
Compositional Probabilistic and Causal Inference using Tractable Circuit Models

Benjie Wang
 Department of Computer Science
 University of Oxford

Marta Kwiatkowska
 Department of Computer Science
 University of Oxford

Abstract

Probabilistic circuits (PCs) are a class of tractable probabilistic models, which admit efficient inference routines depending on their structural properties. In this paper, we introduce *md-vtrees*, a novel structural formulation of (marginal) determinism in structured decomposable PCs, which generalizes previously proposed classes such as probabilistic sentential decision diagrams. Crucially, we show how *md-vtrees* can be used to derive tractability conditions and efficient algorithms for advanced inference queries expressed as *arbitrary compositions* of basic probabilistic operations, such as marginalization, multiplication and reciprocals, in a sound and generalizable manner. In particular, we derive the first polytime algorithms for causal inference queries such as backdoor adjustment on PCs. As a practical instantiation of the framework, we propose MDNets, a novel PC architecture using *md-vtrees*, and empirically demonstrate their application to causal inference.

of PCs. Meanwhile, probabilistic circuit structures have recently been shown to scale to high-dimensional datasets such as CelebA (Peharz et al., 2020a).

A distinguishing feature of the PC framework is the ability to trade off *expressive efficiency* for *tractability* by imposing various properties on the PC. Broadly, these properties can be divided into *scope properties*, such as decomposability and structured decomposability (Pipatsrisawat and Darwiche, 2008), and *support properties*, such as determinism, strong determinism (Pipatsrisawat and Darwiche, 2010), and marginal determinism (Choi et al., 2020). As we impose more properties on a PC, more inference tasks become tractable (i.e. computable in polynomial time), but we also lose some expressive efficiency and generality.

In this paper, we aim to extend the boundaries of advanced inference queries that can be tackled with PCs. In particular, we consider probabilistic inference queries specified as compositions of basic operations such as marginalization, products, and reciprocals, building upon the approach of Vergari et al. (2021). We find that current PC classes are not sufficient to analyze *arbitrary* compositions of these operations, and thus propose a novel class of circuits (*md-vtrees* & MDNet) and accompanying rules (MD-calculus) to derive in a sound and generalizable manner tractability conditions and algorithms for such compositions. Exploiting this, we design the first efficient exact algorithms for causal inference on probabilistic circuits.

Our first contribution is to introduce a unifying formulation of support properties in structured decomposable circuits using *md-vtrees*. We show that *md-vtrees* generalize previously proposed PC families such as probabilistic sentential decision diagrams (PSDD) (Kisa et al., 2014) and structured decomposable and deterministic circuits (Dang et al., 2020; Di Mauro et al., 2021). Notably, we also show that PSDDs are not *optimal* in that we can impose weaker support properties while maintaining tractability for the same inference tasks.

Next, as a practical instantiation of the framework, we propose MDNets, a novel architecture for PCs which can be easily configured to conform to any *md-vtree*. Cru-

1 INTRODUCTION

Probabilistic circuits (PC) (Choi et al., 2020) are a broad family of tractable probabilistic models that are known for their ability to perform *exact* and *efficient* probabilistic inference. For example, in contrast to neural probabilistic models such as variational autoencoders (VAE) (Kingma and Welling, 2013), generative adversarial networks (GAN) (Goodfellow et al., 2014), and normalizing flows (NF) (Rezende and Mohamed, 2015), linear-time exact algorithms on PCs are available for important inference tasks such as computing marginal probabilities, or the maximum probability assignment of variables, for certain types

cially, this allows us to enforce arbitrary support properties needed for tractable inference, including those not covered by existing PC architectures. We show how to learn MD-Nets simply and efficiently using randomized structures and parameter learning (Peharz et al., 2020b,a; Di Mauro et al., 2021).

Finally, for inference, we derive a set of rules for analyzing arbitrary compositions of basic operations using md-vtrees, which we call the MD-calculus. In particular, MD-calculus rules can be applied backward through a given composition, to derive sufficient conditions for tractability on the inputs to the query, which we can enforce during learning through our MDNets. As an application, we demonstrate how the MD-calculus can be applied to derive tractability conditions and algorithms for causal inference estimands on PCs, including the backdoor and frontdoor formulae and (an extension of) the napkin formula.

