such as Corollary 26 is that the definition of $\tilde{\Gamma}(E)$ is binary: either flaw $f$ cannot possibly cause $E$, or every occurrence of $f$ must be tracked to determine if it caused $E$. The next results allow us to take account of flaws which can "partially" cause $E$.

For flaw $f$ and event $E$, let us define

$$
\kappa(f, E)=\max _{\sigma \in f \cap \bar{E}} \frac{e_{\sigma}^{\top} A_{f} e_{E}}{e_{\sigma}^{\top} A_{f} e_{E \cup \bar{f}}}
$$

Note that $\kappa(f, E)=0$ for $f \notin \tilde{\Gamma}(E)$, and $\kappa(f, E) \leq 1$ always. Thus, $\kappa(f, E)$ is a weighted measure of the extent to which $f$ causes $E$. Also note that usually $e_{\sigma}^{\top} A_{f} e_{f}$ is small, and the denominator in the definition of $\kappa(f, E)$ is close to one.

- Theorem 28. $P(E) \leq \mu(E)+\sum_{f \in \mathcal{G}} \kappa(f, E) \cdot \min _{F \supseteq \bar{E}} N_{\mathcal{G}}(F \cap f)$ where $\mathcal{G}=\{f: f \nsubseteq E\}$. Proof. See Appendix B.

We remark that to obtain Theorem 28, we needed to bound $N(F \cap f)$; bounds on $P(F \cap f)$ alone would not have been enough. This explains why we analyzed the more general quantity. By applying Theorem 28 to the event $\bar{E}$, we can obtain a lower bound on the probability of E:

- Corollary 29. $P(E) \geq \mu(E)-\sum_{f \in \mathcal{G}} \kappa(f, \bar{E}) \cdot \min _{F \subseteq E} N_{\mathcal{G}}(f-E)$ where $\mathcal{G}=\{f: f \cap E \neq \emptyset\}$.

For example, consider the permutation setting, where the probability space $\Omega$ is the uniform distribution on permutations on $n$ letters, and each flaw has the form $g_{1} \cap \cdots \cap g_{k}$, where each $g_{i}$ is an atomic event of the form $\pi x_{i}=y_{i}$. We then get the following distributional result:

- Theorem 30 ([14]). In the permutation setting, consider an event $E=g_{1} \cap \cdots \cap g_{k}$ where each $g_{i}$ is an atomic event. We have $N(E) \leq \frac{(n-k)!}{n!} \prod_{i=1}^{k}\left(1+\sum_{f \in \mathcal{F}: f \sim g_{i}} \Psi(f)\right)$.

As another example, consider the setting where the underlying probability space $\Omega$ is the uniform distribution on perfect matchings on the clique $K_{n}$, and each flaw has the form $g_{1} \cap \cdots \cap g_{k}$, where each $g_{i}$ is an atomic event of the form $\left\{x_{i}, y_{i}\right\} \subseteq M$. We then get the following distributional result:

- Theorem 31. In the settings of perfect matchings of the clique, consider an event $E=$ $g_{1} \cap \cdots \cap g_{k}$ where each $g_{i}$ is an atomic event. We have $N(E) \leq \frac{(n-2 k-1)!!}{(n-1)!!} \prod_{i=1}^{k}(1+$ $\left.\sum_{f \in \mathcal{F}: f \sim g_{i}} \Psi(f)\right)$.

The work [14] showed (what is essentially) Theorem 30 using a complicated and ad-hoc analysis based on a variant of witness trees, while Theorem 31 is new. The proofs are deferred to Appendix C.

Using these bounds, we can show the following estimates on individual entries of $\pi$ :

- Theorem 32. If each color appears at most $\Delta=\frac{27}{256} n$ times in the array, then the Search Algorithm generates a latin transversal where, for each cell $x, y$, there holds

$$
\frac{17}{32 n} \leq P(\pi x=y) \leq \frac{173}{128 n}
$$

Proof. Define the event $E$ that $\pi x=y$. For the upper bound, Theorem 28 (with $F=\Omega$ ) gives $P(\pi x=y) \leq \mu(E)+\sum_{f \in \mathcal{F}} N_{\mathcal{F}}(f) \kappa(f, E)$. Now, consider some flaw $f=\left[\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)\right] \in$ $\mathcal{F}$. The flaw $f$ must involve cell $\left(x, y^{\prime}\right)$ or $\left(x^{\prime}, y\right)$, else $\eta(f, E, \Omega)=0$. For a flaw that does so, we can see that there is a probability of at most $1 /(n-1)$ that the resampling causes $E$, since there is at most one possible choice that can cause $\pi x=y$. Thus $\eta(f, E, \Omega) \leq \frac{1 /(n-1)}{1-1 / n(n-1)}$. By Theorem 30, we have $N_{\mathcal{F}}(f) \leq \frac{(n-2)!}{n!}\left(1+\sum_{f^{\prime} \in \mathcal{G}: f^{\prime} \sim g_{i}} \Psi\left(f^{\prime}\right)\right)$ where $g_{1}, g_{2}$ are the two atoms in $f$. Since $\Psi\left(f^{\prime}\right) \leq \gamma=\frac{256}{81 n^{2}}$, and there are at most $2 n(\Delta-1)$ choices for $f^{\prime}$, this is overall at most $\frac{1}{n(n-1)}(1+2 n(\Delta-1) \gamma)$.

Since either $x_{1}=x$ or $y_{1}=y$, there are $2 n(\Delta-1)$ choices for $f$. Summing over these, we get

$$
P(E) \leq \frac{1}{n}+2 n(\Delta-1) \cdot \frac{1}{n(n-1)}(1+2 n(\Delta-1) \gamma) \cdot \frac{1 /(n-1)}{1-1 / n(n-1)} \leq \frac{173}{128 n}
$$

For the lower bound, we use Corollary 29. Letting $\mathcal{G}$ denote the flaws which do not involve cells $\left(x, y^{\prime}\right)$ for $y^{\prime} \neq y$, or $\left(x^{\prime}, y\right)$ for $x^{\prime} \neq x$ and setting $F=E$, we have $P(E) \geq$ $\mu(E)-\sum_{f \in \mathcal{F}} N(E \cap f) \kappa(f, \bar{E})$. Now, consider some such flaw $f=\left[\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)\right]$. If $\left(x_{1}, y_{2}\right)=(x, y)$, then in this case, $f \cap E=f$ and so Theorem 30 implies that $N_{\mathcal{G}}(f \cap E) \leq$ $\frac{1}{n(n-1)}(1+(\Delta-1) \gamma)(1+2 n(\Delta-1) \gamma)$. (We emphasize that, because we are restricting to $\mathcal{G}$, there are no neighbors which involve cells $\left(x, y^{\prime}\right)$ etc.) Otherwise, if $(x, y)$ is distinct from $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right)$, then $f \cap E=\left[\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right),(x, y)\right]$ and Theorem 30 implies that $N_{\mathcal{G}}(f \cap E) \leq \frac{1}{n(n-1)(n-2)}(1+(\Delta-1) \gamma)(1+2 n(\Delta-1) \gamma)^{2}$.

