Tractable minor-free generalization of planar zero-field Ising models

Valerii Likhosherstov\textsuperscript{1,}\textsuperscript{*}, Yury Maximov\textsuperscript{2} and Michael Chertkov\textsuperscript{3}

\textsuperscript{1} Department of Engineering, University of Cambridge, Cambridge, United Kingdom
\textsuperscript{2} Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, United States of America
\textsuperscript{3} Graduate Program in Applied Mathematics, University of Arizona, Tucson, AZ, United States of America

E-mail: vl304@cam.ac.uk

Received 25 October 2020
Accepted for publication 17 November 2020
Published 21 December 2020

Abstract. We present a new family of zero-field Ising models over $N$ binary variables/spins obtained by consecutive ‘gluing’ of planar and $O(1)$-sized components and subsets of at most three vertices into a tree. The polynomial-time algorithm of the dynamic programming type for solving exact inference (computing partition function) and exact sampling (generating i.i.d. samples) consists of sequential application of an efficient (for planar) or brute-force (for $O(1)$-sized) inference and sampling to the components as a black box. To illustrate the utility of the new family of tractable graphical models, we first build a polynomial algorithm for inference and sampling of zero-field Ising models over $K_{33}$-minor-free topologies and over $K_5$-minor-free topologies—both of which are extensions of the planar zero-field Ising models—which are neither genus- nor treewidth-bounded. Second, we empirically demonstrate an improvement in the approximation quality of the NP-hard problem of inference over the square-grid Ising model in a node-dependent nonzero ‘magnetic’ field.

Keywords: analysis of algorithms, inference of graphical models, statistical inference

\textsuperscript{*}Author to whom any correspondence should be addressed.
Contents

1. Introduction 3

2. Planar topology 6
   2.1. Expanded dual graph ................................................. 6
   2.2. Perfect matching (PM) model ....................................... 6

3. \(c\)-nice decomposition of the topology 8
   3.1. Decomposition tree and the key result (of the manuscript) .......... 8
   3.2. Inference and sampling conditioned on one, two, or three vertices/spins 10
   3.3. Inference algorithm .................................................. 10
   3.4. Sampling algorithm ................................................... 12

4. Minor-free extension of planar zero-field Ising models 13
   4.1. Zero-field Ising models over \(K_{33}\)-free graphs ...................... 14
   4.2. \(K_{33}\)-free zero-field Ising models: implementation and tests .......... 14
   4.3. Zero-field Ising models over \(K_{5}\)-free graphs ..................... 15

5. Approximate inference of square-grid Ising model 16

6. Conclusion 18

Appendix A. Lemma proofs 18
   A.1. Lemma 1 .................................................................. 18
   A.2. Lemma 2 .................................................................. 18
   A.3. Lemma 6 .................................................................. 19

Appendix B. Theorem 3 proof 21
   B.1. Counting PMs of planar \(\hat{G}\) in \(O\tilde{N}^{2}\) time ...................... 21
      B.1.1. Pfaffian orientation ................................................. 21
      B.1.2. Computing \(\text{det}\kappa\) .............................................. 21
      B.1.3. Nested dissection .................................................. 22
   B.2. Sampling PMs of planar \(\hat{G}\) in \(O\tilde{N}^{2}\) time (Wilson’s algorithm) .......... 23
      B.2.1. Structure of the algorithm ......................................... 23
      B.2.2. Drawing perfect matchings ....................................... 23
      B.2.3. Step 1: computing lower-right submatrix of \(\hat{K}^{-1}\) ............. 24
      B.2.4. Step 2: sampling edges in the separator ......................... 25
      B.3. Step 3: recursion .................................................... 26

Appendix C. Theorem 7 proof 26

Appendix D. Proof for theorem 10 28

Appendix E. Random graph generation 29

https://doi.org/10.1088/1742-5468/abcaf1
1. Introduction

Let $G = (V(G), E(G))$ be an undirected graph with a set of vertices $V(G)$ and a set of normal edges $E(G) \subseteq \binom{V(G)}{2}$ (no loops or multiple edges). We discuss Ising models that associate the following probability to each random $N \triangleq |V(G)|$-dimensional binary variable/spin configuration $X \in \{\pm 1\}^N$:

$$P(X) \triangleq \frac{W(X)}{Z},$$

where

$$W(X) \triangleq \exp \left( \sum_{v \in V(G)} \mu_v x_v + \sum_{e=\{v,w\} \in E(G)} J_{exvw} x_v x_w \right) \quad \text{and} \quad Z \triangleq \sum_{X \in \{\pm 1\}^N} W(X).$$

Here, $\mu = (\mu_v, v \in V(G))$ is a vector of (magnetic) fields, $J = (J_e, e \in E(G))$ is a vector of the (pairwise) spin interactions, and the normalization constant $Z$, which is defined as a sum over $2^N$ spin configurations, is referred to as the partition function. Given the model specification $I = \langle G, \mu, J \rangle$, we address the tasks of finding the exact value of $Z$ (inference) and drawing exact samples with the probability (1).

**Related work.** It has been known since the seminal contributions of Fisher (1966) and Kasteleyn (1963) that computation of the partition function in the zero-field ($\mu = 0$) Ising model over a planar graph and sampling from the respective probability distribution are both tractable; that is, these are tasks of polynomial complexity in $N$. As shown by Barahona (1982), even when $G$ is planar or when $\mu = 0$ (zero field), the positive results are hard to generalize—both addition of the nonzero (magnetic) field and the extension beyond planar graphs make the computation of the partition function NP-hard. These results are also consistent with the statement from Jerrum and Sinclair (1993) that computation of the partition function of the zero-field Ising model is a #P-complete problem, even in the ferromagnetic case in which all components of $J$ are positive. Therefore, describing $\langle G, \mu, J \rangle$ families for which computations of the partition function and sampling are tractable remains an open question.

The simplest tractable (i.e. inference and sampling are polynomial in $N$) example is one in which $G$ is a tree, and the corresponding inference algorithm, known as dynamic programming and/or belief propagation, has a long history in physics (Bethe 1935, Peierls 1936), optimal control (Bellman 1952), information theory (Gallager 1963), and artificial intelligence (Pearl 1982). The extension to the case in which $G$ is a tree of $(t + 1)$-sized
cliques ‘glued’ together, or more formally when $G$ is of a \textit{treewidth} $t$, is known as the \textit{junction tree algorithm} (Verner Jensen \textit{et al} 1990), which has complexity of counting and sampling that grow exponentially with $t$.

Another insight originates from the foundational statistical physics literature of the last century related to the zero-field version of (1), i.e. when $\mu = 0$, over planar $G$. Onsager (1944) found a closed-form solution of (1) in the case of a homogeneous Ising model over an infinite two-dimensional square grid. Kac and Ward (1952) reduced the inference of (1) over a finite square lattice to compute a determinant. Kasteleyn (1963) generalized this result to an arbitrary (finite) planar graph. Kasteleyn’s approach consists of expanding each vertex of $G$ into a gadget and reducing the Ising model inference to the problem of counting perfect matchings over the expanded graph. Kasteleyn’s construction was simplified by Fisher (1966). The tightest running time estimate for Kasteleyn’s method gives $O(N^{3/2})$. Kasteleyn proposed the conjecture, which was later proven by Gallucio and Loebl (1999), that the approach extends to the case of the zero-field Ising model over graphs embedded in a surface of genus $g$ with a multiplicative $O(4^g)$ penalty.

A parallel way of reducing the planar zero-field Ising model to a perfect matching counting problem consists of constructing the so-called expanded dual graph (Bieche \textit{et al} 1980, Barahona 1982, Schraudolph and Kamenetsky 2009). This approach is advantageous because using the expanded dual graph allows a one-to-one correspondence between spin configurations and perfect matchings. An extra advantage of this approach is that the reduction allows us to develop an exact efficient sampling. Based on linear algebra and planar separator theory (Lipton and Tarjan 1979), Wilson (1997) introduced an algorithm that allows the sampling of perfect matchings over planar graphs in $O\left(N^{3/2}\right)$ time. The algorithms were implemented by Thomas and Middleton (2009, 2013) for Ising model sampling; however, the implementation was limited to only the special case of a square lattice. Thomas and Middleton (2009) also suggested a simple extension of Wilson’s algorithm to the case of bounded genus graphs, again with the $4^g$ factor in complexity. Note that imposing the zero-field condition is critical, as otherwise, the Ising model over a planar graph is NP-hard (Barahona 1982). On the other hand, even in the case of zero magnetic field the Ising models over general graphs are difficult (Barahona 1982).

Wagner’s theorem (Diestel 2006, chapter 4.4) states that $G$ is planar only if it does not have $K_{3,3}$ and $K_5$ as minors (figure 2(b)). Both families of $K_{3,3}$-free and $K_5$-free graphs generalize and extend the family of planar graphs, since $K_{3,3}$ ($K_5$) is nonplanar but $K_5$-free ($K_{3,3}$-free). Both families are genus-unbounded, since a disconnected set of $gK_{3,3}$ ($K_5$) graphs has a genus of $g$ (Battle \textit{et al} 1962) and is $K_5$-free ($K_{3,3}$-free). Moreover, both families are treewidth-unbounded, since a planar square grid of size $t \times t$ has a treewidth of $t$ (Bodlaender 1998). Therefore, the question of interest becomes generalizing tractable inference and sampling in the zero-field Ising model over a $K_{3,3}$-free or $K_5$-free graph.

To extend the tractability of special cases as an approximation to a more general class of inference problems, it is natural to consider a family of tractable spanning subgraphs and then exploit the fact that the log-partition function $\log Z(\mu, J)$ is convex and hence can be upper-bounded by a linear combination of tractable partition functions.
tree-rewighted (TRW) approximation (Wainwright et al 2005) was the first example in the literature in which such upper-bounding was constructed with the trees used as a basic element. The upper-bound TRW approach (Wainwright et al 2005) was extended by Globerson and Jaakkola (2007), in which utilizing a planar spanning subgraph (and not a tree) as the basic (tractable) element was suggested.

**Contribution.** In this manuscript, we first compile results that were scattered across the literature on (at least) \( O(N^{3/2}) \)-efficient exact sampling and exact inference in the zero-field Ising model over planar graphs. To the best of our knowledge, we are the first to present a complete and mathematically accurate description of the tight asymptotic bounds.

Then, we describe a new family of zero-field Ising models on graphs that are more general than planar. Given a tree decomposition of such graphs into planar and ‘small’ \( O(1) \)-sized) components ‘glued’ together along sets of at most three vertices, inference and sampling over the new family of models is of polynomial time. We further show that all the \( K_{33} \)-free or \( K_5 \)-free graphs are included in this family and, moreover, their aforementioned tree decomposition can be constructed with \( O(N) \) efforts. This allows us to prove an \( O(N^{3/2}) \) upper bound on run time complexity for exact inference and sampling of the \( K_5 \)-free or \( K_{33} \)-free zero-field Ising models.

Finally, we show how the newly introduced tractable family of zero-field Ising models allows extension of the approach of Globerson and Jaakkola (2007) resulting in an upper-bound for log-partition function over general Ising models, including non-planar and nonzero magnetic field. Instead of using planar spanning subgraphs as in the work of Globerson and Jaakkola (2007), we use more general (non-planar) basic tractable elements. Using the methodology of Globerson and Jaakkola (2007), we illustrate the approach through experiments with a nonzero-field Ising model on a square grid for which exact inference is known to be NP-hard (Barahona 1982).

**Relation to other algorithms.** The result presented in this manuscript is similar to the approach used to count perfect matchings in \( K_5 \)-free graphs (Curticapean 2014, Straub et al 2014). However, we do not use a transition to perfect matching counting as it is typically done in studies of zero-field Ising models over planar graphs (Fisher 1966, Kasteleyn 1963, Thomas and Middleton 2009). Presumably, a direct transition to perfect matching counting can be done via construction of an expanded graph in the fashion of Fisher (1966), Kasteleyn (1963). However, this results in a size increase and, more importantly, there is no direct correspondence between spin configurations and perfect matchings; therefore, exact sampling is not supported.