1.1 Related Work

Support Properties The property of determinism was first introduced in the context of Boolean circuits, specifically, those in negation normal form (Darwiche, 2001; Darwiche and Marquis, 2002), before being naturally extended to arithmetic/probabilistic circuits (Darwiche, 2003). Later, a stronger property known as *strong determinism* was introduced (Pipatsrisawat and Darwiche, 2010; Darwiche, 2011; Kisa et al., 2014) as a convenient means of enforcing determinism in structured decomposable circuits by tying determinism to the scope decomposition, resulting in the (probabilistic) sentential decision diagram (SDD). Oztok et al. (2016) further introduced the notion of *constrained vtrees*, which restricts the structure (scopes, and thus support) of the SDD vtree in order to solve problems on weighted Boolean formulae. Finally, Choi et al. (2020) recently introduced a more general support property called *marginal determinism*, which applies to general probabilistic circuits and is not directly tied to the scope decomposition; our work shows how to construct marginal deterministic circuits, previously considered an intractable task (Choi et al., 2022). Marginal determinism is sufficient for tractability of some marginal MAP queries (Huang et al., 2006).

Causality and Probabilistic Circuits The relationship between probabilistic circuits and causality has its roots in the seminal *compilation* methods of Darwiche (2003), which described an inference approach for (causal) Bayesian networks that involved compiling their graphs into tractable arithmetic circuits; subsequent work has further examined causality and compiled circuits (Butz et al., 2020; Wang et al., 2021; Darwiche, 2021; Chen and Darwiche, 2022). However, obtaining an exact causal interpretation of more general, learned probabilistic circuits has remained an open problem (Zhao et al., 2015; Papantonis and Belle, 2020). The only practical prior causal inference

method for such circuits is the neural parameterization of Zecevic et al. (2021), but this lacks exactness guarantees and is only applicable to fully observed settings. In contrast, we consider exact causal inference using do-calculus, where circuits encode the observed probability distribution.

2 PRELIMINARIES

Notation We use uppercase to denote a random variable (e.g., V) and lowercase for an instantiation of a variable (e.g., v). Sets of variables (and their instantiations) are denoted using bold font (e.g., \mathbf{V} , \mathbf{v}), and we use *val* for the set of all instantiations of a set of variables (e.g., $\text{val}(\mathbf{V})$).

Probabilistic circuits (PC) (Choi et al., 2020) are computational graphs which encode a non-negative function over a set of variables; in particular, they are often used to model (possibly unnormalized) probability distributions.

Definition 1 (Probabilistic Circuit). *A circuit C over variables \mathbf{V} is a parameterized rooted graph, consisting of three types of nodes N : leaf L , sum T and product P . Leaf nodes L are leaves of the graph, while each internal node (sum or product) N has a set of children, denoted $\text{ch}(N)$. Sum nodes have a parameter/weight $\theta_i \in \mathbb{R}^{\geq 0}$ associated with each of their children N_i .*

Each leaf node L encodes a non-negative function $p_L : \phi(L) \rightarrow \mathbb{R}^{\geq 0}$ over a subset of variables $\phi(L) \subseteq \mathbf{V}$, known as its scope. The function encoded by each internal node N is then given by:

$$p_N(\mathbf{V}) := \begin{cases} \prod_{N_i \in \text{ch}(N)} p_{N_i}(\mathbf{V}) & \text{if } N \text{ is a product}^1 \\ \sum_{N_i \in \text{ch}(N)} \theta_i p_{N_i}(\mathbf{V}) & \text{if } N \text{ is a sum} \end{cases}$$

The function encoded by the circuit, $p_C(\mathbf{V})$, is defined to be the function encoded by its root node R . The size of a circuit, denoted $|C|$, is defined to be the number of edges in the circuit.

Definition 2 (Scope and support of PC node). *The scope of an internal node N is the set of variables p_N specifies a function over, recursively defined by $\phi(N) := \bigcup_{N_i \in \text{ch}(N)} \phi(N_i)$. The support of any node N is the set of all instantiations of its scope s.t. p_N is positive, defined as $\text{supp}(N) := \{\mathbf{w} \in \text{val}(\phi(N)) : p_N(\mathbf{w}) > 0\}$.*

The tractability of probabilistic circuits depends on the scope and support properties they satisfy. A PC is *decomposable* if the children of a product node have distinct scopes (and thus partition the scope of the product node), and is *smooth* if the children of a sum node have the same scope. Decomposability and smoothness together enable

¹We assume in this paper that each product node has exactly two children; this does not lose generality as any product node can be converted into a sequence of binary products.