There are at most $(\Delta-1)$ flaws in the first category, and each trivially has $\kappa(f, \bar{E}) \leq 1$. There are at most $n^{2}(\Delta-1) / 2$ flaws in the second category; each such flaw $f$ has $\kappa(f, \bar{E}) \leq$ $\frac{2 / n}{1-1 / n(n-1)}$, since there are two choices for the cell to swap and in each case there is at most one way to get $\pi x=y$ in a swap.

Putting all terms together, and with some algebraic simplifications, we get $P(E) \geq$ $\frac{17}{32 n}$.

An analogous result can be shown for perfect matchings of the clique; we omit the proof here.

- Theorem 33. Consider an edge-coloring $C$ of the clique $K_{n}$, for $n$ an even integer, such that each color appears on at most $\Delta=\frac{27}{256} n$ edges. Then the Search Algorithm generates a perfect matching $M$ such that $C(e) \neq C\left(e^{\prime}\right)$ for all distinct edges e, $e^{\prime}$ of $M$. Moreover, for each edge $e$, the probability there holds $\frac{17}{32(n-1)} \leq P(e \in M) \leq \frac{173}{128(n-1)}$.


## 7 Compositional properties for resampling oracles

The flaws and their resampling oracles are often built out of a collection of simpler, "atomic" events. For example, in the permutation LLL setting, these would be events of the form $\pi x=y$. In [15], Harris described a generic construction when the atomic events satisfy an additional property referred to as obliviousness. Let us now review this construction, and how it works with commutativity.

Consider a set $\mathcal{A}$ of events, along with a resampling oracle $\mathfrak{R}$ and a dependency relation $\sim$. It is allowed, but not required, to have $f \sim f$ for $f \in \mathcal{A}$. For the compositional construction, we must define explicitly how the resampling oracle $\mathfrak{R}_{f}$ uses the random seed. Specifically, to resample $\sigma^{\prime} \leftarrow \mathfrak{R}_{f}(\sigma)$, we first draw a random seed $r$ from some probability space $R_{f}$, and then set $\sigma^{\prime}=F(\sigma, r)$ for some deterministic function $F$. For brevity, we write this as $\sigma^{\prime}=r \sigma$.

We refer to the elements of $\mathcal{A}$ as atoms. These should be thought of as "pre-flaws", that is, they have the structural algebraic properties of a resampling oracle, but do not necessarily satisfy any convergence condition such as the LLLL. We have the following key definition:

- Definition 34 (Oblivious resampling oracle [15]). The resampling oracle $\mathfrak{R}$ is called oblivious if for every pair $f, g \in \mathcal{A}$ with $f \nsim g$ and for each $r \in R_{f}$, one of the following two properties holds:
- For all $\sigma \in f \cap g$ we have $r \sigma \in g$
- For all $\sigma \in f \cap g$ we have $r \sigma \notin g$

We assume throughout this section that $\mathfrak{R}$ is oblivious. For each $f \in \mathcal{A}$ and $g_{1}, \ldots, g_{s} \in \mathcal{A}$ with $g_{i} \nsim f$, we define $R_{f ; g_{1}, \ldots, g_{s}}$ to be the set of values $r \in R_{f}$ such that $r \sigma \in g_{1} \cap \cdots \cap g_{t}$. With some abuse of notation, we also use $R_{f ; g_{1}, \ldots, g_{s}}$ to refer to the probability distribution of drawing $r$ from $R_{f}$, conditioned on having $r$ in the set $R_{f ; g_{1}, \ldots, g_{s}}$. Note that in light of Definition 34 this is well-defined irrespective of $\sigma$.

For a stable set $C \subseteq \mathcal{A}$, we define $\langle C\rangle$ to be the intersection of the events in $C$, i.e., $\langle C\rangle=\bigcap_{f \in C} f$. From $\mathcal{A}$, one can construct an enlarged set of events $\overline{\mathcal{A}}=\{\langle C\rangle \mid$ $C$ a stable subset of $\mathcal{A}\}$. We define the relation $\sim$ on $\overline{\mathcal{A}}$ by setting $\langle C\rangle \sim\left\langle C^{\prime}\right\rangle$ iff either (i) $C=C^{\prime}$ or (ii) there exist $f \in C, f^{\prime} \in C^{\prime}$ with $f \sim f^{\prime}$. We also define a corresponding resampling oracle $\mathfrak{R}$ on $\overline{\mathcal{A}}$ which will satisfy all its required structural properties. The intent is to choose the flaw set $\mathcal{F}$ to be some arbitrary subset of $\overline{\mathcal{A}}$.; as before, $\overline{\mathcal{A}}$ does not necessarily satisfy any LLLL convergence criterion.

To determine $\mathfrak{R}$, consider some $g=\langle C\rangle$ for a stable set $C$, with some arbitrary enumeration $C=\left\{f_{1}, \ldots, f_{t}\right\}$. We define $R_{g}$ to be the probability distribution on tuples $r=\left(r_{1}, \ldots, r_{t}\right)$ wherein each $r_{i}$ is drawn independently from $R_{f_{i} ; f_{i+1}, \ldots, f_{s}}$, and we set $r \sigma=r_{t} \ldots r_{1} \sigma$.

- Theorem 35 ([15]). Suppose that $\mathfrak{R}$ is an oblivious resampling oracle for $\mathcal{A}$, which is not necessarily commutative. Then: $\mathfrak{R}$ with dependency relation $\sim$ provides an oblivious resampling oracle for $\overline{\mathcal{A}}$. If $\mathfrak{\Re}$ is regenerating on $\mathcal{A}$, then the resampling oracle on $\overline{\mathcal{A}}$ is also regenerating.

It would seem reasonable that if $\mathcal{A}$ is commutative, then $\overline{\mathcal{A}}$ would be as well. Unfortunately, we do not know how to show this for the commutativity definition of [23]. For our commutativity definition, this is easy to show; in addition, $\overline{\mathcal{A}}$ will inherit a number of other nice properties. This is a good illustration of how the new definition of commutativity is easier to work with, beyond its advantage of greater generality.

- Proposition 36. Suppose that $\mathcal{A}$ is oblivious but not necessarily commutative. For a flaw $g=\langle C\rangle$, suppose that we have fixed an enumeration $C=\left\{f_{1}, \ldots, f_{t}\right\}$ to define $\mathfrak{R}_{g}$. Then $A_{g} \propto A_{f_{1}} \ldots A_{f_{t}}$.

Proof. By definition of $\mathfrak{R}_{g}$, we have $A_{g}\left[\sigma, \sigma^{\prime}\right]=\operatorname{Pr}\left(r_{t} \ldots r_{1} \sigma=\sigma^{\prime}\right)$, where each $r_{i}$ is drawn independently from $R_{f_{i} ; f_{i+1}, \ldots, f_{t}}$. Let us define $R_{i}^{\prime}=R_{f_{i} ; f_{i+1}, \ldots, f_{t}}$ and $\sigma_{i}=r_{i} \ldots r_{1} \sigma$ for $i=0, \ldots, t$ (where $\sigma_{0}=\sigma$ ). By enumerating over possible values for $\sigma_{1}, \ldots, \sigma_{t}$, we get $A_{g}\left[\sigma, \sigma^{\prime}\right]=\sum \underset{\substack{\sigma_{1}, \ldots, \sigma_{t} \\ \sigma_{t}=\sigma^{\prime}}}{ } \prod_{i=1}^{t} \operatorname{Pr}_{r_{i} \sim R_{i}^{\prime}}\left(r_{i} \sigma_{i-1}=\sigma_{i}\right)$.