**Structure.** Section 2 states the problems of exact inference and sampling for planar zero-field Ising models. In section 3 we introduce the concept of \( c \)-nice decomposition of graphs, and then formulate and prove tractability of the zero-field Ising models over graphs that are \( c \)-nice decomposable. Section 4 is devoted to application of the algorithm introduced in the preceding section to examples of the zero-field Ising model over the \( K_{33} \)-free (but possibly \( K - 5 \) containing) and \( K_5 \)-free (but possibly \( K_{3,3} \) containing) graphs. Section 5 presents an empirical application of the newly introduced family of tractable models to an upper-bounding log-partition function of a broader family of intractable graphical models (planar nonzero-field Ising models). Section 6 is reserved for conclusions.

https://doi.org/10.1088/1742-5468/abcaf1
Throughout the text, we use common graph-theoretic notations and definitions (Diestel 2006) and also restate the most important concepts briefly.

2. Planar topology

In this section, we consider the special $I = \langle G, 0, J \rangle$ case of the zero-field Ising model over a planar graph and introduce a transition from $I$ to the perfect matching model over a different (derived from $G$) planar graph. One-to-one correspondence between a spin configuration over the Ising model and corresponding perfect matching configuration over the derived graph translate the exact inference and exact sampling over $I$ to the corresponding exact inference and sampling in the derived perfect matching model.

2.1. Expanded dual graph

The graph is planar when it can be drawn on (embedded into) a plane without edge intersections. We assume that the planar embedding of $G$ is given (and if not, that it can be found in $O(N)$ time according to Boyer and Myrvold (2004)). In this section we follow the constructions of Schraudolph and Kamenetsky (2009).

Let us first triangulate $G$ by successively triangulating each face of the original graph and then setting $J_e = 0$ for all the newly added edges $e \in E(G)$. The complexity of the triangulation is $O(N)$; see Schraudolph and Kamenetsky (2009) for an example. (For convenience, we will then use the same notation for the derived, triangulated graph as for the original graph.)

Second, we construct a new graph, $G_F$, where each vertex $f \in V(G_F)$ is a face of $G$, and there is an edge $e = \{f_1, f_2\}$ in $E(G_F)$ only if $f_1$ and $f_2$ share an edge in $G$. By construction, $G_F$ is planar, and it is embedded in the same plane as $G$, so that each new edge $e = \{f_1, f_2\} \in E(G_F)$ intersects the respective old edge. We refer to $G_F$ as a dual graph of $G$. Since $G$ is triangulated, each $f \in V(G_F)$ has degree 3 in $G_F$.

Third, we obtain a planar graph $G^*$ and its embedding from $G_F$ by substituting each $f \in V(G_F)$ by a $K_3$ triangle so that each vertex of the triangle is incident to one edge, going outside the triangle (see figure 1(a) for illustration). We refer to $G^*$ as the expanded dual graph of $G$.

Newly introduced triangles of $G^*$, substituting $G_F$’s vertices, are called Fisher cities (Fisher 1966). We refer to edges outside triangles as intercity edges and denote their set as $E'_I$. The set $E(G^*) \setminus E'_I$ of Fisher city edges is denoted as $E'_C$. Note that $e^* \in E'_I$ intersects exactly one $e \in E(G)$ and vice versa, which defines a bijection between $E'_I$ and $E(G)$; this is denoted by $g : E'_I \rightarrow E(G)$. We observe that $|E'_I| = |E(G)| \leq 3N - 6$, where $N$ is the size (cardinality) of $G$.

A set $E' \subseteq E(G)$ is called a perfect matching (PM) of $G$, if edges of $E'$ are disjoint and their union equals $V$. Let $PM(G)$ denote the set of all perfect matchings (PM) of $G$. Notice that $E'_I$ is a PM of $G^*$, and thus $|V(G^*)| = 2|E'_I| = O(N)$. Since $G^*$ is planar, one also finds that $|E(G^*)| = O(N)$. Constructing $G^*$ requires $O(N)$ steps.

https://doi.org/10.1088/1742-5468/abcaf1
Figure 1. (a) A fragment of $G$’s embedding after triangulation (black); expanded dual graph $G^*$ (red). (b) Possible $X$ configurations and corresponding $M(X)$ (wavy lines) on a single face of $G$. Rotation symmetric and reverse sign configurations are omitted.

2.2. Perfect matching (PM) model

For every spin configuration $X \in \{\pm 1\}^N$, let $I(X)$ be a set $\{e \in E_f | g(e) = \{v, w\}, x_v = x_w\}$. Each Fisher city is incident to an odd number of edges in $I(X)$. Thus, $I(X)$ can be uniquely completed to a PM by edges from $E^*_C$. We denote the resulting PM as $M(X) \in \text{PM}(G^*)$ (see figure 1(b) for an illustration). Let $C_+ = \{+1\} \times \{\pm 1\}^{N-1}$.

**Lemma 1.** $M$ is a bijection between $C_+$ and $\text{PM}(G^*)$.

We define the weights on $G^*$ according to

$$\forall e^* \in E(G^*): c_{e^*} \triangleq \begin{cases} \exp(2J_g(e^*)), & e^* \in E_i^* \\ 1, & e^* \in E_c^* \end{cases}$$ (3)

**Lemma 2.** For $E' \in \text{PM}(G^*)$ it holds that

$$\mathbb{P}(M(X) = E') = \frac{1}{Z^*} \prod_{e^* \in E'} c_{e^*},$$ (4)

where

$$Z^* \triangleq \sum_{E' \in \text{PM}(G^*)} \prod_{e^* \in E'} c_{e^*} = \frac{1}{2} Z \exp \left( \sum_{e \in E(G)} J_e \right)$$ (5)

is the partition function of the PM distribution (PM model) defined by (4).

Proofs of lemmas 1 and 2 are shown in appendix A. The second transition of (5) reduces the problem of computing $Z$ to computing $Z^*$. Furthermore, only two equiprobable spin configurations $X'$ and $-X'$ (one of which is in $C_+$) correspond to $E'$, and they can be recovered from $E'$ in $O(N)$ steps, thus resulting in the statement that one samples from $I$ if sampling from (4) is known.
The PM model can be defined for an arbitrary graph $\hat{G}$, $\hat{N} = |V(\hat{G})|$ with positive weights $c_e, e \in E'$, as a probability distribution over $\hat{M} \in \text{PM}(\hat{G})$: $P(\hat{M}) \propto \prod_{e \in \hat{M}} c_e$. Our subsequent derivations are based on the following.

**Theorem 3.** Given the PM model defined on planar graph $\hat{G}$ of size $\hat{N}$ with positive edge weights $\{c_e\}$, one can find its partition function and sample from it in $O(\hat{N}^3)$ time (steps).

Algorithms constructively proving the theorem are directly inferred from Wilson (1997), Thomas and Middleton (2009), with minor changes/generalizations. We describe the algorithms in appendix B.

**Corollary 4.** Exact inference and sampling of the PM model over $G^*$ (and, hence, zero-field Ising model $I$ over the planar graph $G$) take $O(N^3)$ time.

### 3. c-nice decomposition of the topology

We commence by introducing the concept of $c$-nice decomposition of a graph and stating the main result on the tractability of the new family of Ising models in subsection 3.1. Then, we proceed to build a helpful ‘conditioning’ machinery in subsection 3.2 and subsequently describe algorithms for the efficient exact inference (subsection 3.3) and sampling (subsection 3.4), therefore proving the aforementioned statement constructively.

#### 3.1. Decomposition tree and the key result (of the manuscript)

We mainly follow Curticapean (2014), Reed and Li (2008) in the definition of the decomposition tree and its properties sufficient for our goals. (Let us also reiterate that we consider here graphs containing no self-loops or multiple edges.)

Graph $G'$ is a subgraph of $G$ whenever $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. For two subgraphs $G'$ and $G''$ of $G$, let $G' \cup G'' = (V(G') \cup V(G''), E(G') \cup E(G''))$ (graph union).

Consider a tree decomposition $T = (T, G)$ of a graph $G$ into a set of subgraphs $G \triangleq \{G_t\}$ of $G$, where $t$ are nodes of a tree $T$, that is, $t \in V(T)$. One of the nodes of the tree, $r \in V(T)$, is selected as the root. For each node $t \in V(T)$, its parent is the first node on the unique path from $t$ to $r$. $G_{\leq t}$ denotes the graph union of $G_r$ for all the nodes $t'$ in $V(T)$ that are $t$ or its descendants. $G_{\geq t}$ denotes the graph union of $G_t$ for all the nodes $t'$ in $V(T)$ that are neither $t$ nor descendants of $t$. For two neighboring nodes of the tree, $t, p \in V(T)$ and $\{t, p\} \in E(T)$, the set of overlapping vertices of $G_t$ and $G_p$, $K \triangleq V(G_t) \cap V(G_p)$, is called an attachment set of $t$ or $p$. If $p$ is a parent of $t$, then $K$ is a navel of $t$. We assume that the navel of the root is empty.

$T$ is a c-nice decomposition of $G$ if the following requirements are satisfied:

(a) $\forall t \in V(T)$ with a navel $K$, it holds that $K = V(G_{\leq t}) \cap V(G_{\geq t})$. 

https://doi.org/10.1088/1742-5468/abcaf1
Figure 2. (a) Exemplar graph $G$ and its eight-nice decomposition $\mathcal{T}$, where $t \in \{1, \ldots, 7\}$ labels nodes of the decomposition tree $T$ and node 4 is chosen as the root ($r = 4$). Identical vertices of $G$ in its subgraphs $G_t$ are shown connected by dashed lines. Navels of size 1, 2, and 3 are highlighted. Component $G_5$ is nonplanar, and $G_4$ becomes nonplanar when all attachment edges are added (according to the fourth item of the definition of the $c$-nice decomposition). $G_{\leq 3}$ and $G_{\geq 3}$ are shown with dotted lines. Note that the decomposition is non-unique for the graph. For instance, edges that belong to the attachment set can go to either of the two subgraphs containing this set or even repeat in both. (b) Minors $K_5$ and $K_{33}$ are forbidden in the planar graphs. The Möbius ladder and its subgraphs are the only nonplanar graphs allowed in the eight-nice decomposition of a $K_5$-free graph. (c) The left panel is an example of conditioning on three vertices/spins in the center of a graph. The right panel shows a modified graph where the three vertices (from the left panel) are reduced to one vertex, then leading to a modification of the pairwise interactions within the associated zero-field Ising model over the reduced graph. (d) Example of a graph that contains $K_5$ as a minor: by contracting the highlighted groups of vertices and deleting the remaining vertices, one arrives at the $K_5$ graph.

(b) Every attachment set $K$ is of size 0, 1, 2, or 3.
(c) $\forall t \in V(T)$, either $|V(G_t)| \leq c$ or $G_t$ is planar.
(d) If $t \in V(T)$ is such that $|V(G_t)| > c$, addition of all edges of type $e = \{v, w\}$, where $v, w$ belong to the same attachment set of $t$ (if $e$ is not yet in $E(G_t)$), does not destroy planarity of $G_t$.

Stated informally, the $c$-nice decomposition of $G$ is a tree decomposition of $G$ into planar and ‘small’ (of size at most $c$) subgraphs $G_t$, ‘glued’ via subsets of at most three vertices of $G$. Figure 2(a) shows an example of a $c$-nice decomposition with $c = 8$. There are various similar ways to define a graph decomposition in the literature, and the one presented above is customized (to our purposes) to include only properties significant for our consecutive analysis.

The remainder of this section is devoted to a constructive proof of the following key statement of the manuscript.

https://doi.org/10.1088/1742-5468/abcaf1
Theorem 5. Let \( \mathcal{I} = \langle G, 0, J \rangle \) be any zero-field Ising model where there exists a \( c \)-nice decomposition \( \mathcal{T} \) of \( G \), where \( c \) is an absolute constant. Then, there is an algorithm which, given \( \mathcal{I}, \mathcal{T} \) as an input: (1) finds \( Z \) and (2) samples a configuration from \( \mathcal{I} \) in time \( O(\sum_{t \in V(T)} |V(G_t)|^3) \).

3.2. Inference and sampling conditioned on one, two, or three vertices/spins

Before presenting the algorithm that proves theorem 5 constructively, let us introduce the auxiliary machinery of ‘conditioning’, which describes the partition function of a zero-field Ising model over a planar graph conditioned on one, two, or three spins. We consider a zero-field Ising model \( \mathcal{I} = \langle G, 0, J \rangle \) defined over a planar graph \( G \). We intend to use the algorithm for efficient inference and sampling of \( \mathcal{I} \) as a black box in our subsequent derivations.