tractable marginal inference; that is, for any subset $\mathbf{W} \subseteq \phi(N)$ of the scope of a node N , we can compute $p_N(\mathbf{W})$ efficiently, where $p_N(\mathbf{W}) := \sum_{\phi(N) \setminus \mathbf{W}} p_N(\phi(N))$ is the *marginal* of the function. A stronger version of decomposability known as *structured decomposability* (Pipatsrisawat and Darwiche, 2008; Kisa et al., 2014) requires the scope of product nodes to decompose according to a *vtree*. Structured decomposability enables efficient computation of additional operations/queries, notably the product of two circuits respecting the same vtree (Pipatsrisawat and Darwiche, 2008; Shen et al., 2016). As for support properties, a PC is *deterministic* if, for every instantiation \mathbf{w} of the scope of a sum node, at most one of its children N_i evaluates to a non-zero value $p_{N_i}(\mathbf{w})$ (equivalently, the supports of the children are distinct). Determinism enables tractability of the MAP inference query, i.e. computing $\max_{\mathbf{V} \setminus \mathbf{E}} p_N(\mathbf{V} \setminus \mathbf{E}, \mathbf{e})$ for some instantiation \mathbf{e} of a set of evidence variables $\mathbf{E} \subseteq \mathbf{V}$.

3 A UNIFYING FRAMEWORK FOR SUPPORT PROPERTIES IN STRUCTURED DECOMPOSABLE CIRCUITS

In this section, we describe our md-vtree framework, which integrates support properties into the vtree formulation of structured decomposable circuits. Using this unifying perspective, we derive a trade-off between the generality of md-vtree circuit classes and tractability, and necessary conditions for optimality of this trade-off.

3.1 Structured Decomposability

The property of structured decomposability is defined with respect to a variable tree known as a *vtree*.

Definition 3 (Vtree). *A vtree $v = (M, E)$ for a set of variables \mathbf{V} is a rooted binary tree with nodes M and edges E , whose leaves m each correspond to a subset $\phi(m) \subseteq \mathbf{V}$, such that the subsets for all leaves form a partition of \mathbf{V} .*

We define the *scope* of a leaf m to be $\phi(m)$, and the scope of any other node to be $\phi(m) = \cup_{m_i \in \text{ch}(m)} \phi(m_i)$. Further, we write $v_m = (M_m, E_m)$ to denote the vtree rooted at m .

Intuitively, a vtree specifies how the scope of product nodes decompose in a circuit. However, our definition of structured decomposability differs from typical recursive definitions (Pipatsrisawat and Darwiche, 2008; Darwiche, 2011; Shih et al., 2019) in that the key condition is on the scopes of the sum/leaf nodes, without directly placing conditions on the product nodes (besides decomposability):

Definition 4 (PC respecting vtree). *Let C be a PC and $v = (M, E)$ be a vtree, both over variables \mathbf{V} . We say that C respects v if (1) C is smooth² and decomposable; and (2)*

for every leaf node and non-trivial³ sum node $N \in C$, there exists a vtree node $m \in M$ such that $\phi(N) = \phi(m)$.

Definition 5 (Structured Decomposability). *A PC C is structured decomposable if it respects some vtree v .*

Structured decomposability enables tractable products of circuits respecting compatible vtrees, i.e. those which have the same structure when projected onto their common variables $\mathbf{C} := \mathbf{V}^{(1)} \cap \mathbf{V}^{(2)}$ (see Appendix for details). In the next subsection, we will show how support properties, which are also specified on the sum nodes in the circuit, can be neatly integrated into vtrees.

3.2 Structured Marginal Determinism

Our new definition of structured decomposability based on sum nodes provides the basis for us to specify a novel systematic characterization of support properties in structured decomposable circuits, which we call *structured marginal determinism*. First, we reformulate the definitions of marginal determinism from Choi et al. (2020).

Definition 6 (Restricted Scope). *For a PC node N (resp. vtree node m), and given a set $\mathbf{C} \subseteq \mathbf{V}$, the restricted scope is defined as $\phi_{\mathbf{C}}(N) := \phi(N) \cap \mathbf{C}$ (resp. $\phi_{\mathbf{C}}(m) := \phi(m) \cap \mathbf{C}$).*

Definition 7 (Marginalized Support). *For any PC node N and subset of variables $\mathbf{Q} \subseteq \mathbf{V}$, we define the marginalized support of N with respect to \mathbf{Q} as $\text{supp}_{\mathbf{Q}}(N) := \{\mathbf{q} \in \text{val}(\mathbf{Q}) : p_N(\mathbf{q}) > 0\}$.*

Note that \mathbf{Q} can contain variables outside of $\phi(N)$; in a slight abuse of notation, we write $p_N(\mathbf{q})$ for $p_N(\mathbf{q} \cap \phi(N))$.