Note that if $\sigma_{i} \notin f_{j}$ for some $j>i$, then the term $\operatorname{Pr}_{r_{i} \sim R_{i}^{\prime}}\left(r_{i} \sigma_{i-1}=\sigma_{i}\right)$ must be zero, since $r_{i} \in R_{i}^{\prime} \subseteq R_{f_{i} ; f_{j}}$. So we may restrict the sum to terms with $\sigma_{i} \in f_{i+1} \cap \cdots \cap f_{t}$ for all $i=0, \ldots, t$. For each such term, we have $\operatorname{Pr}_{r_{i} \sim R_{i}^{\prime}}\left(r_{i} \sigma_{i-1}=\sigma_{i}\right)=\frac{\operatorname{Pr}_{r_{i} \sim R_{i}}\left(r_{i} \sigma_{i-1}=\sigma_{i}\right)}{\operatorname{Pr}_{r_{i} \sim R_{i}}\left(r_{i} \in R_{i}^{\prime}\right)}=$ $\frac{A_{f_{i}}\left[\sigma_{i-1}, \sigma\right]}{\operatorname{Pr}_{r_{i}} \sim R_{i}\left(r_{i} \in R_{i}^{\prime}\right)}$. So:

$$
\begin{aligned}
& A_{g}\left[\sigma, \sigma^{\prime}\right]=\sum_{\substack{\sigma_{1}, \ldots, \sigma_{t} \\
\sigma_{t}=\sigma^{\prime}}} \frac{A_{f_{1}}\left[\sigma_{0}, \sigma_{1}\right] \ldots, A_{f_{t}}\left[\sigma_{t-1}, \sigma_{t}\right]}{\prod_{i=1}^{t} \operatorname{Pr}_{r_{i} \sim R_{i}}\left(r_{i} \in R_{i}^{\prime}\right)}=\frac{\sum c_{1}, \ldots, \sigma_{t}}{\sigma_{t}=\sigma^{\prime}} A_{f_{1}}\left[\sigma_{0}, \sigma_{1}\right] \ldots, A_{f_{t}}\left[\sigma_{t-1}, \sigma_{t}\right] \\
& \prod_{i=1}^{t} \operatorname{Pr}_{r_{i} \sim R_{i}}\left(r_{i} \in R_{i}^{\prime}\right) \\
&=\left(c A_{f_{1}} \ldots A_{f_{t}}\right)\left[\sigma, \sigma^{\prime}\right]
\end{aligned}
$$

- Proposition 37. If $\mathcal{A}$ is commutative, then the transition matrix $A_{g}$ for a flaw $g=\langle C\rangle$ does not depend on the chosen enumeration $C=\left\{f_{1}, \ldots, f_{t}\right\}$.

Proof. By Proposition 36, we have $A_{g}=c A_{g}^{\prime}$ for $A_{g}^{\prime}=A_{f_{1}} \ldots A_{f_{t}}$. Since the matrices $A_{f_{i}}$ all commute, $A_{g}^{\prime}$ does not depend on the enumeration of $C$. Furthermore, the constant $c$ can be determined from $A_{g}^{\prime}$ by choosing an arbitrary state $\sigma \in g$ and setting $c=\frac{1}{\sum_{\sigma^{\prime}} A_{g}^{\prime}\left[\sigma, \sigma^{\prime}\right]}$.

- Theorem 38. If the resampling oracle is commutative on $\mathcal{A}$, then it is also commutative on $\overline{\mathcal{A}}$.

Proof. Let $g=\langle C\rangle$ and $g^{\prime}=\left\langle C^{\prime}\right\rangle$ for stable sets $C, C^{\prime}$ such that $g \nsim g^{\prime}$. So $f \nsim f^{\prime}$ for all $f \in C$ and $f^{\prime} \in C^{\prime}$. By Proposition 36 we have

$$
A_{g} A_{g^{\prime}}=c_{g} c_{g^{\prime}}\left(\prod_{f \in C} A_{f} \prod_{f^{\prime} \in C^{\prime}} A_{f^{\prime}}\right), \quad \quad A_{g^{\prime}} A_{g}=c_{g^{\prime}} c_{g}\left(\prod_{f^{\prime} \in C^{\prime}} A_{f^{\prime}} \prod_{f \in C} A_{f}\right)
$$

for scalar constants $c_{g}, c_{g^{\prime}}$. All these matrices $A_{f}, A_{f^{\prime}}$ commute, so both quantities are equal.

Another useful property for such resampling oracles is idempotence. We say that $\mathcal{A}$ is idempotent if $A_{f}^{2} \propto A_{f}$ for all $f \in \mathcal{A}$. Most of the known commutative resampling oracles, resampling oracles have this property, including the variable LLLL and the permutation LLL.

- Proposition 39. If the resampling oracle is commutative and idempotent on $\mathcal{A}$, then it is also idempotent on $\overline{\mathcal{A}}$. Furthermore, for any stable set $I=\left\{\left\langle C_{1}\right\rangle, \ldots,\left\langle C_{k}\right\rangle\right\}$ of $\overline{\mathcal{A}}$ and stable set $J=C_{1} \cup \cdots \cup C_{k}$ of $\mathcal{A}$, there holds $A_{I} \propto A_{J}$.

Proof. First, let $f=\langle C\rangle$ for stable set $C=\left\{g_{1}, \ldots, g_{k}\right\}$. Proposition 36 gives $A_{f}^{2} \propto$ $\left(A_{g_{1}} \cdots A_{g_{k}}\right)^{2}$. Since the matrices $A_{g_{i}}$ commute with each other, this gives $A_{f}^{2} \propto A_{g_{1}}^{2} \cdots A_{g_{k}}^{2}$. Since $\mathcal{A}$ is idempotent, this is proportional to $A_{g_{1}} \cdots A_{g_{k}}$, which again by Proposition 36 is proportional to $A_{f}$.

For the second result, Proposition 36 gives $A_{I} \propto \prod_{i=1}^{k} \prod_{g \in C_{i}} A_{g}=\prod_{g \in J} A_{g}^{n_{g}}$ where $n_{g} \geq 1$ is the number of copies of $g$ appearing in $C_{1}, \ldots, C_{k}$. Since $\mathcal{A}$ is idempotent, each term $A_{g}^{n_{g}}$ is proportional to $A_{g}$. Hence we have $A_{I} \propto \prod_{g \in J} A_{g}=A_{J}$.

## References

1 Dimitris Achlioptas and Fotis Iliopoulos. Random walks that find perfect objects and the Lovász local lemma. Journal of the ACM, 63(3):Article \#22, 2016. doi:10.1145/2818352.

2 Dimitris Achlioptas, Fotis Iliopoulos, and Vladimir Kolmogorov. A local lemma for focused stochastic algorithms. SIAM Journal on Computing, 48(5):1583-1602, 2019. doi:10.1137/ 16M109332X.
3 Dimitris Achlioptas, Fotis Iliopoulos, and Alistair Sinclair. Beyond the Lovász local lemma: Point to set correlations and their algorithmic applications. In Proc. 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 725-744, 2019.
4 Noga Alon, Joel Spencer, and Prasad Tetali. Covering with latin transversals. Discrete applied mathematics, 57(1):1-10, 1995.
5 Rodrigo Bissacot, Roberto Fernández, Aldo Procacci, and Benedetto Scoppola. An improvement of the Lovász local lemma via cluster expansion. Combinatorics, Probability \& Computing, 20(5):709-719, 2011. doi:10.1017/S0963548311000253.
6 Karthekeyan Chandrasekaran, Navin Goyal, and Bernhard Haeupler. Deterministic algorithms for the Lovász local lemma. SIAM Journal on Computing, 42(6):2132-2155, 2013. doi: 10.1137/100799642.