Let us now introduce the notion of conditioning. We consider a spin configuration \( X \in \{\pm 1\}^N \), a subset \( V = \{v^{(1)}, \ldots, v^{(\omega)}\} \subseteq V(G) \), and define a condition \( S = \{x_{v^{(i)}} = s^{(i)}, \ldots, x_{v^{(\omega)}} = s^{(\omega)}\} \) on \( V \), where \( s^{(1)}, \ldots, s^{(\omega)} = \pm 1 \) are fixed values. Conditional versions of the probability distribution (1–2) and the conditional partition function become

\[
\mathbb{P}(X|S) \triangleq \frac{W(X) \times \mathbb{I}(X|S)}{Z|S}, \quad \mathbb{I}(X|S) \triangleq \left\{ \begin{array}{ll}
1, & x_{v^{(i)}} = s^{(i)}, \ldots, x_{v^{(\omega)}} = s^{(\omega)} \\
0, & \text{otherwise}
\end{array} \right.
\]

(6)

where \( Z|S \triangleq \sum_{X \in \{\pm 1\}^N} W(X) \times \mathbb{I}(X|S) \).

Note that when \( \omega = 0 \), \( S = \{\} \) and (6–7) is reduced to (1–2). The subset of \( V(G) \) is connected whenever the subgraph induced by this subset is connected. Inference and sampling of \( \mathcal{I} \) can be extended as follows (a formal proof can be found in appendix A).

Lemma 6. Given \( \mathcal{I} = \langle G, 0, J \rangle \), where \( G \) is planar and a condition \( S \) on a connected subset \( V' \subseteq V(G) \), \( |V'| \leq 3 \), computing the conditional partition function \( Z|S \) and sampling from \( \mathbb{P}(X|S) \) are tasks of \( O(N^{\frac{3}{2}}) \) complexity.

Intuitively, the conditioning algorithm proving the lemma takes the subset of connected vertices and ‘collapses’ them into a single vertex. The graph remains planar and the task is reduced to conditioning on one vertex, which is an elementary operation given the algorithm from section 2. (See figure 2(c) for an illustration.)

3.3. Inference algorithm

This subsection constructively proves the inference part of theorem 5. For each \( t \in V(T) \), let \( \mathcal{I}_{\leq t} \triangleq \langle G_{\leq t}, 0, \{J_e \mid e \in E(G_{\leq t}) \subseteq E(G)\} \rangle \) denote a zero-field Ising submodel induced...
by $G_{<t}$. We denote the partition function and subvector of $X$ related to $I_{<t}$ as $Z_{<t}$ and $X_{<t} \triangleq \{ x_v | v \in V(G_{<t}) \}$, respectively.

Further, let $K$ be $t$’s navel and let $S = \{ \forall v \in K : x_v = s(v) \}$ denote some condition on $K$. Recall that $|K| \leq 3$. For each $t$, the algorithm computes conditional partition functions $Z_{<t}^{S}$ for all choices of condition spin values $\{ s(v) = \pm 1 \}$. Each $t$ is processed only when its children have already been processed, so the algorithm starts at the leaf and ends at the root. If $r \in G(T)$ is a root, its navel is empty and $G_{<r} = G$; hence, $Z = Z_{<r}^{S}$ is computed after $r$’s processing.

Suppose all children of $t$, $c_1, \ldots, c_m \in V(T)$ with navel $K_1, \ldots, K_m \subseteq V(G_t)$ have already been processed, and now $t$ itself is considered. We denote a spin configuration on $G_t$ by $Y_t \triangleq \{ y_v = \pm 1 | v \in V(G_t) \}$. $I_{<t}, \ldots, I_{<m}$ are $I_{<t}$’s submodels induced by $G_{<c_1}, \ldots, G_{<c_m}$, which can only intersect at their navels in $G_t$. Based on this, one states the following dynamic programming relation:

$$Z_{|S|}^{<t} = \sum_{Y_t \in \{ \pm 1 \}^{V(G_t)}} \prod_{i=1}^{m} Z_{|S_i|}^{c_i},$$

(8)

Here, $S_i[Y_t]$ denotes a condition $\{ \forall v \in K_i : x_v = y_v \}$ on $K_i$. The goal is to efficiently perform summation in (8). Let $I^{(0)}, I^{(1)}, I^{(2)}, I^{(3)}$ be a partition of $\{ 1, \ldots, m \}$ by navel sizes. Figures 3(a) and (b) illustrate inference in $t$.

(a) **Navels of size 0, 1.** Note that if $i \in I^{(0)}$, then $Z_{|I^{(0)}|}^{c_i} = Z_{<c_i}$ is constant, as computed before. The same is true for $i \in I^{(1)}$ and $Z_{|S_0|}^{c_i} = \frac{1}{2} Z_{<c_i}$.

(b) **Navels of size 2.** Let $i \in I^{(2)}$ denote $K_i = \{ u', q' \}$ and simplify notation $Z_{|S_i|}^{c_i} \triangleq Z_{x_i=1,y_i=1}^{c_i}$ for convenience. Note that $Z_{|S_i|}^{c_i}$ is strictly positive, and due to the zero-field nature of $I_{<c_i}$, one finds $Z_{|S_i|}^{c_i} = Z_{|S_i|}^{c_i} = Z_{|S_i|}^{c_i}$. Then, one arrives at $\log Z_{|S_i|}^{c_i} = A_i + B_i y_i y_i'$, where $A_i \triangleq \log Z_{|S_i|}^{c_i}$ and $B_i \triangleq \log Z_{|S_i|}^{c_i}$. Due to the zero-field nature of $I_{<c_i}$, it holds that $Z_{|S_i|}^{c_i} = Z_{|S_i|}^{c_i}$. We observe that there are such $A_i, B_i, C_i, D_i$ that $Z_{|S_i|}^{c_i} = A_i + B_i y_1 y_2 + C_i y_1 y_3 + D_i y_2 y_3$ for all $y_1, y_2, y_3 = \pm 1$, which is guaranteed since the following system of equations has a solution:

$$\begin{bmatrix}
\log Z_{|S_i|}^{c_i} \\
\log Z_{|S_i|}^{c_i} \\
\log Z_{|S_i|}^{c_i} \\
\log Z_{|S_i|}^{c_i}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix} \times \begin{bmatrix}
A_i \\
B_i \\
C_i \\
D_i
\end{bmatrix}. $$

(9)

https://doi.org/10.1088/1742-5468/abcaf1
Figure 3. (a) Example of inference at node $t$ with children $c_1, c_2, c_3, c_4$. Navel sets $K_1 = \{u^1, q^1, h^1\}$, $K_2 = \{u^2, q^2, h^2\}$, $K_3 = \{u^3, q^3\}$, $K_4 = \{u^4\}$, and $K = \{u, q, h\}$ are highlighted. Fragments of $I^c \leq t$ are shown with dotted lines. Here, $I^{(0)} = \emptyset$, $I^{(1)} = \{4\}$, $I^{(2)} = \{3\}$, and $I^{(3)} = \{1, 2\}$, indicating that one child is glued over one node, one child is glued over two nodes, and two children are glued over three nodes. (b) ‘Aggregated’ Ising model $I_t$ and its pairwise interactions are shown. Both (c) and (d) illustrate sampling over $I_t$. One sample spins in $I_t$ conditioned on $S^{(t)}$ and then repeats the procedure at the child nodes.

Considering these three cases, one rewrites equation (8) as

$$Z^{(t)}_{I_S} = M \cdot \prod_{Y_t} \left( \sum_{y_S} \exp \left( \sum_{e \in E(G_t)} J_{eyv}y_vy_w + \sum_{i \in I^{(2)} \cup I^{(3)}} B_i y_{ii'}y_{ii'} + \sum_{i \in I^{(3)}} (C_i y_{ii'}y_{ii'} + D_i y_{ii'}y_{ii'}) \right) \right),$$

where $M \triangleq 2^{-|I^{(3)}|} \cdot (\prod_{j \in I^{(0)} \cup I^{(1)}} Z^{(t)}_{I_S}) \cdot \exp(\sum_{i \in I^{(2)} \cup I^{(3)}} A_i)$. The sum in equation (10) is simply a conditional partition function of a zero-field Ising model $I_t$ defined over a graph $G_t$ with pairwise interactions of $I$ adjusted by the addition of $B_i, C_i$, and $D_i$ summands at the appropriate navel edges (if a corresponding edge is not present in $G_t$, it has to be added). If $|V(G_t)| \leq c$, then (10) is computed a maximum of four times (depending on navel size) by brute force ($O(1)$ time). Otherwise, if $K$ is a disconnected set in $G_t$, we add zero-interaction edges inside it to make it connected. Possible addition of edges inside $K, K_1, \ldots, K_m$ does not destroy planarity according to the fourth item in the definition of the $c$-nice decomposition above. Finally, we compute (10) using lemma 6 in time $O(|V(G_t)|^3)$.

The inference part of theorem 5 follows directly from the procedure just described.

### 3.4. Sampling algorithm

Next, we address the sampling part of theorem 5. We extend the algorithm from section 3.3 so that it supports efficient sampling from $I$. We assume that the inference pass through $T$ (from leaves to root) has been done so that $I_t$ for all $t \in V(T)$ are computed. We denote $X_t \triangleq \{x_v | v \in V(G_t)\}$. The sampling algorithm runs backward,
first drawing spin values $X_r$ at the root $r$ of $T$ from the marginal distribution $\mathbb{P}(X_r)$, and then processing each node $t$ of $T$ after its parent $p$ is processed. Processing consists of drawing spins $X_t$ from $\mathbb{P}(X_t \mid X_p) = \mathbb{P}(X_t \mid X(t)) = \{x_v \mid v \in K\}$, where $K$ is a navel of $t$. This marginal-conditional scheme generates the correct sample $X$ of spins over $G$.

Let $\mathbb{P}_{\leq t}(X_{\leq t})$ define a spin distribution of $I_{\leq t}$. Because the Ising model is an example of a Markov random field, it holds that $\mathbb{P}_{\leq t}(X_{\leq t} \mid X(t)) = \mathbb{P}(X_{\leq t} \mid X(t))$. We further derive

$$
\mathbb{P}(X_t \mid X(t)) = \mathbb{P}_{\leq t}(X_t \mid X(t)) = \frac{1}{Z_{\leq t}} \sum_{X_t \setminus X_t} \exp \left( \sum_{e = \{v,w\} \in E(G_{\leq t})} J_{e}x_vx_w \right)
$$

$$
= \frac{1}{Z_{\leq t}} \cdot \exp \left( \sum_{e = \{v,w\} \in E(G_{t})} J_{e}x_vx_w \right) \cdot \prod_{i=1}^{m} \mathbb{P}^{\leq c_i}{_{\mid S_i \mid X_t}}
$$

$$
\propto \exp \left( \sum_{e = \{v,w\} \in E(G_{t})} J_{e}x_vx_w + \sum_{i \in I(2) \cup I(3)} B_{i}x_{ui}x_{qi} + \sum_{i \in I(3)} (C_{i}x_{ui}x_{hi} + D_{i}x_{qi}x_{hi}) \right).
$$

In other words, sampling from $\mathbb{P}(X_t \mid X(t))$ is reduced to sampling from $\mathcal{I}_t$ conditional on spins $X(t)$ in the navel $K$. This is done via brute force if $|V(G_{t})| \leq c$; otherwise, lemma 6 allows one to draw $X_t$ in $O(|V(G_t)|^4)$, since $|K| \leq 3$. Sampling efforts cost as much as inference, which concludes the proof of theorem 5. Figures 3(c) and (d) illustrate sampling in $t$.

4. Minor-free extension of planar zero-field Ising models

Contraction is an operation of removing two adjacent vertices $v$ and $u$ (and all edges incident to them) from the graph and adding a new vertex $w$ adjacent to all neighbors of $v$ and $u$. For two graphs $G$ and $H$, $H$ is $G$’s minor if it is isomorphic to a graph obtained from $G$’s subgraph by a series of contractions (figure 2(d)). $G$ is $H$-free if $H$ is not $G$’s minor.