Definition 8 (Marginal Determinism). *A sum node is marginal deterministic with respect to a subset $\mathbf{Q} \subseteq \mathbf{V}$ (written \mathbf{Q} -deterministic) if the children of the sum node have distinct marginalized support, i.e. $\text{supp}_{\mathbf{Q}}(N_i) \cap \text{supp}_{\mathbf{Q}}(N_j) = \emptyset$ for N_i, N_j distinct children of T .*

Definition 9 (Marginal Determinism of PC). *A PC is marginal deterministic with respect to a subset $\mathbf{Q} \subseteq \mathbf{V}$ (written \mathbf{Q} -deterministic) if for every sum node T , either:*

- $\phi(T)$ does not overlap with \mathbf{Q} , i.e. $\phi_{\mathbf{Q}}(T) = \emptyset$; or
- The sum node T is \mathbf{Q} -deterministic.

For example, normal determinism is equivalent to C being marginal deterministic with respect to \mathbf{V} . In general, there is no straightforward relation between \mathbf{Q} -determinism and \mathbf{Q}' -determinism for different sets \mathbf{Q}, \mathbf{Q}' . In particular, neither determinism (i.e. \mathbf{V} -determinism) nor \mathbf{Q} -determinism imply each other in general; for example, a circuit can be \mathbf{Q} -deterministic but not deterministic if there exist some

²A structured decomposable circuit can be smoothed in near-linear time (Shih et al., 2019).

³A sum node is non-trivial iff it has more than one child.

²A structured decomposable circuit can be smoothed in near-

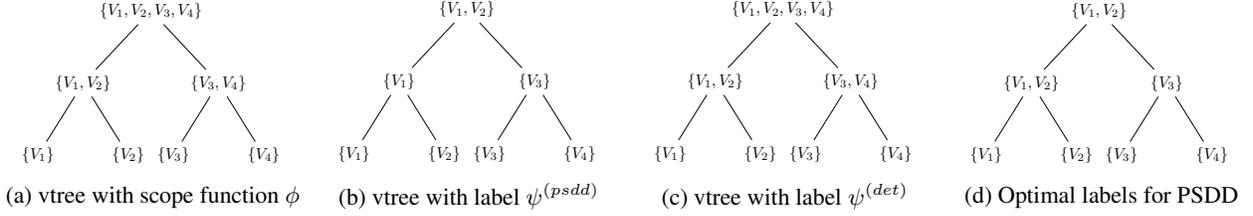


Figure 1: Example of md-vtree with scope function, and three different labelling functions.

sum nodes with $\phi_Q(T) = \emptyset$. Thus, we use $\mathcal{Q}(C)$ to denote the set of all sets $Q \subseteq V$ such that C is Q -deterministic; this provides a characterization of the support properties of the circuit.

Now, for a given PC C , and any sum node T in that PC, let $\psi(T)$ be the set of all sets Q such that T is Q -deterministic; we call this a *labelling function*. Note that the label function ψ is a *specification* of marginal determinism for the circuit; that is, it is sufficient to deduce $\mathcal{Q}(C)$. We make two observations that allow us to simplify the labelling function, one straightforward, and one more subtle. Firstly, we note that Q -determinism for the circuit imposes the same requirement on all nodes with the same scope; thus we restrict ψ to have the same value for all sum nodes with the same scope. For structured decomposable circuits, we can thus write $\psi(m)$ as a function of the vtree node m .

The second observation is that, under some assumptions, we can actually specify $\psi(m)$ using a *single* set $Q \subseteq V$.

Proposition 1 (Conflicting Q -Determinisms for Sum Nodes). *Let C be a PC, and let $Q, Q' \subseteq V$ such that neither is a subset of the other. Suppose that there exists a non-trivial sum node T in C that is Q -deterministic and Q' -deterministic, but not $(Q \cap Q')$ -deterministic. Then the circuit rooted at T , C_T , cannot have full support.*

Proposition 1 says that, if we want $\psi(m)$ to contain two sets Q, Q' which are not subsets of each other, then this necessarily restricts the support of the circuit. While it can be beneficial to enforce a restricted support on a PC if we have prior knowledge (Kisa et al., 2014), it is undesirable in our case where restricting support comes as a *side effect* of enforcing tractability, as this can result in bias when learning. As such, we only consider labellings $\psi(m)$ where, for every $Q, Q' \in \psi(m)$, we have $Q \subseteq Q'$ or $Q' \subseteq Q$.