7 Antares Chen, David G. Harris, and Aravind Srinivasan. Partial resampling to approximate covering integer programs. Random Structures \& Algorithms, pages 69-93, 2021.
8 Kai-Min Chung, Seth Pettie, and Hsin-Hao Su. Distributed algorithms for the Lovász local lemma and graph coloring. Distributed Computing, 30(4):261-280, 2017.
9 Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, pages 609-627. Colloq. Math. Soc. János Bolyai, Vol. 10. 1975.
10 Paul Erdös and Joel Spencer. Lopsided Lovász local lemma and latin transversals. Discrete Applied Mathematics, 30(2-3):151-154, 1991. doi:10.1016/0166-218X (91) 90040-4.
11 Bernhard Haeupler and David G. Harris. Parallel algorithms and concentration bounds for the Lovász local lemma via witness DAGs. ACM Transactions on Algorithms, 13(4):Article \#25, 2017.
12 Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the Lovász local lemma. Journal of the ACM, 58(6):Article \#28, 2011. doi:10.1145/2049697. 2049702.

13 David G. Harris. Lopsidependency in the Moser-Tardos framework: Beyond the lopsided Lovász local lemma. ACM Transactions on Algorithms, 13(1):Article \#17, 2016. doi: 10.1145/3015762.

14 David G. Harris. New bounds for the Moser-Tardos distribution. Random Structures $\&$ Algorithms, 57(1):97-131, 2020.
15 David G. Harris. Oblivious resampling oracles and parallel algorithms for the Lopsided Lovász Local Lemma. ACM Transactions on Algorithms, 17(1):Article \#1, 2021.
16 David G. Harris and Aravind Srinivasan. Algorithmic and enumerative aspects of the MoserTardos distribution. ACM Transactions on Algorithms, 13(3):Article \#33, 2017.
17 David G. Harris and Aravind Srinivasan. A constructive Lovász Local Lemma for permutations. Theory of Computing, 13(1):Article \#17, 2017.
18 David G. Harris and Aravind Srinivasan. The Moser-Tardos framework with partial resampling. Journal of the ACM, 66(5):Article \#36, 2019.
19 Nicholas J. A. Harvey and Jan Vondrák. An algorithmic proof of the Lovász local lemma via resampling oracles. SIAM Journal on Computing, 49(2):394-428, 2020.
20 Fotis Iliopoulos. Commutative algorithms approximate the LLL-distribution. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 2018.
21 Fotis Iliopoulos and Alistair Sinclair. Efficiently list-edge coloring multigraphs asymptotically optimally. In Proc. 14 th annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 2319-2336, 2020.
22 Kashyap Kolipaka and Mario Szegedy. Moser and Tardos meet Lovász. In Proc. 43 rd annual ACM Symposium on Theory of Computing (STOC), pages 235-244, 2011.

23 Vladimir Kolmogorov. Commutativity in the algorithmic Lovász local lemma. SIAM Journal on Computing, 47(6):2029-2056, 2018.
24 László Lovász. Submodular functions and convexity. In Mathematical Programming: the State of the Art, pages 235-257. Springer, 1983.
25 Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. Journal of the ACM, 57(2):Article \#11, 2010. doi:10.1145/1667053.1667060.
26 Wesley Pegden. An extension of the Moser-Tardos algorithmic local lemma. SIAM Journal on Discrete Mathematics, 28(2):911-917, 2014. doi:10.1137/110828290.

## A Necessity of transition matrix commutativity for Lemma 5

Consider a set of events $\mathcal{B}^{*}$ with a dependency relation $\sim$. We say that $\mathcal{B}^{*}$ is complete if for each $\sigma \in \Omega$ there exists a flaw $h_{\sigma}=\{\sigma\} \in \mathcal{B}^{*}$, and with $h_{\sigma} \sim g$ for all $g \in \mathcal{B}^{*}$. Note that this definition is satisfied if $\mathcal{B}^{*}$ is generated by atomic events corresponding to permutations, perfect matchings of hypergraphs, or spanning trees.

We show now that if transition matrix commutativity fails in a complete set of events, even for a single pair of flaws, then some wdags may appear with probability arbitrarily higher than their weight. It can be checked that preconditions of the lemma (and thus its conclusion) apply to some existing resampling oracles, such as the oracle for spanning trees from [19] and the oracle for perfect matchings of complete $s$-uniform hypergraphs (for $s \geq 3$ ) from [15].

- Theorem 40. Suppose that $\mathcal{B}^{*}$ is complete, regenerating, and contains a pair $f, g \in \mathcal{B}^{*}$ with $f \nsim g$ and $A_{f} A_{g} \neq A_{g} A_{f}$. Then for any $C>0$ there exists a set of flaws $\mathcal{B} \subseteq \mathcal{B}^{*}$ with $|\mathcal{B}|=3$, wdag $H$ with a single sink and a flaw resampling strategy $S$ such that the probability that $H$ appears in the execution of the algorithm is at least $C \cdot w(H)=C \cdot \prod_{v \in H} \mu(L(v))$.

Proof. Consider states $\sigma, \tau$ with $A_{f} A_{g}[\sigma, \tau] \neq A_{g} A_{f}[\sigma, \tau]$. Denote $x=A_{f} A_{g} e_{\tau}$ and $y=$ $A_{g} A_{f} e_{\tau}$, and assume w.l.o.g. that $x[\sigma]<y[\sigma]$. Note that $\mu^{\top} A_{f} A_{g}=\mu^{\top} A_{g} A_{f}=\gamma_{f} \gamma_{g} \cdot \mu^{\top}$ since the oracles are regenerating, and therefore $\mu^{\top} x=\mu^{\top} y=\gamma_{f} \gamma_{g} \cdot \mu[\tau]=\gamma_{f} \gamma_{g} \gamma_{h}$.

Consider the following strategy $S$ given a current state $\sigma_{1}$ : (i) if $\sigma_{1} \neq \sigma$ then prioritize flaws $f, g, h$ at steps $1,2,3$ respectively; (ii) if $\sigma_{1}=\sigma$ then prioritize flaws $g, f, h$ at steps $1,2,3$ respectively. We say that the run succeeds if the sequence of addressed flaws is $(f, g, h)$ in the first case and ( $g, f, h$ ) in the second case. Clearly, the probability of success equals $e_{\sigma_{1}}^{\top} A_{f} A_{g} e_{\tau}=e_{\sigma_{1}}^{\top} x$ in the first case and $e_{\sigma_{1}}^{\top} A_{g} A_{f} e_{\tau}=e_{\sigma_{1}}^{\top} y$ in the second case. If $\sigma_{1}$ is distributed according to $\mu$ then the probability of success is

$$
\begin{aligned}
p & =\mu[\sigma] \cdot e_{\sigma}^{\top} y+\sum_{\sigma_{1} \in \Omega-\{\sigma\}} \mu\left[\sigma_{1}\right] \cdot e_{\sigma_{1}}^{\top} x=\mu[\sigma] \cdot\left(e_{\sigma}^{\top} y-e_{\sigma}^{\top} x\right)+\sum_{\sigma_{1} \in \Omega} \mu\left[\sigma_{1}\right] \cdot e_{\sigma_{1}}^{\top} x \\
& =\mu^{\top} x+\mu[\sigma] \cdot(y[\sigma]-x[\sigma])>\gamma_{f} \gamma_{g} \gamma_{h}
\end{aligned}
$$

Furthermore, if the run succeeds then the last state is distributed according to $\mu$ (since step 3 resamples $h$ at state $\tau$, and the oracles are regenerating).