According to Wagner’s theorem (Diestel 2006, chapter 4.4), a set of planar graphs coincides with an intersection of $K_{33}$-free graphs and $K_{5,5}$-free graphs. Some nonplanar graphs are $K_{33}$-free ($K_{5,5}$-free); for example, $K_{5}$ ($K_{33}$). $K_{33}$-free ($K_{5,5}$-free) graphs are not genus-bounded (a disconnected set of $gK_{5}$ ($K_{33}$) graphs is $K_{33}$-free ($K_{5,5}$-free) and has a genus of $g$ (Battle et al 1962)). $K_{33}$-free ($K_{5,5}$-free) graphs are treewidth-unbounded as well (a planar square grid of size $t \times t$ is $K_{33}$-free and $K_{5,5}$-free and has a treewidth of $t$ (Bodlaender 1998)). In the remainder of the section we show that a $c$-nice decomposition of $K_{33}$-free graphs and $K_{5,5}$-free graphs can be computed in polynomial time and, hence, inference and sampling of zero-field Ising models on these graph families can be performed efficiently.

https://doi.org/10.1088/1742-5468/abcaf1
4.1. Zero-field Ising models over $K_{33}$-free graphs

Even though $K_{33}$-free graphs are Pfaffian-orientable (with the Pfaffian orientation computable in polynomial time; see Vazirani (1989)), the expanded dual graph—introduced to map the zero-field Ising model to the respective PM problem—is not necessarily $K_{33}$-free. Therefore, the latter is generally not Pfaffian-orientable. Hence, the reduction to a well-studied perfect matching counting problem is not straightforward.

**Theorem 7.** Let $G$ be $K_{33}$-free graph of size $N$ with no loops or multiple edges. Then, the five-nice decomposition $T$ of $G$ exists and can be computed in time $O(N)$.

**Proof (sketch).** An equivalent decomposition is constructed by Hopcroft and Tarjan (1973), Gutwenger and Mutzel (2001), Vo (1983) in time $O(N)$. We present a formal proof in appendix C.

**Remark 8.** The $O(N)$ construction time of $T$ guarantees that $\sum_{t \in V(T)} |V(G_t)| = O(N)$. All nonplanar components in $T$ are isomorphic to $K_5$ or its subgraph.

Therefore, if $G$ is $K_{33}$-free, it satisfies all the conditions needed for efficient inference and sampling, as described in section 3.

**Theorem 9.** For any $I = \langle G, 0, J \rangle$ where $G$ is $K_{33}$-free, inference or sampling of $I$ takes $O\left(\frac{N^3}{2}\right)$ steps.

**Proof.** Finding five-nice $T$ for $G$ is the $O(N)$ operation. Provided with $T$, inference and sampling take at most

$$O\left(\sum_{t \in V(T)} |V(G_t)|^{\frac{3}{2}}\right) = O\left(\left(\sum_{t \in V(T)} |V(G_t)|\right)^{\frac{3}{2}}\right) = O\left(N^{\frac{3}{2}}\right)$$

(12)

where we apply convexity of $f(z) = z^{\frac{3}{2}}$ and the remark after theorem 7.

4.2. $K_{33}$-free zero-field Ising models: implementation and tests

In addition to theoretical justification, which is fully presented in this manuscript, we perform empirical simulations to validate the correctness of the inference and sampling algorithm for $K_{33}$-free zero-field Ising models.

To test the correctness of inference, we generate random $K_{33}$-free models of a given size and then compare the value of PF computed in a brute-force way (tractable for sufficiently small graphs) and by our algorithm. See the graph generation algorithm in appendix E. We simulate samples of sizes from \{10, \ldots, 15\} (1000 samples per size) and verify that respective expressions coincide.

When testing sampling implementation, we take for granted that the produced samples do not correlate given that the sampling procedure accepts the Ising model as an input and uses an independent random number generator inside. The construction does not have any memory; therefore, it generates statistically independent samples. To test that the empirical distribution is approaching a theoretical one (in the limit of an infinite number of samples), we draw different numbers $m$ of samples from a model of size...
Tractable minor-free generalization of planar zero-field Ising models

Figure 4. (a) KL-divergence of the model probability distribution compared with the empirical probability distribution. $N, m$ are the model size and the number of samples, respectively. (b) Execution time of inference (red dots) and sampling (blue dots) depending on $N$, shown on a logarithmic scale. The black line corresponds to $O(N^{3/2})$.

Then, we find Kullback–Leibler divergence between the probability distribution of the model (here we use our inference algorithm to compute the normalization $Z$) and the empirical probability obtained from samples. Figure 4(a) shows that KL-divergence converges to zero as the sample size increases. Zero KL-divergence corresponds to equal distributions.

Finally, we simulate inference and sampling for random models of different sizes $N$ and observe that the computational time (efforts) scales as $O\left(N^{3/2}\right)$ (figure 4(b))

4.3. Zero-field Ising models over $K_5$-free graphs

It can be shown that a result similar to the one described above for the $K_{33}$-free graphs also holds for the $K_5$-free graphs.

**Theorem 10.** Let $G$ be a $K_5$-free graph of size $N$ with no loops or multiple edges. Then, the eight-nice decomposition $T$ of $G$ exists and can be computed in time $O(N)$.

**Proof (sketch).** An equivalent decomposition is constructed by Reed and Li (2008) in time $O(N)$. See appendix D for formal proof.

**Remark 11.** The $O(N)$ construction time of $T$ guarantees that $\sum_{t \in V(T)} |V(G_t)| = O(N)$. All nonplanar components in $T$ are isomorphic to the Möbius ladder (figure 2(b)) or its subgraph.

The graph in figure 2(a) is actually $K_5$-free. Theorems 5 and 10 allow us to conclude:

**Theorem 12.** Given $\mathcal{I} = (G, 0, J)$ with $K_5$-free $G$ of size $N$, finding $Z$ and sampling from $\mathcal{I}$ take $O\left(N^{3/2}\right)$ total time.

\footnote{Implementation of the algorithms is available at https://github.com/ValeryTyumen/planarising.}

https://doi.org/10.1088/1742-5468/abcaf1
Figure 5. Construction of graphs used for approximate inference on a rectangular lattice. For better visualization, vertices connected to an apex are colored white. (a) $G'$ graph. (b) One of the planar $G^{(r)}$ graphs used in Globerson and Jaakkola (2007). This ‘separator’ pattern is repeated for each column and row, resulting in $2(H - 1)$ graphs in $\{G^{(r)}\}$. In addition, Globerson and Jaakkola (2007) add an independent variables graph in which only apex edges are drawn. (c) A modified ‘separator’ pattern we propose. Again, the pattern is repeated horizontally and vertically resulting in $2(H - 2)$ graphs + independent variables graph. This pattern covers more magnetic fields and connects separated parts. Dashed edges indicate the structure of ten-nice decomposition used for inference. (A nonplanar node of size 10 is illustrated on the right.)

**Proof.** Analogous to the proof of theorem 9. □

5. Approximate inference of square-grid Ising model

In this section, we consider $\mathcal{I} = \langle G, \mu, J \rangle$ such that $G$ is a square-grid graph of size $H \times H$. Finding $Z(G, \mu, J)$ for arbitrary $\mu, J$ is an NP-hard problem (Barahona 1982) in such a setting. We construct $G'$ by adding an apex vertex connected to all $G$’s vertices by edge (figure 5(a)). Now it can easily be seen that $Z(G, \mu, J) = \frac{1}{2}Z(G', 0, J' = (J_\mu \cup J))$, where $J_\mu = \mu$ are interactions assigned to apex edges.

Let $\{G^{(r)}\}$ be a family of spanning graphs $(V(G^{(r)}) = V(G'), \ E(G^{(r)}) \subseteq E(G'))$ and let $J^{(r)}$ be interaction values on $G^{(r)}$. We also denote $J^{(r)} = J^{(r)} \cup \{0, e \in E(G') \setminus E(G^{(r)})\}$. Assuming that $\log Z(G^{(r)}, 0, J^{(r)})$ are tractable, the convexity of $\log Z(G', 0, J')$ allows one to write the following upper bound:

$$
\log Z(G', 0, J') \leq \min_{\rho(r) \geq 0, \sum \rho(r) = 1} \sum_r \rho(r) \log Z(G^{(r)}, 0, J^{(r)}).
$$

After graph set $\{G^{(r)}\}$ has been fixed, one can numerically optimize the right-hand side of (13), as shown in Globerson and Jaakkola (2007) for planar $G^{(r)}$. The extension of the basic planar case is straightforward and is detailed in appendix F. The appendix also
Figure 6. Comparison of TRW, PSG, and DSG approaches. The first plot is for normalized log-partition error, the second is for error in pairwise marginals, and the third is for error in a singleton central marginal. Standard errors over 100 trials are shown as error bars. An asterisk ‘∗’ indicates the statistically significant improvement of DSG over PSG, with a p-value smaller than 0.01 according to the Wilcoxon test with the Bonferroni correction (Wilcoxon 1945).

contains the description of marginal probabilities approximation suggested in Globerson and Jaakkola (2007), Wainwright et al (2005).

The choice of a planar spanning graph (PSG) family \( \{G^{(r)}\} \) of Globerson and Jaakkola (2007) is illustrated in figure 5(b). A tractable decomposition-based extension of the planar case presented in this manuscript suggests a more advanced construction—decomposition-based spanning graphs (DSG) (figure 5(c)). We compare performance of both PSG and DSG approaches as well as the performance of the TRW approximation (Wainwright et al 2005) in the following setting of varying interaction:

\[
\mu \sim \mathcal{U}(-0.5, 0.5), \quad J \sim \mathcal{U}(-\alpha, \alpha), \quad \alpha \in \{1, 1.2, 1.4, \ldots, 3\}.
\]

We optimize for grid size \( H = 15 \) (225 vertices, 420 edges) and compare the upper bounds and marginal probability approximations (superscript \textit{alg}) with the exact values obtained using a junction tree algorithm (Verner Jensen et al 1990) (superscript \textit{true}). We compute three types of error:

(a) Normalized log-partition error \( \frac{1}{|E(G)|} \sum_{\{v,w\} \in E(G)} \log |P^{\text{true}}(x_v x_w = 1) - P^{\text{true}}(x_v x_w = 1)| \),

(b) Error in pairwise marginals \( \frac{1}{|E(G)|} \sum_{\{v,w\} \in E(G)} |P^{\text{true}}(x_v x_w = 1) - P^{\text{true}}(x_v x_w = 1)| \),

and

(c) Error in singleton central marginal \( |P^{\text{true}}(x_v = 1) - P^{\text{true}}(x_v = 1)| \), where \( v \) is a vertex of \( G \) with coordinates (8, 8).

We average results over 100 trials (see figure 6). We use the same quasi-Newton algorithm (Bertsekas 1999) and parameters when optimizing (13) for PSG and DSG, but for most settings, DSG outperforms PSG and TRW. Cases with smaller TRW error can be explained by the fact that TRW implicitly optimizes (13) over the family of all spanning trees, which can be exponentially big in size, whereas for PSG and DSG we only use \( O(H) \) spanning graphs.

\[ ^5 \text{Hardware used: 24-core Intel® Xeon® Gold 6136 CPU@3.00 GHz.} \]

\[ ^6 \text{Implementation of the algorithms is available at https://github.com/ValeryTyumen/planarising.} \]

https://doi.org/10.1088/1742-5468/abcaf1
Because PSG and DSG approaches come close to each other, we additionally test for each value of $\alpha$ on each plot, whether the difference $\text{err}_{\text{PSG}} - \text{err}_{\text{DSG}}$ is bigger than zero. We apply a one-sided Wilcoxon test (Wilcoxon 1945) together with the Bonferroni correction because we test 33 times (Jean Dunn 1961). In most settings, the improvement is statistically significant (figure 6).