Proposition 2 (Superset Q -Determinisms for Sum Nodes). *Suppose that a sum node T is Q -deterministic. Then it is also Q' -deterministic for any $Q \subseteq Q' \subseteq V$.*

Using Proposition 2, it now follows that $\psi(m)$ must take the form $\{Q' | Q \subseteq Q' \subseteq V\}$ for some Q . As a result, we can just label our vtree node m with Q , i.e. $\psi(m) = Q$.

This motivates our characterization of structural marginal determinism based on the concept of a *md-vtree*, which provides a means of specifying the support properties that

a structured decomposable circuit satisfies.

Definition 10 (md-vtree). *A md-vtree $w = (v, \psi)$ for a set of variables V consists of a vtree $v = (M, E)$ over V , together with a labelling function ψ .*

The labelling function maps a vtree node $m \in M$ to some element in $\mathcal{P}(\phi(m)) \cup \{U\}$, where U is the universal set.⁴

Definition 11 (PC respecting md-vtree). *Let C be a PC and $w = (v, \psi)$ be a md-vtree, both over variables V . Then we say that C respects w if 1) C respects v ; and 2) for any sum unit $T \in C$, T is marginally deterministic with respect to $\psi(m)$, where m is the vtree node such that $\phi(T) = \phi(m)$.*

We denote the class of circuits respecting w by \mathcal{C}_w .

Intuitively, md-vtrees capture both structured decomposability, as well as a marginal determinism “pattern” that the circuit must follow:

Definition 12 (Implied Q -Determinisms). *For any set $Q \subseteq V$, we say that Q -determinism is implied by a md-vtree w if, for every vtree node $m \in M$ such that $\phi(m) \cap Q \neq \emptyset$, it is the case that $Q \supseteq \psi(m)$. We write $\mathcal{Q}(w)$ to denote the set of all sets Q s.t. Q -determinism is implied by w .*

Proposition 3 (Validity of Implied Q -Determinisms). *For any PC C respecting md-vtree w , both over V , and any $Q \subseteq V$ s.t. w implies Q -determinism, it follows that C is Q -deterministic.*

Notice that there is a trade-off between generality (expressivity) of the PC class, and the support properties it supports. Increasing the size of the labelling sets will improve the former, but hurt the latter.

Theorem 1 (Generality-Tractability Tradeoff). *Let $w = (v, \psi)$ and $w' = (v, \psi')$ be two md-vtrees, such that $\psi'(m) \supseteq \psi(m)$ for all $m \in M$. Then we have that $\mathcal{Q}(w) \supseteq \mathcal{Q}(w')$, and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$.*

It is worth commenting on the two extremes of possible labels; namely, the universal set, and the empty set. The role of the universal set label U is to indicate that no Q -determinism properties hold (including normal determinism). On the other hand, the empty set indicates that any sum node T corresponding to m can only have one child

⁴The universal set satisfies, for any set S , $U \supseteq S$, $U \not\subseteq S$ (unless S is U), $U \cap S = S$, and $U \cup S = U$.

N_i which is not zero, i.e. $p_{N_i} \equiv 0$; thus the sum node must be trivial. In practice, this means that it must represent a factorized distribution with factors corresponding to the scopes of the children of m .

Examples To the best of our knowledge, the only concrete type of probabilistic circuit proposed in the literature that imposes non-trivial marginal determinism constraints (i.e., not just normal determinism) is the probabilistic sentential decision diagram (PSDD) (Kisa et al., 2014). PSDDs satisfy structured decomposability and a property known as strong determinism. In the language of md-vtrees, this corresponds to requiring that, for any non-leaf vtree node m , and children m_1, m_2 of m , the label $\psi(m)$ is either $\phi(m_1)$ or $\phi(m_2)$ (which is then referred to as the *left* child). For example, for the vtree over $V = \{V_1, V_2, V_3, V_4\}$, shown in Figure 1a together with the scopes for each vtree node, the label function $\psi^{(psdd)}$ is given on the right in Figure 1b.