Now consider the trajectory which repeats the sequence $f, g, h$ for $n$ times, and the corresponding wdag $H=G_{3 n}^{T}$ which has a single sink node labeled $h$. Let $S^{n}$ be the strategy $S$ repeated cyclically. From the previous paragraph, the probability that the run starting with some distribution $\mu$ produces $H$ is given by $c_{\mu} \cdot p^{n-1}$, where $c_{\mu}$ depends only on the initial distribution. Note that $w(H)=\left(\gamma_{f} \gamma_{g} \gamma_{h}\right)^{n}$. Choosing $n$ sufficiently large now gives the claim.

## B Proof of Theorem 28

First, there is a probability of $\mu(E)$ that $E$ is true at time 0 . To bound the probability the $E$ becomes true later, let us say that a pair $(f, t)$ is good if (i) $E$ is false at time $t$; (ii) $f$ is reampled at time $t$; and (iii) either this is the first resampling of $f$, or if the most recent resampling of $f$ had occured at time $t^{\prime}<t$, then $f$ had been false at some intermediate time between $t^{\prime}+1$ and $t$. If $(f, t)$ is good, we say it is finalized at time $s \geq t$ if (i) $f$ is resampled at time $s$; (ii) $E$ has been false at all times prior to $s$; and (iii) $f$ is true at times $t, \ldots, s$.

We claim that if $E$ becomes true, then there is some pair $(f, t)$ which is good and is finalized at some time $s \geq t$. For, suppose that $E$ first becomes true at time $s \geq 1$ and that $f$ is resampled at time $s$. Going backward in time, look for the earliest time $t$ such that the same flaw $f$ is resampled at time $t$ and that $f$ remained true between $t$ and $s$ (possibly $t=s)$. Then $(f, t)$ is good and is finalized at time $s$.

Now let us fix good pair $(f, t)$. For each $s \geq t$, let $\mathcal{E}_{s}$ be the event that $(f, t)$ is finalized by some $s^{\prime} \geq s$. We claim that $\operatorname{Pr}\left(\mathcal{E}_{s}\right) \leq \kappa(f, E)$; furthermore, this probability bound holds conditional on the full state of the system at times up to $s$.

By a limiting argument, it suffices to show this bound if we restrict to $s \leq s_{\max }$ for arbitrary integer $s_{\max }$. For fixed $s_{\max }$, we show it by induction backward on $s$. The claim follows immediately from induction if $f$ is not being resampled at time $s$ or if $s=s_{\max }$. If $E$ is true at time $s$, then $\mathcal{E}_{s}$ is impossible (since $(f, t)$ would have needed to be finalized at some earlier time $s^{\prime}<s$ ). Likewise, if there was an intervening time between $t$ and $s$ where $f$ was false, then $\mathcal{E}_{s}$ is impossible.

So, suppose we resample $f$ at time $s$ while $E$ is false and $f$ has remained true for times $t, \ldots, s$. Let $\tau \in f \cap \bar{E}$ be the state at time $s$. There are three things that can happen when resampling $f$ :

- $E$ becomes true. This has probability $e_{\tau}^{\top} A_{f} e_{E}$. In this case, event $\mathcal{E}_{s}$ may have occurred. - $E$ remains false, and $f$ becomes false. This has probability $e_{\tau}^{\top} A_{f} e_{\bar{E} \cap \bar{f}}$. In this case, event $\mathcal{E}_{s}$ is impossible.
- $E$ becomes false, and $f$ remains true. This has probability $e_{\tau}^{\top} A_{f} e_{\bar{E} \cap f}$. In this case, in order to $\mathcal{E}_{s}$ to occur, it must be that $\mathcal{E}_{s+1}$ holds after resampling $f$. By induction hypothesis, this has probability at most $\kappa(f, E)$.

Overall, we have $\operatorname{Pr}\left(\mathcal{E}_{s}\right) \leq e_{\tau}^{\top} A_{f} e_{E} \cdot 1+e_{\tau}^{\top} A_{f} e_{\bar{E} \cap f} \cdot \kappa(f, E)$. This is at most $\kappa(f, E) e_{\tau} A_{f} e_{E \cup \bar{f}}+e_{\tau}^{\top} A_{f} e_{\bar{E} \cap f} \cdot \kappa(f, E) \leq \kappa(f, E)$. This concludes the induction.

Now consider any good pair $(f, t)$. Because of the claim, we know that that the probability that $(f, t)$ is finalized (by any time $s \geq t$ ) is at most $\kappa(f, E)$, conditional on all other state at time $t$. Since this is a necessary condition for $E$ to become true, the overall probability that $E$ becomes true is at most $\mathbf{E}\left[L_{f}\right] \cdot \kappa(f, E)$, where $L$ is the number of good pairs $(f, t)$. Note that, between any good pairs $(f, t)$ and $\left(f, t^{\prime}\right)$ for $t^{\prime}>t$, the event $F \cap f$ is false at least once, where $F$ is an arbitrary event with $F \supseteq \bar{E}$. However, at times $t$ and $t^{\prime}$, the event $F \cap f$ is true. Thus, $F \cap f$ is caused to become true at least $L_{f}$ times, and so by Proposition 22, we have $\mathbf{E}\left[L_{f}\right] \leq N_{\mathcal{G}}(F \cap f)$.

## C Proof of Theorem 30 and Theorem 31

We begin by considering the setting where $\Omega$ is the uniform distribution on the permutations $\pi$ on $[n]$. The set $\mathcal{A}$ is defined as follows: for each pair $(x, y) \in[n] \times[n]$, there is atom $\pi x=y$, which we denote by $[x, y]$. The resampling oracle here, for such an event, is to update the state $\pi \leftarrow(y z) \pi$, where $z$ is uniformly drawn from $[n]$. (Here and throughout the section, $(y z)$ denotes the permutation which swaps $y$ and $z$.) We have $[x, y] \sim\left[x^{\prime}, y^{\prime}\right]$ if exactly one
of the following holds: (i) $x=x^{\prime}$ or (ii) $y=y^{\prime}$. Equivalently, this holds iff $[x, y] \cap\left[x^{\prime}, y^{\prime}\right]=\emptyset$. See [15] for further details, including a proof that the resampling oracle is commutative and oblivious.

We begin with a basic observation on how atoms of $\mathcal{A}$ interact with events in $\overline{\mathcal{A}}$.