6. Conclusion

In this manuscript, we first describe an algorithm for $O\left(N^{3/2}\right)$ inference and sampling of planar zero-field Ising models on $N$ spins. Then, we introduce a new family of zero-field Ising models composed of planar components and graphs of $O(1)$ size. For these models, we describe a polynomial algorithm for exact inference and sampling provided that the decomposition tree is also in the input. A theoretical application is an $O(N^{3/2})$ inference and sampling algorithm for $K_{33}$-free or $K_{5}$-free zero-field Ising models—both families are supersets of the family of planar zero-field models, and both are neither treewidth- nor genus-bounded. We show that our scheme offers an improvement on the approximate inference scheme for arbitrary topologies. The suggested improvement is based on the PSG ideas from Globerson and Jaakkola (2007), but we use tractable spanning decomposition-based graphs instead of planar graphs. (That is, we keep the algorithm of Globerson and Jaakkola (2007), but substitute planar graphs with a family of spanning decomposition-based graphs that are tractable.) This improvement of Globerson and Jaakkola (2007) results in a tighter upper bound on the true partition function and a more precise approximation of marginal probabilities.

Appendix A. Lemma proofs

A.1. Lemma 1

**Proof.** Let $E' \in \text{PM}(G^*)$. We refer to $e \in E$ as saturated if it intersects an edge from $E' \cap E^*_I$. Each Fisher city is incident to an odd number of edges in $E' \cap E^*_I$. Thus, each face of $G$ has an even number of unsaturated edges. This property is preserved when two faces/cycles are merged into one by evaluating the respective symmetric difference. Therefore, one finds that any cycle in $G$ has an even number of unsaturated edges.

For each $i$ we define $x_i := -1^r_i$, where $r_i$ is the number of unsaturated edges on the path connecting $v_1$ and $v_i$. The definition is consistent due to the aforementioned cycle property. Now for each $e = \{v, w\} \in E(G)$, $x_v = x_w$ only if $e$ is saturated. To conclude, we construct $X$ such that $E' = M(X)$. This $X$ is unique, because the parity of unsaturated edges on a path between $v_1$ and $v_i$ uniquely determines the relationship between $x_1$ and $x_i$ and $x_1$ is always +1.

A.2. Lemma 2

**Proof.** Let $X' = (x'_1, \ldots, x'_N) \in C_+$, $M(X') = E'$. The statement is justified by the following chain of transitions:
Tractable minor-free generalization of planar zero-field Ising models

\[ \mathbb{P}(M(S) = E') = \mathbb{P}(S = X') + \mathbb{P}(S = -X') \]

\[ = \frac{2}{Z} \exp \left( \sum_{e = \{v, w\} \in E(G)} J_{e} x'_{v} x'_{w} \right) \]

\[ = \frac{2}{Z} \exp \left( \sum_{e^* \in E' \cap E'_I} 2J_{g(e^*)} - \sum_{e \in E(G)} J_{e} \right) \]

\[ = \frac{2}{Z} \exp \left( -\sum_{e \in E(G)} J_{e} \right) \prod_{e^* \in E' \cap E'_I} c_{e^*} \]

\[ = \frac{1}{Z} \prod_{e^* \in E'} c_{e^*}. \]  \hspace{1cm} (14)

\[ \square \]

A.3. Lemma 6

**Proof.** We consider cases depending on \( \omega \) and consequently reduce each case to a simpler one. For convenience, in cases where the following apply we denote \( u \triangleq v^{(1)}, h \triangleq v^{(2)}, q \triangleq v^{(3)} \):

(a) **Conditioning on \( \omega = 0 \) spins.** Trivial given the algorithm described in section 2.

(b) **Conditioning on \( \omega = 1 \) spin.** Since configurations \( X \) and \( -X \) have the same probability in \( I \), one deduces that \( Z_{| x_u = s^{(1)}}, x_h = s^{(2)} } = \frac{1}{2} Z \).

One also deduces that sampling \( X \) from \( \mathbb{P}(X \mid x_u = s^{(1)}) \) is reduced to (1) drawing \( \overline{X} = \{ \overline{x}_v = \pm 1 \} \) from \( \mathbb{P}(X) \) and then (2) returning \( X = (s^{(1)} \overline{x}_u) \cdot \overline{X} \) as a result.

(c) **Conditioning on \( \omega = 2 \) spins.** There is an edge \( e^0 = \{u, h\} \in E(G) \). The following expansion holds:

\[ Z_{| x_u = s^{(1)}, x_h = s^{(2)}} = \sum_{X, x_u = s^{(1)}, x_h = s^{(2)}} \exp \left( \sum_{e = \{v, w\} \in E(G)} J_{e} x_{v} x_{w} \right) \]

\[ = \exp(J_{e^0}s^{(1)}s^{(2)}) \cdot \sum_{X, x_u = s^{(1)}, x_h = s^{(2)}} \exp \left( \sum_{e = \{v, w\} \in E(G) \setminus e^0} J_{e} x_{v} x_{w} \right) \]
We obtain graph $G'$ from $G$ by contracting $u, h$ into $z$. $G'$ is still planar and has $N-1$ vertices. We preserve the pairwise interactions of edges that were not deleted after contraction. For each edge $e = \{u, v\}$, $v \neq h$ we set $J_{\{z, v\}} = J_e s^{(1)}$, and for each edge $e = \{h, v\}$, $v \neq u$ we set $J_{\{z, v\}} = J_e s^{(2)}$. We collapse the double edges in $G'$ that were possibly created by transforming into single edges. A pairwise interaction of the result edge is set to the sum of collapsed interactions.

We define a zero-field Ising model $\mathcal{Z}'$ on the resulting graph $G'$ with its pairwise interactions, inducing a distribution $\mathbb{P}'(X' = \{x'_v = \pm 1 | v \in V(G')\})$. Let $\mathcal{Z}'$ denote $\mathcal{Z}'$’s partition function. A closer look at (15) reveals that

$\left| Z_{x_u = s^{(1)}, x_h = s^{(2)}} = \exp(J_{e} s^{(1)} s^{(2)}) \cdot Z'_{x'_z = 1} \right|_{x'_z = 1}$

where $Z'_{x'_z = 1}$ is a partition function conditioned on a single spin and can be found efficiently as shown above.

Since the equality of sums (16) holds summand-wise, for a given $X'' = \{x''_v = \pm 1 | v \in V(G) \setminus \{u, h\}\}$ the probabilities $\mathbb{P}(X'' \cup \{x_u = s^{(1)}, x_h = s^{(2)}\} | x_u = s^{(1)}, x_h = s^{(2)})$ and $\mathbb{P}(X'' \cup \{x'_v = 1\} | x'_v = 1)$ are the same. Hence, sampling from $\mathbb{P}(X | x_u = s^{(1)}, x_h = s^{(2)})$ is reduced to conditional sampling from a planar zero-field Ising model $\mathbb{P}'(X' | x'_z = 1)$ of size $N-1$.

(d) **Conditioning on $w = 3$ spins.** Without loss of generality, we assume that $u, h$ are connected by an edge $e$ in $G$. A derivation similar to (15) and (16) reveals that (preserving the notation of case 2):

$\left| Z_{x_u = s^{(1)}, x_h = s^{(2)}, x_q = s^{(3)}} = \exp(J_{e} s^{(1)} s^{(2)}) \cdot Z'_{x'_z = 1, x'_w = s^{(3)}} \right|_{x'_z = 1, x'_w = s^{(3)}}$

which reduces the inference conditional on three vertices to a simpler case of two vertices. Again, sampling from $\mathbb{P}(X | x_u = s^{(1)}, x_j = s^{(2)}, x_q = s^{(3)})$ is reduced to a more basic sampling from $\mathbb{P}'(X' | x'_z = 1, x'_w = s^{(3)})$.

In principle, lemma 6 can be extended to arbitrarily large $\omega$ leaving a certain freedom for the Ising model conditioning framework. However, in this manuscript we focus on a given special case, which is enough for our goals.

https://doi.org/10.1088/1742-5468/abcaf1
Appendix B. Theorem 3 proof

B.1. Counting PMs of planar \( \hat{G} \) in \( O(\hat{N}^2) \) time

This section addresses the inference part of theorem 3.

B.1.1. Pfaffian orientation. We consider an orientation on \( \hat{G} \). \( \hat{G} \)'s cycle of even length (built on an even number of vertices) is said to be odd-oriented if, when all edges along the cycle are traversed in any direction, an odd number of edges are directed along the traversal. For \( X \subseteq V(\hat{G}) \) let \( \hat{G}(X) \) denote a graph \((X, \{e \in E(\hat{G}) | e \subseteq X\})\). An orientation of \( \hat{G} \) is called Pfaffian if all cycles \( C \), such that \( \text{PM}(\hat{G}(V(\hat{G}) - C)) \neq \emptyset \), are odd-oriented.

We will need \( \hat{G} \) to contain a Pfaffian orientation; moreover, the construction is easy.

Theorem 13. Pfaffian orientation of \( \hat{G} \) can be constructed in \( O(\hat{N}) \).

Proof. This theorem is proven constructively; see e.g. Wilson (1997), Vazirani (1989), or Schraudolph and Kamenetsky (2009), where the latter construction is based on specifics of the expanded dual graph. □

We construct a skew-symmetric sparse matrix \( \mathcal{K} \in \mathbb{R}^{\hat{N} \times \hat{N}} \) (\( \to \) denotes the orientation of edges):

\[
\mathcal{K}_{ij} = \begin{cases} 
  c_e & \text{if } \{v_i, v_j\} \in E(\hat{G}), v_i \to v_j \\
  -c_e & \text{if } \{v_i, v_j\} \in E(\hat{G}), v_j \to v_i \\
  0 & \text{if } \{v_i, v_j\} \notin E(\hat{G}).
\end{cases} 
\] (18)

The next result allows us to compute \( PF \hat{Z} \) of the PM model on \( \hat{G} \) in a polynomial time.

Theorem 14. \( \det \mathcal{K} > 0, \hat{Z} = \sqrt{\det \mathcal{K}}. \)

Proof. See e.g. Wilson (1997) or Kasteleyn (1963). □

B.1.2. Computing \( \det \mathcal{K} \). \( LU \)-decomposition of a matrix \( A = LU \), found via Gaussian elimination, where \( L \) is a lower-triangular matrix with unit diagonals and \( U \) is an upper-triangular matrix, would be a standard way of computing \( \det A \), which is then equal to a product of the diagonal elements of \( U \). However, this standard way of constructing the \( LU \) decomposition applies only if all \( A \)'s leading principal submatrices are nonsingular (see e.g. Horn and Johnson 2012, section 3.5, for detailed discussions). The \( 1 \times 1 \) leading principal submatrix of \( \mathcal{K} \) is already zero/singular.

Luckily, this difficulty can be resolved through the following construction. Take \( \hat{G} \)'s arbitrary perfect matching \( E' \in \text{PM}(\hat{G}) \). In the case of a general planar graph, \( E' \) can be found, e.g. via Blum’s algorithm (Blum 1990) in \( O(\sqrt{\hat{N}|E(\hat{G})|}) = O(\hat{N}^{3/2}) \) time, while for graph \( G^* \) appearing in this paper, \( E' \) can be found in \( O(\hat{N}) \) from a spin configuration using \( M \) mapping (e.g. \( E' = E'_f = M(\{+1, \ldots, +1\}) \in \text{PM}(G^*) \)). We modify the ordering of vertices, \( V(\hat{G}) = \{v_1, v_2, \ldots, v_{\hat{N}}\} \), such that \( E' = \{v_1, v_2, \ldots, \{v_{\hat{N}-1}, v_{\hat{N}}\}\}. \) We build \( \mathcal{K} \) according to definition (18). We obtain \( \overline{\mathcal{K}} \) from \( \mathcal{K} \) by swapping column
1 with column 2, 3 with 4, etc. This results in \( \det \mathcal{K} = |\det \mathcal{K}| \), where the new \( \mathcal{K} \) is properly conditioned.

**Lemma 15.** \( \mathcal{K} \)'s leading principal submatrices are nonsingular.

**Proof.** The proof, presented in Wilson (1997) for the case of unit weights \( c_e \), generalizes to arbitrary positive \( c_e \).

Note that in the general case (of a matrix represented in terms of a general graph) the complexity of the \( LU \)-decomposition is cubic in the size of the matrix. Fortunately, the nested dissection technique, discussed in the following subsection, allows us to reduce the complexity of computing \( \hat{Z} \) to \( O\left(\hat{N}^{\frac{3}{2}}\right) \).

**B.1.3. Nested dissection.** The partition \( P_1, P_2, P_3 \) of set \( V(\hat{G}) \) is a separation of \( \hat{G} \) if for any \( v \in P_1, w \in P_2 \) it holds that \( \{v, w\} \notin E(\hat{G}) \). We refer to \( P_1, P_2 \) as the parts, and to \( P_3 \) as the separator.