Despite implementing strong determinism, recent work has shown that almost all of the tractable queries and operations that PSDDs support require only structured decomposability and determinism (Dang et al., 2020). This raises the question of whether strong determinism adds anything. To analyse this, in Figure 1c we show the labelling function $\psi^{(det)}$ which defines a deterministic circuit. With these representations, we can deduce the Q -determinism properties that any PSDD, or structured decomposable and deterministic circuit, must satisfy, by finding the set $\mathcal{Q}(w)$ of sets Q which its md-vtree implies. In this example, by enumerating all sets $Q \subseteq V$ and checking the condition, we can see that $\mathcal{Q}(w^{psdd}) = \{\{V_1, V_2\}, \{V_1, V_2, V_3\}, \{V_1, V_2, V_3, V_4\}\}$, while $\mathcal{Q}(w^{det}) = \{\{V_1, V_2, V_3, V_4\}\}$. This shows that PSDDs do have additional Q -determinisms, which means, for example, that they are more tractable with regards to MMAP queries. In fact, in the particular case of (P)SDDs, the implied Q -determinisms $\mathcal{Q}(w^{psdd})$ coincide with the definition of Q -constrained vtrees (Oztok et al., 2016).

3.3 Regular md-vtrees

Given the trade-off between expressivity and tractability for md-vtrees, one might ask how to choose the labelling function in practice. One reasonable strategy would be to enforce some marginal determinism properties that we require for tractability of some inference task, and optimize for expressivity within this constraint.

Problem 1 (Labelling Selection). *Given a vtree v , and a set S of subsets $Q \subseteq V$, choose a labelling function ψ such that $w = (v, \psi)$ implies Q -determinism for all $Q \in S$, i.e. $\mathcal{Q}(w) \supseteq S$, while maximizing expressivity.*

We propose a simple algorithm to tackle this problem (Algorithm 1), which directly enforces the necessary labels for

Algorithm 1: Optimal Labelling

Input: vtree $v = (M, E)$, required set of marginal determinisms S

Result: labelling function ψ for v

```

1 for  $m \in M$  do
2    $\psi(m) \leftarrow U$ 
3   for  $Q \in S$  do
4     if  $Q \cap \phi(m) \neq \emptyset$  then
5       |  $\psi(m) \leftarrow \psi(m) \cap Q$ 
6   if  $\psi(m) \neq U$  then
7     |  $\psi(m) \leftarrow \psi(m) \cap \phi(m)$ 
8 Return  $w = (v, \psi)$ 

```

each vtree nodes; it can be seen that it is optimally expressive, in the sense that increasing the size of any labelling set will result in some losing Q -determinism for some $Q \in S$. It turns out that such labelling functions have a very specific structure:

Definition 13 (Regular md-vtree). *We say that a md-vtree $w = (v, \psi)$ is regular if for every non-leaf node m , and its children m_1, m_2 , it holds that either $\psi(m) = \psi(m_1)$, $\psi(m) = \psi(m_2)$, or $\psi(m) = \psi(m_1) \cup \psi(m_2)$.*

Proposition 4 (Regularity of Algorithm 1). *The output of Algorithm 1 is a regular md-vtree.*

The following Theorem shows that regular md-vtrees are optimal in the sense that for any given marginal determinism requirement S , (one of) the most expressive md-vtree is always a regular md-vtree. This means that we can restrict our attention to the much smaller space of regular md-vtrees. In particular, the labelling function of a regular md-vtree is entirely determined by the labelling of each leaf node $\psi(m_{leaf})$, and a ternary variable over values $\{f, s, b\}$ for each non-leaf node, indicating whether the label depends on the label of the first child, second child, or both.

Theorem 2 (Optimal md-vtrees). *Let $w = (v, \psi)$ be any md-vtree. Then there exists a regular md-vtree $w' = (v, \psi')$ such that $\mathcal{Q}(w) = \mathcal{Q}(w')$, and $\mathcal{C}_{w'} \supseteq \mathcal{C}_w$.*

While Algorithm 1 always returns an optimal labelling for a given vtree satisfying the required marginal determinisms, the expressivity of the circuit class may differ depending on the vtree. For example, for some vtrees, Algorithm 1 may output a labelling function such that $\psi(m)$ is empty for some vtree nodes m , i.e. a factorized distribution. We leave designing optimally expressive vtrees for a given set S of marginal determinisms as an open problem.

Examples Let us return to the PSDD example from Figure 1. It can be easily checked that the corresponding md-vtree w^{psdd} in Figure 1b is not regular. Following Algorithm 1, we therefore construct in Figure 1d an regular md-vtree w^{opt} that retains the same $\mathcal{Q}(w^{opt}) = \mathcal{Q}(w^{psdd}) =$

$\{\{V_1, V_2\}, \{V_1, V_2, V_3\}, \{V_1, V_2, V_3, V_4\}\}$. We should prefer w^{opt} over w^{psdd} as it imposes less constraints/more circuits respect w^{opt} . In other words, PSDDs impose more constraints than they “need to” to obtain their marginal determinism properties.