- Proposition 41. Let $f=[x, y]$ be an atom and let $E=\langle C\rangle$ where $C$ is a stable set of $\mathcal{A}$. Then $A_{f} e_{E} \propto e_{E^{\prime}}$, where $E^{\prime}=\left\langle C^{\prime}\right\rangle$ and stable set $C^{\prime}$ is obtained from $C$ as follows:
- If $C$ contains exactly two atoms $f_{1}, f_{2}$ which are neighbors of $f$, i.e. $f_{1}=\left[x, y_{1}\right]$ and $f_{2}=\left[x_{2}, y\right]$, then $C^{\prime}=C-\left\{f_{1}, f_{2}\right\} \cup\left\{[x, y],\left[x_{2}, y_{1}\right]\right\}$.
- If $C$ contains exactly one atom $f_{1}$ which is a neighbor of $f$, then $C^{\prime}=C-f_{1} \cup\{f\}$.
- Otherwise, if $C$ contains no neighbors of $f$, then $C^{\prime}=C \cup\{f\}$.

Proof. Consider state $\pi$, and suppose we resample $f$ to obtain $\pi^{\prime}=(y z) \pi$. If $\pi \notin f$, then $e_{\pi}^{\top} A_{f} e_{E}=0=e_{\pi}^{\top} e_{E^{\prime}}$. Similarly, for each $f^{\prime} \in C$ which is not a neighbor of $f$, we must have $\pi \in f^{\prime}$ as otherwise $e_{\pi}^{\top} A_{f} e_{E}=0=e_{\pi}^{\top} e_{E^{\prime}}$. In these cases, we also automatically have $\pi^{\prime} \in f^{\prime}$ for all such $f^{\prime}$. Thus, we suppose that $\pi \in f$ and also $\pi \in f^{\prime}$ for all $f^{\prime} \in C-\Gamma(f)$. We consider the following cases in turn:

- If $C$ contains two atoms $f_{1}, f_{2}$, then we claim that $\pi^{\prime} \in E$ precisely when $z=y_{1}$ and $\pi x_{2}=y$. For, in order to have $\pi^{\prime} \in f_{1}$, we must have $\pi^{\prime} x=y_{1}$. Since $\pi x=y$, this implies that $(y z) y=y_{1}$, i.e. $z=y_{1}$. Thus, $\pi^{\prime}=\binom{y}{y_{1}} \pi$. To satisfy $f_{2}$, we must have $y=\pi^{\prime} x_{2}=\left(\begin{array}{ll}y & y_{1}\end{array}\right) \pi x_{2}$, i.e. $\pi x_{2}=y$. In this case, we see that $e_{\pi}^{\top} A_{f} e_{E}=1 / n$ for all $\pi$ and also $e_{\pi}^{\top} e_{E^{\prime}}=1$.
- Suppose that $C$ contains a neighbor $f_{1}=\left[x, y_{1}\right]$. In this case, we have $\pi^{\prime} \in f_{1}$ precisely if $y_{1}=z$. Similarly, suppose that $C$ contains a neighbor $f_{2}=\left[x_{2}, y\right]$. In this case, we have $\pi^{\prime} \in f_{2}$ precisely if $z=\pi x_{2}$. Thus, we have $e_{\pi}^{\top} A_{f} e_{E}=1 / n$ for all such $\pi$ and also $e_{\pi}^{\top} e_{E^{\prime}}=1$.
- If $C$ has no neighbors of $f$, then $\pi^{\prime}$ is in $E$ iff $z \notin\left\{y_{1}, \ldots, y_{k}\right\}$ where $C=$ $\left\{\left[x_{1}, y_{1}\right], \ldots,\left[x_{k}, y_{k}\right]\right\}$. Thus $e_{\pi}^{\top} A_{f} e_{E}=\frac{n-k}{n}$ and $e_{\pi}^{\top} e_{E^{\prime}}=1$

To understand more complex, multi-atom interactions, let us fix event $E=\langle C\rangle$ for a stable set $C$. For a stable set $I \subseteq \mathcal{A}$, we can form an associated bipartite graph $G_{I}$, as follows: the left vertices correspond to $C$ (we call these $C$-nodes), and the right vertices correspond to $I$ (we call these $I$-nodes). It has an edge between $f$ and $f^{\prime}$ iff $f \sim f^{\prime}$. Observe that since $C$ and $I$ are stable, the graph $G_{I}$ has degree at most two - each node $[x, y]$ can have one neighbor of the form $\left[x^{\prime}, y\right]$ and another neighbor of the form $\left[x, y^{\prime}\right]$. So, $G_{I}$ decomposes into paths and cycles.

We define $\tau(I)$ to be the size of a maximum matching in $G_{I}$. We also define the active conditions for $I$, denoted $\operatorname{Active}(I) \subseteq \mathcal{A}$, as follows. First, for each $f \in I$, we also place $f$ into $\operatorname{Active}(I)$. Second, consider some maximal path of $G_{I}$ starting and ending at $C$-nodes (which we call a $C$-path). The path can be written (in one of its two orientations) as

$$
\left[x_{1}, y_{1}\right],\left[x_{1}, y_{2}\right],\left[x_{2}, y_{2}\right], \ldots,\left[x_{k}, y_{k-1}\right],\left[x_{k}, y_{k}\right]
$$

In this case, we also put $\left[x_{k}, y_{1}\right]$ into $\operatorname{Active}(I)$. (It is possible that $k=1$, in which case $\left[x_{1}, y_{1}\right]$ is an isolated $C$-node.)

For brevity, we define $\alpha(I)$ to be the event $\langle\operatorname{Active}(I)\rangle$ in $\overline{\mathcal{A}}$. The active conditions determine the vector $A_{I} e_{E}$, and also have a number of nice combinatorial properties.

- Proposition 42. Let $I$ be a stable set of $\mathcal{A}$. Then the following properties hold:

1. $A_{I} e_{E} \propto e_{\alpha(I)}$.
2. $|\operatorname{Active}(I)|=|C|+|I|-\tau(I)$.
3. Any $f \in I$ with $\tau(I)=\tau(I-f)$ has Active $(I)=\operatorname{Active}(I-f) \cup\{f\}$.