Lipton and Tarjan (LT) (Lipton and Tarjan 1979) found an \( O(\hat{N}) \) algorithm, which finds a separation \( P_1, P_2, P_3 \) such that \( \max(|P_1|, |P_2|) \leq \frac{3}{4} \hat{N} \) and \( |P_3| \leq 2\sqrt{N} \). The LT algorithm can be used to construct the so-called nested dissection ordering of \( V(\hat{G}) \). The ordering is built recursively, by first placing vertices of \( P_1 \), then \( P_2 \) and \( P_3 \), and finally permuting indices of \( P_1 \) and \( P_2 \) recursively according to the ordering of \( \hat{G}(P_1) \) and \( \hat{G}(P_2) \) (see Lipton et al (1979) for accurate description of details, definitions and analysis of the nested dissection ordering). As shown by Lipton et al (1979), the complexity of finding the nested dissection ordering is \( O(\hat{N} \log \hat{N}) \).

Let \( A \) be an \( \hat{N} \times \hat{N} \) matrix with a sparsity pattern of \( \hat{G} \). That is, \( A_{ij} \) can be nonzero only if \( i = j \) or \( \{v_i, v_j\} \in \hat{E} \).

**Theorem 16 (Lipton et al 1979).** If \( \hat{V} \) is ordered according to the nested dissection and \( A \)'s leading principal submatrices are nonsingular, computing the \( LU \)-decomposition of \( A \) becomes a problem of \( O\left(\hat{N}^{\frac{3}{2}}\right) \) complexity.

Note, however, that we cannot directly apply the theorem to \( \mathcal{K} \), because the sparsity pattern of \( \mathcal{K} \) is asymmetric and does not correspond, in general, to any graph.

Let \( G^{**} \) be a planar graph, obtained from \( \hat{G} \), by contracting each edge in \( E' \), \( |V(G^{**})| = |E'| = \frac{1}{2} \hat{N} \). We find and fix a nested dissection ordering over \( V(G^{**}) \) (it takes \( O(\hat{N} \log \hat{N}) \) steps) and let the \( \{v_1, v_2\}, \ldots, \{v_{\hat{N}-1}, v_{\hat{N}}\} \) enumeration of \( E' \) correspond to this ordering. We split \( K \) into \( 2 \times 2 \) cells and consider the sparsity pattern of the nonzero cells. One observes that the resulting sparsity pattern coincides with the sparsity patterns of \( \mathcal{K} \) and \( G^{**} \). Since \( LU \)-decomposition can be stated in the \( 2 \times 2 \) block elimination form, its complexity is reduced to \( O\left(\hat{N}^{\frac{3}{2}}\right) \).

This concludes the construction of an efficient inference (counting) algorithm for the planar PM model.
B.2. Sampling PMs of planar $\hat{G}$ in $O\left( N^{\frac{3}{2}} \right)$ time (Wilson’s algorithm)

This section addresses the sampling part of theorem 3. In this section we assume that
the degrees of $\hat{G}$’s vertices are upper-bounded by 3. This is true for $G^*$—the only type
of PM model appearing in the paper. Any other constant substituting 3 would not
affect the analysis of complexity. Moreover, Wilson (1997) shows that any PM model
on a planar graph can be reduced to a bounded-degree planar model without affecting
$O(\hat{N}^{\frac{3}{2}})$ complexity.

B.2.1. Structure of the algorithm. We denote a sampled PM as $M$, $P(M) = \hat{Z}^{-1} \prod_{e \in M} c_e$. Wilson’s algorithm first applies the LT algorithm of Lipton and Tarjan
(1979) to find a separation $P_1, P_2, P_3$ of $\hat{G}$ ($\max(|P_1|, |P_2|) \leq \frac{3}{2} \hat{N}$, $|P_3| \leq 2^{\frac{3}{2}} \sqrt{\hat{N}}$). Then, it iterates over $v \in P_3$ and for each $v$ it draws an edge of $M$, saturating $v$. It then appears
that, given this intermediate result, drawing the remaining edges of $M$ may be split into
two independent drawings over $\hat{G}(P_1)$ and $\hat{G}(P_2)$, respectively, and then the process is
repeated recursively.

It takes $O(\hat{N}^{\frac{3}{2}})$ steps to sample the edges attached to $P_3$ at the first step of the
recursion; therefore, the overall complexity of Wilson’s algorithm is also $O(\hat{N}^{\frac{3}{2}})$.

Subsection B.2.2 introduces the probabilities required to draw the aforementioned
PM samples. Subsections B.2.3 and B.2.4 describe how to sample edges attached to the
separator, while subsection B.3 focuses on describing the recursion.

B.2.2. Drawing perfect matchings. For some $Q \in E(\hat{G})$ we consider the probability of
obtaining $Q$ as a subset of $M$:

$$P(Q \subseteq M) = \frac{1}{\hat{Z}} \sum_{M' \in \text{PM}(\hat{G})} \left( \prod_{e \in M'} c_e \right)$$

$$= \frac{1}{\hat{Z}} \left( \prod_{e \in Q} c_e \right) \cdot \sum_{M' \in \text{PM}(\hat{G})} \left( \prod_{e \in M' \setminus Q} c_e \right).$$  \hspace{1cm} (19)

Let $V_Q = \cup_{e \in Q} e$ and $\hat{G}_{\setminus Q} = \hat{G}(V(\hat{G}) \setminus V_Q)$. Then, the set $\{M' \setminus Q \mid M' \in \text{PM}(\hat{G})\}$
coincides with PM($\hat{G}_{\setminus Q}$). This yields the following expression:

$$P(Q \subseteq M) = \frac{\hat{Z}_{\setminus Q}}{\hat{Z}} \left( \prod_{e \in Q} c_e \right)$$  \hspace{1cm} (20)

where

$$\hat{Z}_{\setminus Q} = \sum_{M'' \in \text{PM}(\hat{G}_{\setminus Q})} \left( \prod_{e \in M''} c_e \right)$$  \hspace{1cm} (21)

is a PF of the PM model on $\hat{G}_{\setminus Q}$ induced by the edge weights $c_e$. 

https://doi.org/10.1088/1742-5468/abcaf1
For a square matrix $A$ let $A_{r_1, \ldots, r_l}^{c_1, \ldots, c_l}$ denote the matrix obtained by deleting rows $r_1, \ldots, r_l$ and columns $c_1, \ldots, c_l$ from $A$. Let $[A]_{c_1, \ldots, c_l}^{r_1, \ldots, r_l}$ be obtained by leaving only rows $r_1, \ldots, r_l$ and columns $c_1, \ldots, c_l$ of $A$ and placing them in this order.

Now let $V_Q = \{v_i, \ldots, v_j\}$, $i_1 < \cdots < i_r$. A simple check demonstrates that deleting a vertex from a graph preserves the Pfaffian orientation. By induction this holds for any number of vertices deleted. From that it follows that $\mathcal{K}_{i_1, \ldots, i_r}$ is a Kasteleyn matrix for $\hat{G}_Q$ and then

$$Z_{\hat{G}} = \text{pf}(\mathcal{K}_{i_1, \ldots, i_r}) = \sqrt{\det K_{i_1, \ldots, i_r}}$$

resulting in

$$\mathbb{P}(Q \subseteq M) = \sqrt{\frac{\det K_{i_1, \ldots, i_r}}{\det K}} \left(\prod_{e \in Q} e_c\right).$$

Linear algebra transformations described by Wilson (1997) suggest that if $A$ is non-singular, then

$$\frac{\det A_{i_1, \ldots, i_r}}{\det A} = \pm \det [A^{-1}]_{i_1, \ldots, i_r}. \quad (24)$$

This observation allows us to express probability (19) as

$$\mathbb{P}(Q \subseteq M) = \sqrt{\frac{\det \mathcal{K}_{i_1, \ldots, i_r}}{\det \mathcal{K}}} \left(\prod_{e \in Q} e_c\right). \quad (25)$$

Now we are in the position to describe the first step of Wilson’s recursion.

B.2.3. Step 1: computing lower-right submatrix of $\overline{\mathcal{K}}^{-1}$. We find a separation $P_1, P_2, P_3$ of $\hat{G}$. The goal is to sample an edge from every $v \in P_3$.

Let $W$ be a set of vertices from $P_3$ and their neighbors; then, $|W| \leq 3|P_3|$ because each vertex in $\hat{G}$ is of degree 3 at most. Let $W^* \subseteq V(G^*)$ be a set of the contracted edges (recall $G^*$ definition from subsection B.1.3), containing at least one vertex from $W$, $|W^*| \leq |W|$. Then, $W^*$ is a separator of $G^*$ such that

$$|W^*| \leq |W| \leq 3|P_3| \leq 3 \cdot 2^3 \sqrt{\hat{N}} \leq 3 \cdot 2^2 \sqrt{|V(G^*)|} \quad (26)$$

where one uses the fact that $|V(G^*)| = \frac{\hat{N}}{2}$. We find a nested dissection ordering (subsection B.1.3) of $V(G^*)$ with $W^*$ as a top-level separator. This is a correct nested dissection due to equation (26).

Utilizing this ordering, we construct $\overline{\mathcal{K}}$. We compute $L$ and $U$—$LU$-decomposition of $\overline{\mathcal{K}}$ ($O(\hat{N}^2)$ time). Let $\gamma = 2|W^*| \leq 3 \cdot 2^3 \sqrt{\hat{N}}$ and let $\mathcal{I}$ be a shorthand notation for $(\hat{N} - \gamma + 1, \ldots, \hat{N})$. Using $L$ and $U$, we find $D = [\overline{\mathcal{K}}^{-1}]_{\mathcal{I}^2}$, which is a lower-right $\overline{\mathcal{K}}^{-1}$ submatrix of size $\gamma \times \gamma$.

It is straightforward to observe that the $i$th column of $D$, $d_i$, satisfies

$$[L]_{\mathcal{I}^2} \times ([U]_{\mathcal{I}^2} \times d_i) = e_i, \quad (27)$$

https://doi.org/10.1088/1742-5468/abcaf1
where $e_i$ is a zero vector with unity at the $i$th position. Therefore, constructing $D$ is reduced to solving $2\gamma$ triangular systems, each of size $\gamma \times \gamma$, resulting in $O(\gamma^3) = O(\hat{N}^2)$ required steps.

B.2.4. Step 2: sampling edges in the separator. Now, progressing iteratively, one can find $v \in P_3$ which is not yet paired and draw an edge emanating from it. Suppose that the edges, $e_1 = \{v_{j_1}, v_{j_2}\}$, $\ldots$, $e_k = \{v_{j_{2k-1}}, v_{j_{2k}}\}$, are already sampled. We assume that by this point we have also computed $LU$-decomposition $A_k = [\mathcal{K}^{-1}]_{j_1,\ldots,j_{2k}} = L_k U_k$ and we will update it to $A_{k+1}$ when the new edge is drawn. Then,

$$\text{P}(e_1, \ldots, e_k \in M) = \sqrt{|\det A_k|} \prod_{j=1}^{k} c_{e_j}.$$  \hfill (28)

Next we choose $j_{2k+1}$ so that $v_{j_{2k+1}}$ is not saturated yet. We iterate over $v_{j_{2k+1}}$’s neighbors considered as candidates for becoming $v_{j_{2k+2}}$. Letting $v_j$ become the next candidate, we denote $e_{k+1} = \{v_{j_{2k+1}}, v_j\}$. For $n \in \mathbb{N}$ let $\alpha(n) = n + 1$ if $n$ is odd and $\alpha(n) = n - 1$ if $n$ is even. Then, the identity

$$\mathcal{K}^{-1} = [\mathcal{K}^{-1}]_{1,\ldots,\hat{N}}$$  \hfill (29)

follows from the definition of $\mathcal{K}$. One deduces from equation (29) that

$$A_{k+1} = [\mathcal{K}^{-1}]_{j_1,\ldots,j_{2k+1}j} = [\mathcal{K}^{-1}]_{j_1,\ldots,j_{2k+1}j}^{\alpha(j_1),\ldots,\alpha(j_{2k+1})\alpha(j)}.$$  \hfill (30)

Constructing $W^{**}$, one has $j_1, j_2, j_3, j_4, \alpha(j_1), \ldots, \alpha(j_{2k+1}), \alpha(j) > \hat{N} - t$. This means that $A_{k+1}$ is a submatrix of $D$ with permuted rows and columns; hence, $A_{k+1}$ is known.