4 MDNETS: A PRACTICAL ARCHITECTURE FOR MD-VTREES

In this section, we show how to construct and learn a probabilistic circuit that respects a particular md-vtree. It is worth noting that, as special cases of md-vtrees, we can use existing architectures and learning algorithms for PCs such as PSDDs (Liang et al., 2017) and structured decomposable and deterministic circuits (Dang et al., 2020; Di Mauro et al., 2021). However, we have seen that PSDDs are not optimally expressive, and to enforce tractability, we may need to target md-vtrees which do not fall into these existing categories, such as those generated from Algorithm 1. We thus propose a novel PC architecture, MDNet, which enforces a given regular md-vtree *by design*.

The key component of MDNets is the *node group*, which is a vector of sum nodes with the same scope (i.e. corresponding to the same vtree node m) with the property that the nodes in the group have disjoint marginalized support $\text{supp}_{\psi(m)}(N)$. Intuitively, the sum nodes in a group provide a partition of the domain of $\psi(m)$, which we use as an invariant in order to enforce the required marginal determinisms throughout the circuit. More formally, suppose that we have a non-leaf vtree node m , and let m_l, m_r be its children. We refer to all sum nodes corresponding to a vtree node as being a *layer*, and assign G groups T_1, \dots, T_G to the layer for vtree node m , and similarly $T_1^{(l)}, \dots, T_{G_l}^{(l)}, T_1^{(r)}, \dots, T_{G_r}^{(r)}$ to the layers for m_l, m_r . For regular md-vtrees, the label $\psi(m)$ is either equal to $\psi(m_l)$ or $\psi(m_r)$, or is their union. We handle these cases separately, as *mixing* and *synthesizing* layers.

Mixing Layer If $\psi(m) = \psi(m_l)$ or $\psi(m_r)$, we implement a *mixing* layer. W.l.o.g. we assume $\psi(m) = \psi(m_l)$. For each group $T_i = (T_{i,1}, \dots, T_{i,k})$, and for each sum node $T_{i,k}$ in the group, we assign a set of product nodes $P_{i,k}$ to $T_{i,k}$. Each product node has two children; the left child being a node from $T_1^{(l)}, \dots, T_{G_l}^{(l)}$ and the right child being a node from $T_1^{(r)}, \dots, T_{G_r}^{(r)}$. We place the following restriction: the left children of the product nodes $\cup_k P_{i,k}$ must all be distinct, and all come from a single group. This ensures that all of the sum nodes are $\psi(m) = \psi(m_l)$ -deterministic, and further that each group T_i satisfies the invariant.

Synthesizing Layer If $\psi(m) = \psi(m_l) \cup \psi(m_r)$, we assign product nodes $P_{i,k}$ to each sum node $T_{i,k}$ as before, but with a different restriction; we now require that both the left children and right children of the product nodes $\cup_k P_{i,k}$

Operation	Requirements	Output Encodes
MARG($C; \mathbf{W}$)	-	$\sum_{\mathbf{W}} p_C(\mathbf{V})$
INST($C; \mathbf{w}$)	-	$p_C(\mathbf{w}, \mathbf{V} \setminus \mathbf{W})$
PROD(C_1, C_2)	Cmp. Vtrees	$p_{C_1}(\mathbf{V}) \times p_{C_2}(\mathbf{V})$
POW($C; \alpha$)	Det	$p_C(\mathbf{V})^\alpha _{\text{supp}(C)}$
MAX(C)	Det	$\max_{\mathbf{V}} p_C(\mathbf{V})$
LOG(C)	Det	$\log p_C(\mathbf{V}) _{\text{supp}(C)}$

Table 1: Definitions of basic operations.

are nodes coming from a single group from their respective layers, and that each product node has a unique combination of children from these groups.

Intuitively, mixing layers achieve their marginal determinism by “copying” the marginal determinism of one of their child layers, while mixing over groups in the other child layer. On the other hand, synthesizing layers enforce marginal determinism by combining, or synthesizing, the marginal determinism properties of both of their children.

For simplicity, we propose to learn MDNets exploiting recent advancements in random structures for PC learning (Peharz et al., 2020a,b; Di Mauro et al., 2021): in particular, we propose to choose the MDNet structure randomly within the constraints, and then learn the parameters using standard MLE estimation if the md-vtree implies (\mathbf{V} -)determinism, or use EM otherwise (Peharz, 2015).