## Proof.

1. We show this by induction on $I$. The base case $I=0$ is clear, since then $\operatorname{Active}(I)=C$. For the induction step, consider $f \in I$. We have $A_{I} e_{E}=A_{f} A_{I-f} e_{E}$; by induction hypothesis, this is proportional to $A_{f} e_{\langle U\rangle}$ where $U=$ Active $(I-f)$. By Proposition 41, this in turn is proportional to $e_{\left\langle U^{\prime}\right\rangle}$, where $U^{\prime}$ is formed according to the specific given rules. We thus need to show that $U^{\prime}=\operatorname{Active}(I)$. There are three cases.
If $U$ has no neighbors of $f$, then $U^{\prime}=U \cup\{f\}$. Furthermore, $G_{I}$ has no changes in its $C$-paths compared to $G_{I-f}$, so $\operatorname{Active}(I)=\operatorname{Active}(I-f) \cup\{f\}=U^{\prime}$.
Next suppose $U$ has one neighbor $f^{\prime}=\left[x^{\prime}, y^{\prime}\right]$ of $f$. Since $I$ is a stable set, it must have $f^{\prime} \notin I$, i.e. $G_{I-f}$ contains a $C$-path with endpoints $x^{\prime}, y^{\prime}$. This $C$-path now terminates in a degree-one $I$-node $[x, y]$ in $G_{I}$, and hence it is removed from $G_{I}$. So $\operatorname{Active}(I)=\operatorname{Active}(I-f)-f^{\prime} \cup\{f\}=U^{\prime}$.
Finally, If $U$ has two neighbors $f_{1}=\left[x, y_{1}\right], f_{2}=\left[x_{2}, y\right]$, then again since $I$ is a stable set these must correspond to $C$-paths in $G_{I-f}$. Thus, there are two $C$-paths with endpoints $x, y_{1}$ and $x_{2}, y$ respectively. Now in $G_{I}$, there is a new degree-two $I$-node $[x, y]$. This merges the two $C$-paths into a single new $C$-path with endpoints $x_{2}, y_{1}$. Thus again $\operatorname{Active}(I)=\operatorname{Active}(I-f)-\left\{f_{1}, f_{2}\right\} \cup\left\{f,\left[x_{2}, y_{1}\right]\right\}=U^{\prime}$.
2. Consider some connected component of $G_{I}$; it is a path or cycle with $i$ distinct $I$-nodes and $c$ distinct $C$-nodes, where $c \in\{i-1, i, i+1\}$. If $c=i-1$, then it has maximum matching size $t=i-1$ else it has maximum matching size $t=i$. If $c=i+1$, then it has one additional active condition corresponding to the $C$-path on its nodes, and thus has $a=i+1$ active conditions; else it has $a=i$ active conditions. In all cases, it can be checked that $a=c+i-t$. The claimed formula is obtained by summing over all components.
3. By part (2), we have $\tau(I)=\tau(I-f)$ precisely if $|\operatorname{Active}(I)|=|\operatorname{Active}(I-f)|+1$, i.e. $G_{I}$ has the same number of $C$-paths as $G_{I-f}$. We claim in this case that $G_{I}$ has the same $C$-paths as $G_{I-f}$ as well, which will show the claim.
For, if not, then $G_{I}$ would need to gain, and lose, some $C$-paths compared to $G_{I}$. The new $I$-node $[x, y]$ would need to participate in a new $C$-path. This can only occur if $G_{I-f}$ has two $C$-paths with endpoints $\left[x, y^{\prime}\right]$ and $\left[x^{\prime}, y\right]$. But in this case, these two existing $C$-paths get destroyed in $G_{I}$, and thus in fact $G_{I}$ has strictly fewer $C$-paths compared to $G_{I-f}$.

- Proposition 43. Let $I=\left\{f_{1}, \ldots, f_{k}\right\}$ be a stable set in $\overline{\mathcal{A}}$, where $f_{i}=\left\langle F_{i}\right\rangle$ for each $i$. Consider the stable sets $J^{\prime}=F_{1} \cup \cdots \cup F_{k-1}$ and $J=J^{\prime} \cup F_{k}$ of $\mathcal{A}$. If $\tau\left(J^{\prime}\right)=\tau(J)$, then $f_{k}$ is dominated by $I-f_{k}$ in $\overline{\mathcal{A}}$.

Proof. Let $I^{\prime}=\left\{f_{1}, \ldots, f_{k-1}\right\}$. t is easily seen that the permutation LLL setting is idemptotent. Thus, by Proposition 39, we have $A_{I^{\prime}} \propto A_{J^{\prime}}$. Combined with Proposition 42(1), this implies that there is some scalar value $p \geq 0$ such that $A_{I^{\prime}} e_{E}=p e_{\alpha\left(J^{\prime}\right)}$. We want to show that

$$
\begin{equation*}
e_{\pi}^{\top} A_{I} e_{E} \leq e_{\pi}^{\top} A_{I^{\prime}} e_{E} \tag{4}
\end{equation*}
$$

for any state $\pi$.
By Proposition 39, we have $A_{I} \propto A_{J}$. Thus, the LHS of Eq. (4) is zero if $\pi \notin \alpha(J)$, in which case the inequality clearly holds. So suppose that $\pi \in \alpha(J)$. By Proposition 42(3), we have $\operatorname{Active}\left(J^{\prime}\right) \subseteq \operatorname{Active}(J)$ since $\tau\left(J^{\prime}\right)=\tau(J)$. In this case, also $\pi \in \alpha\left(J^{\prime}\right)$ so the RHS of Eq. (4) is equal to $p$. The LHS can be factored as $e_{\pi}^{\top} A_{I} e_{E}=\sum_{\sigma} A_{f}[\pi, \sigma] \cdot e_{\sigma}^{\top} A_{I^{\prime}} e_{E}=$ $\sum_{\sigma \in \alpha\left(J^{\prime}\right)} A_{f}[\pi, \sigma] p$. Since matrix $A_{f}$ is substochastic, this is at most $p$. This establishes the desired inequality.

- Proposition 44. For any $I \in \mathfrak{O}(E)$, there is an injective function $\phi_{I}: I \rightarrow C$ with $g \sim \phi_{I}(g)$ for all $g \in I$.

Proof. By definition, $I$ can be ordered as $I=\left\{f_{1}, \ldots, f_{k}\right\}$, where $f_{i}=\left\langle F_{i}\right\rangle$ and such that each $f_{i}$ is not dominated by $\left\{f_{1}, \ldots, f_{i-1}\right\}$. Let us define $J_{i}=F_{1} \cup \cdots \cup F_{i}$ for each $i$. By Proposition 43, we must have $\tau\left(J_{i}\right)>\tau\left(J_{i-1}\right)$ for each $i$. Thus, for each $i$, there is some $g_{i} \in F_{i}-J_{i-1}$ and some $F_{i}^{\prime} \subseteq F_{i}-\left\{g_{i}\right\}$ with $\tau\left(J_{i-1} \cup F_{i}^{\prime} \cup\left\{g_{i}\right\}\right)>\tau\left(J_{i-1} \cup F_{i}^{\prime}\right)$. It is known (see, e.g. [24, Example 1.4]) that $\tau$ is a submodular set function. Hence, we have

$$
1=\tau\left(J_{i-1} \cup F_{i}^{\prime} \cup\left\{g_{i}\right\}\right)-\tau\left(J_{i-1} \cup F_{i}^{\prime}\right) \leq \tau\left(\left\{g_{1}, \ldots, g_{i-1}\right\} \cup\left\{g_{i}\right\}\right)-\tau\left(\left\{g_{1}, \ldots, g_{i-1}\right\}\right)
$$

since $\left\{g_{1}, \ldots, g_{i-1}\right\} \subseteq J_{i-1}$.
This implies that $\tau\left(\left\{g_{1}, \ldots, g_{k}\right\}\right)=k$ and $G_{\left\{g_{1}, \ldots, g_{k}\right\}}$ has a matching $M$ of size $k$. We define the function $\phi$ by setting $\phi\left(f_{i}\right)=c_{i}$ where $g_{i}$ is matched to $c_{i}$ in $M$.