We further observe that

$$A_{k+1} = \begin{bmatrix} A_k & y \\ r & d \end{bmatrix} = \begin{bmatrix} L_k & 0 \\ R & 1 \end{bmatrix} \begin{bmatrix} U_k & Y \\ 0 & z \end{bmatrix} = L_{k+1} U_{k+1}.$$  \hfill (31)

Therefore, to update $L_{k+1}$ and $U_{k+1}$, one just solves the triangular system of equations $R U_k = r$ and $L_k Y = y$, where $R^T$, $r^T$, $Y$, $y$ are of size $2k \times 2$ (this is done in $O(k^2)$ steps), and then computes $z = d - R Y$, which is of the size $2 \times 2$, then sets $u = \det z$.

The probability to pair $v_{j_{2k+1}}$ and $v_j$ is

$$\text{P}(e_{k+1} \in M | e_1, \ldots, e_k \in M) = \frac{\text{P}(e_1, \ldots, e_{k+1} \in M)}{\text{P}(e_1, \ldots, e_k \in M)} = \frac{\sqrt{|\det A_{k+1}|} \prod_{j=1}^{k+1} c_{e_j}}{\sqrt{|\det A_k|} \prod_{j=1}^{k} c_{e_j}} = c_{e_{k+1}} \frac{|u| |\det A_k|}{\sqrt{|\det A_k|}} = c_{e_{k+1}} \sqrt{|u|}.\quad (32)$$
Therefore, maintaining \( U_{k+1} \) allows us to compute the required probability and draw a new edge from \( v_{2k+1} \). By construction of \( \hat{G} \), \( v_{2k+1} \) has only three neighbors; therefore, the complexity of this step is \( O \left( \sum_{k=1}^{\left| P_{3} \right|} k^{2} \right) = O \left( \hat{N}^{2} \right) \) because \( |P_{3}| \leq 2^{\hat{N}} \sqrt{\hat{N}} \).

### B.3. Step 3: recursion

Let \( M_{\text{sep}} = \{e_{1}, e_{2}, \ldots\} \) be a set of edges drawn on the previous step, and \( \hat{V}_{\text{sep}} \) be a set of vertices saturated by \( M_{\text{sep}} \). \( P_{3} \subseteq \hat{V}_{\text{sep}} \). Given \( M_{\text{sep}} \), the task of sampling \( M \in \text{PM}(\hat{G}) \) such that \( M_{\text{sep}} \subseteq M \) is reduced to sampling perfect matchings \( M_{1} \) and \( M_{2} \) over \( \hat{G}(P_{1} \setminus \hat{V}_{\text{sep}}) \) and \( \hat{G}(P_{2} \setminus \hat{V}_{\text{sep}}) \), respectively. Then, \( M = M_{1} \cup M_{2} \cup M_{\text{sep}} \) becomes the result of the perfect matching drawn from (4).

Even though only the first step of Wilson’s recursion has been discussed so far, any further step in the recursion is done in exactly the same way with the only exception that vertex degrees may become less than 3, while in \( \hat{G} \) they are exactly 3. Obviously, this does not change the iterative procedure and it also does not affect the complexity analysis.

### Appendix C. Theorem 7 proof

Prior to the proof we introduce a series of definitions and results. We follow Hopcroft and Tarjan (1973), Gutwenger and Mutzel (2001) (see also Mader (2008)) to define the tree of triconnected components. The definitions apply for a biconnected graph \( G \) (see the definitions of a biconnected graph and biconnected component, e.g. in appendix D).

Let \( v, w \in V(G) \). We divide \( E(G) \) into equivalence classes \( E_{1}, \ldots, E_{k} \) so that \( e_{1}, e_{2} \) are in the same class if they lie on a common simple path that has \( v, w \) as endpoints. \( E_{1}, \ldots, E_{k} \) are referred to as separation classes. If \( k \geq 2 \), then \( \{v, w\} \) is a separation pair of \( G \), unless (a) \( k = 2 \) and one of the classes is a single edge or (b) \( k = 3 \) and each class is a single edge. Graph \( G \) is called triconnected if it has no separation pairs.

Let \( \{v, w\} \) be a separation pair in \( G \) with equivalence classes \( E_{1}, \ldots, E_{k} \). Let \( E' = \bigcup_{i=1}^{l} E_{i}, \quad E'' = \bigcup_{i=l+1}^{k} E_{i} \) be such that \( |E'| \geq 2, \quad |E''| \geq 2 \). Then, graphs \( G_{1} = (\bigcup_{e \in E'} e, E' \cup \{e_{v}\}), \quad G_{2} = (\bigcup_{e \in E''} e, E'' \cup \{e_{v}\}) \) are called split graphs of \( G \) with respect to \( \{v, w\} \), and \( e_{v} \) is a virtual edge, which is a new edge between \( v \) and \( w \), identifying the split operation. Due to the addition of \( e_{v} \), \( G_{1} \) and \( G_{2} \) are not normal in general.

Split \( G \) into \( G_{1} \) and \( G_{2} \). Continue splitting \( G_{1}, G_{2} \), and so on, recursively, until no further split operation is possible. The resulting graphs are split components of \( G \). They can either be \( K_{3} \) (triangles), triple bonds, or triconnected normal graphs.

Let \( e_{v} \) be a virtual edge. There are exactly two split components containing \( e_{v} \): \( G_{1} = (V_{1}, E_{1}) \) and \( G_{2} = (V_{2}, E_{2}) \). Replacing \( G_{1} \) and \( G_{2} \) with \( G' = (V_{1} \cup V_{2}, (E_{1} \cup E_{2}) \setminus \{e_{v}\}) \) is called merging \( G_{1} \) and \( G_{2} \). We perform all possible mergings of the cycle graphs (starting from triangles), and then all possible mergings of multiple bonds starting from triple bonds. The components of the resulting set are referred to as the triconnected components of \( G \). We emphasize again that some graphs (i.e. cycles and bonds) in the set of triconnected components are not necessarily triconnected.

https://doi.org/10.1088/1742-5468/abcaf1
Figure C1. (I) Example biconnected graph $G$. (II) Separation pair $\{a, b\}$ of $G$ and separation classes $E_1, E_2, E_3$ associated with $\{a, b\}$. (III) Result of split operation with $E' = E_1 \cup E_2, E'' = E_3$. Dashed lines indicate virtual edges and dotted lines connect equivalent virtual edges in split graphs. (IV) Split components of $G$ (non-unique). (V) Triconnected components of $G$. (VI) Triconnected component tree $T$ of $G$; spatial alignment of $V$ is preserved. ‘G’, ‘B’, and ‘C’ are examples of the ‘triconnected graph’, ‘multiple bond’, and ‘cycle’, respectively.

Lemma 17 (Hopcroft and Tarjan 1973). Triconnected components are unique for $G$. The total number of edges within the triconnected components is at most $3|E| - 6$.

Consider a graph $T'$, where vertices (further referred to as nodes for disambiguation) are triconnected components, and there is an edge between $a$ and $b$ in $T'$, when $a$ and $b$ share a (copied) virtual edge.

Lemma 18 (Hopcroft and Tarjan 1973). $T'$ is a tree.

We will also use the following celebrated result:

Lemma 19 (Hall 1943). Biconnected graph $G$ is $K_{33}$-free only if its nonplanar triconnected components are exactly $K_5$.

The graph in figure C1 is actually $K_{33}$-free according to the lemma. Now we are in the position to give a proof of theorem 7.

Proof. Since $G$ is $K_{33}$-free and has no loops or multiple edges, it holds that $|E(G)| = O(N)$ (Thomason 2001). In time $O(N)$ we can find a forest of $G$’s biconnected components (Tarjan 1971). If we find the five-nice decomposition of each biconnected component, we can trivially combine them into a single five-nice decomposition in time $O(N)$ using navels of size 0 and 1. Hence, we can assume that $G$ is biconnected.

We build a tree of triconnected components for $G$ in time $O(N)$ (Hopcroft and Tarjan 1973, Gutwenger and Mutzel 2001, Vo 1983). Now we delete the virtual edges, which results in a five-nice decomposition of $G$, given the lemma 19. □

https://doi.org/10.1088/1742-5468/abcaf1
Appendix D. Proof for theorem 10

Prior to the proof, we introduce a series of definitions used by Reed and Li (2008). It is assumed that a graph $G = (V, E)$ (no loops and multiple edges) is given.

For any $X \subseteq V(G)$ let $G - X$ denote a graph $(V(G) \setminus X, \{e = \{v, w\} \in E(G) \mid v, w \notin X\})$. $X \subseteq V(G)$ is an $(i, j)$-cut whenever $|X| = i$ and $G - X$ has at least $j$ connected components.

The graph is biconnected whenever it has no $(1, 2)$-cut. A biconnected component of the graph is a maximal biconnected subgraph. Clearly, a pair of biconnected components can intersect in at most one vertex and a graph of components’ intersections is a tree whenever $G$ is connected (a tree of biconnected components). The graph is three-connected whenever it has no $(2, 2)$-cut.

A two-block tree of a biconnected graph $G$, written $\langle T', G' \rangle$, is a tree $T'$ with a set $G' = \{G'_t\}_{t \in V(T')}$ with the following properties:

- $G'_t$ is a graph (possibly with multiple edges) for each $t \in V(T')$.
- If $G$ is three-connected then $T'$ has a single node $r$ which is colored 1 and $G'_r = G$.
- If $G$ is not three-connected then there exists a color 2 node $t \in V(T')$ such that
  * $G'_t$ is a graph with two vertices $u$ and $v$ and no edges for some $(2, 2)$-cut $\{u, v\}$ in $G$.
  * Let $T'_1, \ldots, T'_k$ be the connected components (subtrees) of $T' - t$. Then, $G - \{u, v\}$ has $k$ connected components $U_1, \ldots, U_k$ and there is a labeling of these components such that $T'_i$ is a two-block tree of $G'_i = (V(U_i) \cup \{u, v\}, E(U_i) \cup \{\{u, v\}\}$.
  * For each $i$, there exists exactly one color 1 node $t_i \in V(T'_i)$ such that $\{u, v\} \subseteq V(G'_i)$.
  * For each $i$, $\{t, t_i\} \in E(T)$.

A $3$-block tree of a three-connected graph $G$, written $\langle T'', G'' \rangle$, is a tree $T''$ with a set $G'' = \{G''_t\}_{t \in V(T'')}$ with the following properties:

- $G''_t$ is a graph (possibly with multiple edges) for each $t \in V(T'')$.
- If $G$ has no $(3, 3)$-cut then $T$ has a single node $r$ which is colored 1 and $G''_r = G$.
- If $G$ has a $(3, 3)$-cut then there exists a color 2 node $t \in V(T'')$ such that
  * $G''_t$ is a graph with vertices $u$, $v$, and $w$ and no edges for some $(3, 3)$-cut $\{u, v, w\}$ in $G$.
  * Let $T''_1, \ldots, T''_k$ be the connected components (subtrees) of $T'' - t$. Then $G - \{u, v, w\}$ has $k$ connected components $U_1, \ldots, U_k$ and there is a labeling of these components such that $T''_i$ is a $(3, 3)$-block tree of $G''_i = (V(U_i) \cup \{u, v, w\}, E(U_i) \cup \{\{u, v\}, \{v, w\}, \{u, w\}\}$).
  * For each $i$, there exists exactly one color 1 node $t_i \in V(T''_i)$, such that $\{u, v, w\} \subseteq V(G''_i)$.
  * For each $i$, $\{t, t_i\} \in E(T'')$. 

https://doi.org/10.1088/1742-5468/abcaf1
**Proof.** Since $G$ is $K_5$-free and has no loops or multiple edges, it holds that $|E(G)| = O(N)$ (Thomason 2001). In time $O(N)$ we can find a forest of $G$’s biconnected components (Tarjan 1971). If we find an eight-nice decomposition for each biconnected component, we join them into a single eight-nice decomposition using attachment sets of size 1 for decompositions inside $G$’s connected component and attachment sets of size 0 for decompositions in different connected components. Hence, we further assume that $G$ is biconnected.

The $O(N)$ algorithm of Reed and Li (2008) finds a two-block tree $\langle T', G' \rangle$ for $G$ and then for each color 1 node $G'_t \in G'$ it finds a $(3,3)$-block tree $\langle T'', G'' \rangle$ where all components are either planar or Möbius ladders. To obtain an eight-nice decomposition from each $(3,3)$-block tree, (1) for each color 2 node we contract an edge between it and one of its neighbors in $T''$, and (2) we remove all edges that were only created during $\langle T'', G'' \rangle$ construction (second item of the $(3,3)$-block tree definition).

Now we have to draw additional edges in the forest $F$ of obtained eight-nice decompositions to obtain a single eight-nice decomposition $T$ of $G$. Note that for each pair of adjacent nodes $G'_t, G'_s \in G'$ where $G'_t$ is color 1 node and $G'_s = (\{u, v\}, \emptyset)$ is a color 2 node, $u, v$ are in $V(G'_t)$ and $\{u, v\} \in E(G'_t)$. Hence, there is at least one component $G''_r$ of eight-nice decomposition of $G'_t$ where both $u$ and $v$ are present. For each pair of $s$ and $t$ we draw an edge between $s$ and $r$ in $F$. Then, (1) for each color 2 node in $F$ (such as $s$) we contract an edge between it and one of its neighbors (such as $r$), and (2) we remove all edges that were created during $\langle T', G' \rangle$ construction (second item of two-block tree definition). This results is a correct $c$-nice decomposition for biconnected $G$. □

**Appendix E. Random graph generation**

As our derivations cover the most general case of planar and $K_{33}$-free graphs, we want to test them on graphs that are as general as possible. Based on lemma 19 (note that it provides necessary and sufficient conditions for a graph to be $K_{33}$-free) we implement a randomized construction of $K_{33}$-free graphs, which is assumed to cover most general $K_{33}$-free topologies.

Namely, one generates a set of $K_5$’s and random planar graphs, attaching them by edges to a tree-like structure. Our generation process consists of the following two steps.

(a) **Planar graph generation.** This step accepts $N \geq 3$ as an input and generates a normal biconnected planar graph of size $N$ along with its embedding on a plane.

The details of the construction are as follows.

First, a random embedded tree is drawn iteratively. We start with a single vertex, on each iteration choose a random vertex of an already ‘grown’ tree, and add a new vertex connected only to the chosen vertex. Items (I)–(V) in figure E1 illustrate this step.

Then, we triangulate this tree by adding edges until the graph becomes biconnected and all faces are triangles, as in subsection 2.1 ((VI) in figure E1). Next, to obtain a normal graph, we remove multiple edges possibly produced by triangulation ((VII) in figure E1). At this point the generation process is complete.
(b) **$K_{33}$-free graph generation.** Here we take $N \geq 5$ as the input and generate a normal biconnected $K_{33}$-free graph $G$ in the form of its partially merged decomposition $T'$. Namely, we generate a tree $T$ of graphs in which each node is either a normal biconnected planar graph or $K_5$, and every two adjacent graphs share a virtual edge.

The construction is greedy and is essentially the tree generation process from step 1. We start with the $K_5$ root and then iteratively create and attach new nodes. Let $N' < N$ be the size of the already generated graph, $N' = 5$ at first. Note, that when a node of size $n$ is generated, it contributes $n - 2$ new vertices to $G$.

An elementary step of iteration here is as follows. If $N - N' \geq 3$, a coin is flipped and the type of the new node is chosen—$K_5$ or planar. If $N - N' < 3$, $K_5$ cannot be added, so planar type is chosen. If a planar node is added, its size is drawn uniformly in the range between 3 and $N - N' + 2$ and then the graph itself is drawn as described in step 1. Then, we attach a new node to a randomly chosen free edge of a randomly chosen node of $T'$. We repeat this process until $G$ is of the desired size $N$. Figure E2 illustrates the algorithm.

To obtain an Ising model from $G$, we sample pairwise interactions for each edge of $G$ independently from $\mathcal{N}(0,0.1^2)$. 

https://doi.org/10.1088/1742-5468/abcaf1
Note that the tractable Ising model generation procedure is designed in this section solely for the convenience of testing and it is not claimed to be sampling models of any particular practical interest (e.g. in statistical physics or computer science).

Appendix F. Upper bound minimization and marginal computation in approximation scheme

We denote:

\[
    h(J') \triangleq \min_{\rho(r) \geq 0, \sum_r \rho(r) = 1} g(J', \rho),
\]

\[
    g(J', \rho) \triangleq \min_{\{J^{(r)}\}} \sum_r \rho(r) J^{(r)} \sum_r \rho(r) \log Z(G^{(r)}, 0, J^{(r)})
\]

where \(h(J')\) is a tight upper bound for \(\log Z(G', 0, J')\).

Given a fixed \(\rho\), we compute \(g(J', \rho)\) using L-BFGS-B optimization (Zhu et al 1997) by back-propagating through \(Z(G^{(r)}, 0, J^{(r)})\) and projecting gradients on the constraint linear manifold. On the upper level we also apply the L-BFGS-B algorithm to compute \(h(J')\), which is possible since (Wainwright et al 2005, Globerson and Jaakkola 2007)

\[
    \frac{\partial}{\partial \rho(r)} g(J', \rho) = \log Z(G^{(r)}, 0, J_{\min}^{(r)}) - (M^{(r)})^T J_{\min}^{(r)}, M^{(r)} \triangleq \frac{\partial}{\partial J_{\min}^{(r)}} \log Z(G^{(r)}, 0, J_{\min}^{(r)})
\]

where \(\{J_{\min}^{(r)}\}\) is argmin inside \(g(J', \rho)\)'s definition and \(M^{(r)} = \{M_e^{(r)} | e \in E(G^{(r)})\}\) is a vector of pairwise marginal expectations. We reparameterize \(\rho(r)\) into \(\sum_{r_w'(r)} w'(r)\) where \(w'(r) > 0\).

For \(e = \{v, w\} \in E(G)\) we approximate pairwise marginal probabilities as (Wainwright et al 2005, Globerson and Jaakkola 2007)

\[
    \mathbb{P}_{\text{alg}}(x_v x_w = 1) = \frac{1}{2} \cdot \left[ \sum_r \rho(r) M_e^{(r)} \right] + \frac{1}{2}.
\]

Let \(e_A\) be an edge between the central vertex \(v\) and apex in \(G'\). We approximate the singleton marginal probability at vertex \(v\) as

\[
    \mathbb{P}_{\text{alg}}(x_v = 1) = \frac{1}{2} \cdot \left[ \sum_r \rho(r) M_{e_A}^{(r)} \right] + \frac{1}{2}.
\]

References

Barahona F 1982 On the computational complexity of Ising spin glass models J. Phys. A: Math. Gen. 15 3241
Battle J, Frank H and Kodama Y 1962 Additivity of the genus of a graph Bull. Am. Math. Soc. 68 565–8
Bellman R 1952 On the theory of dynamic programming Proc. Natl Acad. Sci. 38 716–9
Bertsekas D P 1999 Nonlinear Programming (Nashua, NH: Athena Scientific)
Bethe H A 1935 Statistical theory of superlattices Proc. R. Soc. A 150 552

https://doi.org/10.1088/1742-5468/abcaf1

31
Tractable minor-free generalization of planar zero-field Ising models

Bieche L, Uhry J P, Maynard R and Rammal R 1980 On the ground states of the frustration model of a spin glass by a matching method of graph theory J. Phys. A: Math. Gen. 13 2553

Blum N 1990 A new approach to maximum matching in general graphs Automata, Languages and Programming ed M S Paterson (Berlin: Springer) pp 586–97

Bodlaender H L 1998 A partial k-arborescent graphs with bounded treewidth Theor. Comput. Sci. 209 1–45

Boyer J M and Myrvold W J 2004 On the cutting edge: simplified O(n) planarity by edge addition J. Graph Algorithms Appl. 8 241–73

Curticapean R 2014 Counting perfect matchings in graphs that exclude a single-crossing minor (arXiv:1406.4056)

Diestel R 2006 Graph Theory (Electronic Library of Mathematics) (Berlin: Springer)

Fisher M E 1966 On the dimer solution of planar Ising models J. Math. Phys. 7 1776–81

Gallager R G 1963 Low Density Parity Check Codes (Cambridge, MA: MIT Press)

Galluccio A and Loebi M 1999 On the theory of Pfaffian orientations: I. Perfect matchings and permanents Electron. J. Comb. 6 R6

Globerson A and Jaakkola T S 2007 Approximate inference using planar graph decomposition Adv. Neural Inf. Process. Syst. 473–80

Gutwenger C and Mutzel P 2001 A linear time implementation of SPQR-trees Graph Drawing edn J Marks (Berlin: Springer) pp 77–90

Hall D W 1943 A note on primitive skew curves Bull. Am. Math. Soc. 49 935–7

Hopcroft J E and Tarjan R E 1973 Dividing a graph into triconnected components J. Graph Theory 7 197–217

Horng R A and Johnson C R 2012 Matrix Analysis 2nd edn (Cambridge: Cambridge University Press)

Jean Dunm O 1961 Multiple comparisons among means J. Am. Stat. Assoc. 56 52–64

Jerrum M and Sinclair A 1993 Polynomial-time approximation algorithms for the Ising model Algorithmica (Springer) pp 77–90

Kac M and Ward J C 1952 A combinatorial solution of the two-dimensional Ising model Math. Proc. Camb. Phil. Soc. 48 287–93

Kasteleyn P W 1963 Dimer statistics and phase transitions J. Math. Phys. 4 287–93

Lipton R J and Tarjan R E 1979 A separator theorem for planar graphs SIAM J. Comput. 8 197–217

Lipton R J, Rose D J and Tarjan R E 1979 Generalized nested dissection SIAM J. Numer. Anal. 16 346–58

Mader M 2008 Planar graph drawing Master’s Thesis University of Konstanz, Konstanz

Mather M 1999 On the theory of Pfaffian orientations: I. Perfect matchings and permanents Electron. J. Comb. 6 R6

Pearl J 1982 Reverend bayes on inference engines: a distributed hierarchical approach Proc. 2nd AAAI Conf. on Artificial Intelligence, AAAI’82 (AAAI Press) pp 133–6

Peierls R 1936 On Ising’s model of ferromagnetism Math. Proc. Camb. Phil. Soc. 32 477–81

Reed B and Li Z 2008 Optimization and recognition for K5-minor free graphs in linear time LATIN 2008: Theoretical Informatics eds E S Laber, C Bornstein, L T Nogueira and L Faria (Berlin: Springer) pp 206–15

Schraudolph N N and Kamenetsky D 2009 Efficient exact inference in planar Ising models Adv. Neural Inf. Process. Syst. 21 1417–24

Straub S, Thierauf T and Wagner F 2014 Counting the number of perfect matchings in K5-free graphs 2014 IEEE 29th Conf. on Computational Complexity (CCC) pp 66–77

Tarjan R 1971 Depth-first search and linear graph algorithms 12th Annual Symp. on Switching and Automata Theory (SWAT 1971) pp 114–21

Thomas C K and Middleton A A 2009 Exact algorithm for sampling the two-dimensional Ising spin glass Phys. Rev. E 80 046708

Thomas C K and Middleton A A 2013 Numerically exact correlations and sampling in the two-dimensional Ising spin glass Phys. Rev. E 87 043303

Thomas A 2001 The extremal function for complete minors J. Comb. Theory B 81 318–38

Vazirani V K 1999 Finding triconnected components of graphs Linear Multilinear Algebra 13 143–65

Vo K-P 1983 Finding triconnected components of graphs Linear Multilinear Algebra 13 143–65

Wainwright M J, Jaakkola T S and Willsky A S 2005 A new class of upper bounds on the log partition function IEEE Trans. Inform. Theory 51 2313–35

Wilcoxon F 1945 Individual comparisons by ranking methods Biometrics Bull. 1 80–3

Zhu C, Byrd R H, Lu P and Nocedal J 1997 Algorithm 778: L-BFGS-B ACM Trans. Math. Softw. 23 505–60

https://doi.org/10.1088/1742-5468/abcaf1 32