We can now obtain Theorem 30.
Proof of Theorem 30. Clearly $\mu(E)=\frac{(n-k)!}{n!}$. To enumerate a set $I \in \mathfrak{O}(E)$, by Proposition 44, we choose, for each $g \in C$, either zero or one preimages $f=\phi_{I}^{-1}(g)$ in $I$. If we write $I_{g}$ for the set of preimages of $g$, then $\left|I_{g}\right| \leq 1$ for all $g$ and $I=\bigcup_{g \in C} I_{g}$. Overall, this shows that

$$
\sum_{I \in \mathfrak{O}(E)} \Psi(I) \leq \sum_{I_{g_{1}}, \ldots, I_{g_{k}}} \Psi\left(I_{g_{1}} \cup \cdots \cup I_{g_{k}}\right) \leq \sum_{I_{g_{1}}, \ldots, I_{g_{k}}} \Psi\left(I_{g_{1}}\right) \cdots \Psi\left(I_{g_{k}}\right)
$$

where the last inequality follows from log-subadditivity of $\Psi$. This can be written as $\prod_{i=1}^{k} \sum_{I_{g_{i}}} \Psi\left(I_{g_{i}}\right)$. The case of $I_{g_{i}}=\emptyset$ contributes 1 , and the case of $I_{g_{i}}=\{f\}$ contributes $\Psi(f)$.

We next consider the setting where $\Omega$ is the set of perfect matchings $M$ on the clique on vertex set $[n]$, where $n$ is an even integer. The set $\mathcal{A}$ is defined as follows: for each pair $(x, y) \in[n] \times[n]$ with $x \neq y$, there is an atomic event that $M \supseteq\{x, y\}$. We denote this event by $[x, y]$; note that $[x, y]=[y, x]$, which is different from the permutation setting. The resampling oracle, for such an event with $x<y$, is to update the state by drawing $z$ uniformly from $[n]-x$ and setting $M \leftarrow(y z) M$. Here, we are using the natural left-group action of permutations on matchings, i.e. $\sigma M=\left\{\left\{\sigma x^{\prime}, \sigma y^{\prime}\right\} \mid\left\{x^{\prime}, y^{\prime}\right\} \in M\right\}$.

We have $[x, y] \sim\left[x^{\prime}, y^{\prime}\right]$ if $\left|\{x, y\} \cap\left\{x^{\prime}, y^{\prime}\right\}\right|=1$. Equivalently, this holds iff $[x, y] \cap\left[x^{\prime}, y^{\prime}\right]=$ $\emptyset$. See [15] for further details, including a proof that this resampling oracle is commutative and oblivious.

As before, we begin with a basic observation on how atoms of $\mathcal{A}$ interact with events in $\overline{\mathcal{A}}$.

- Proposition 45. Let $f=[x, y]$ be an atom and let $E=\langle C\rangle$ where $C$ is a stable set of $\mathcal{A}$. Then $A_{f} e_{E} \propto e_{E^{\prime}}$, where $E^{\prime}=\left\langle C^{\prime}\right\rangle$ and stable set $C^{\prime}$ is obtained from $C$ as follows:
- If $C$ contains exactly two atoms $f_{1}, f_{2}$ which are neighbors of $f$, i.e. $f_{1}=\left[x, y_{1}\right]$ and $f_{2}=\left[x_{2}, y\right]$, then $C^{\prime}=C-\left\{f_{1}, f_{2}\right\} \cup\left\{[x, y],\left[x_{2}, y_{1}\right]\right\}$.
- If $C$ contains exactly one atom $f_{1}$ which is a neighbor of $f$, then $C^{\prime}=C-f_{1} \cup\{f\}$.
- Otherwise, if $C$ contains no neighbors of $f$, then $C^{\prime}=C \cup\{f\}$.

Proof. Consider state $M \in f$, and suppose we resample $f$ to $M^{\prime}=(y z) M$ where $z$ is drawn from $[n]-x$. For each $f^{\prime} \in C$ which is not a neighbor of $f$, we must have $M \in f^{\prime}$ as otherwise $e_{M}^{\top} A_{f} e_{E}=0=e_{M}^{\top} e_{E^{\prime}}$. In these cases, we also automatically have $M^{\prime} \in f^{\prime}$ for all such $f^{\prime}$. Thus, we suppose that $M \in f$ and also $M \in f^{\prime}$ for all $f^{\prime} \in C-\Gamma(f)$. We consider the following cases:

- Suppose that $C$ contains two atoms $f_{1}, f_{2}$. Then we claim that $M^{\prime} \in E$ precisely when $z=y_{1}$ and $\left\{x_{2}, y\right\} \in M$. First, we have $\left\{x, y_{1}\right\} \in M^{\prime}$ with $M^{\prime}=(y z) M$ iff $z=y_{1}$. Thus, $M^{\prime}=\left(y y_{1}\right) M$. To satisfy $f_{2}$, we must have $\left\{x_{2}, y_{1}\right\} \in M$. In this case, $e_{M}^{\top} A_{f} e_{E}=1 /(n-1)$ and also $e_{M}^{\top} e_{E^{\prime}}=1$.
- Suppose that $C$ contains a neighbor $f_{1}=\left[x, y_{1}\right]$. In this case, we have $M^{\prime} \in f_{1}$ precisely if $y_{1}=z$. Thus, we have $e_{M}^{\top} A_{f} e_{E}=1 /(n-1)$ and also $e_{M}^{\top} e_{E^{\prime}}=1$.
- Suppose $C$ has no neighbors of $f$. Let $C=\left\{\left[x_{1}, x_{2}\right], \ldots,\left[x_{2 k-1}, x_{2 k}\right]\right\}$; we have $M \in E^{\prime}$ precisely if $z \notin\left\{x_{1}, \ldots, x_{2 k}\right\}$. Since $z \neq x$, we thus have $e_{M}^{\top} A_{f} e_{E}=\frac{n-2 k+1}{n-1}$. We also have $e_{M}^{\top} e_{E^{\prime}}=1$.

To understand multi-atom interactions, let us fix event $E=\langle C\rangle$ for a stable set $C$. For a stable set $I \subseteq \mathcal{A}$, we can form an associated bipartite graph $G_{I}$, whose left vertices correspond to $C$ and whose right vertices correspond to $I$. It has an edge between $f$ and $f^{\prime}$ iff $f \sim f^{\prime}$. Observe that since $C$ and $I$ are stable, the graph $G_{I}$ has degree at most two each node $[x, y]$ can have one neighbor of the form $\left[x^{\prime}, y\right]$ and another neighbor of the form $\left[x, y^{\prime}\right]$. So, $G_{I}$ decomposes into paths and cycles.

We define $\tau(I)$ to be the size of a maximum matching in $G_{I}$. We also define Active $(I) \subseteq$ $\mathcal{A}$ as follows. First, for each $f \in I$, we also place $f$ into Active $(I)$. Second, consider some maximal path of $G_{I}$ starting and ending at $C$-nodes; the path can be written as $\left[x_{1}, x_{2}\right],\left[x_{2}, x_{3}\right], \ldots,\left[x_{k-1}, x_{k}\right]$ for even $k$. In this case, we also put $\left[x_{1}, x_{k}\right]$ into $\operatorname{Active}(I)$.

- Proposition 46. Let I be a stable set of $\mathcal{A}$. Then the following properties hold:

1. $A_{I} e_{E} \propto e_{\alpha(I)}$ where $\alpha(I)=\langle\operatorname{Active}(I)\rangle$.
2. $|\operatorname{Active}(I)|=|C|+|I|-\tau(I)$.
3. Any $f \in I$ with $\tau(I)=\tau(I-f)$ has $\operatorname{Active}(I)=\operatorname{Active}(I-f) \cup\{f\}$.

We omit the the proof of Proposition 44, as well as the remainder of the proof of Theorem 31, as they are precisely analogous to the proof of Theorem 30.