5 COMPOSITIONAL INFERENCE USING STRUCTURED MARGINAL DETERMINISM

In this section, we will describe a methodology that exploits our md-vtree framework as a language for deriving tractability conditions for arbitrary compositions of basic operations on probabilistic circuits. In particular, we build upon the work of Vergari et al. (2021), showing how to extend their analysis to compositional queries which include marginalization (integration) operations at arbitrary points in the pipeline, and allow for maximization queries (e.g. marginal MAP).

5.1 Support Properties in Compositional Inference

In Table 1, we define the basic probabilistic inference operations, including marginalization, products, instantiation, powers, maximization, and logarithms, along with the properties (*requirements*) under which there exist efficient (polytime) algorithms for computing them on PCs (Vergari et al., 2021); note that we assume decomposability and smoothness by default. These operations produce a circuit encoding the specified function (or scalar in the

case of MAX). We use $|_{\text{supp}(C)}$ to denote the *restricted* power/logarithm, as these functions are not defined at 0:

$$f(p_C(\mathbf{V}))|_{\text{supp}(C)} := \begin{cases} f(p_C(\mathbf{V})) & \text{if } p_C(\mathbf{V}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

We refer to the basic operations in the bottom half of the table as *deterministic* operations, as their tractability depends on the input circuit being deterministic; they are NP-complete (MAX) or #P-hard (POW, LOG) otherwise (Choi and Darwiche, 2017; Vergari et al., 2021).

Many complex inference queries can be expressed as compositions of these basic operations; we show a selection of examples in Table 2. We can interpret such compositions as *pipelines*, or computational graphs, which specify an algorithm for computing the query that uses the efficient algorithms for each basic operation. To show tractability of a pipeline for given input circuits, one needs to show that the inputs to intermediate operations satisfy the requirement for tractability of that operation. For this purpose, Vergari et al. (2021) derive *input-output conditions* for basic operations, which specify a pair of properties such that the output of the operation is guaranteed to satisfy the output property if the input satisfies the input property. This allows properties to be soundly propagated through the pipeline, from the input circuits.

However, there remains a significant unresolved challenge for analyzing general compositions of basic operations; namely, analyzing how operations affect *support properties* of the circuit beyond just determinism. For example, the marginal MAP problem (MMAP) (Huang et al., 2006; Choi et al., 2020) in Table 2 is a canonical inference task that can be decomposed into a composition of a MARG operation $\sum_{\mathbf{V} \setminus \mathbf{W}} p_C(\mathbf{W}, \mathbf{V} \setminus \mathbf{W})$, and a MAX operation $\max_{\mathbf{W}} p_C(\mathbf{W})$. The MAX operation $\max_{\mathbf{W}} p_C(\mathbf{W})$ is known to be tractable for any deterministic input circuit. Unfortunately, however, ensuring that the output of MARG($C; \mathbf{V} \setminus \mathbf{W}$) is deterministic is known to be NP-hard (Shen et al., 2016), even if C is deterministic. For similar reasons, any compositional inference task in which a deterministic operation appears after a marginalization operation cannot be currently analyzed, notable examples of which we show in the lower half of Table 2.

5.2 Operations on md-vtrees

To tackle these challenges, we apply our md-vtree framework as a unified language for scope and support in structured decomposable circuits. The first problem that we would like to address is that of determining whether a pipeline is tractable for given input md-vtrees:

Problem 2 (Forward Problem). *Given a query expressed as a pipeline of basic operations, and md-vtree(s) that the input circuits respect, determine if the pipeline is tractable.*

To solve this problem, we propose to derive input-output conditions in terms of md-vtrees. In other words, if we have input circuit(s) that respect some given md-vtree(s), can we obtain a md-vtree that the output of a basic operation is guaranteed to respect? In the Appendix, we detail a set of algorithms for each of the operations in Table 1, which take as input md-vtree(s) and return an output md-vtree that provides exactly such a guarantee, based upon the corresponding algorithms on circuits. Then, given any pipeline, and input md-vtree(s), we can determine if the compositional query is tractable, simply by propagating md-vtree(s) *forward* through the pipeline, and checking that the input md-vtree(s) to any intermediate operation satisfy the requirements in Table 1. Importantly, this can be done *without doing the computation of the query pipeline itself*; all of our algorithms run in polytime in the number of variables $|\mathbf{V}|$, which is much smaller than the circuits themselves.

5.3 The MD-calculus

The forward problem allows us to reason about tractability if we already have the input circuits. However, when learning circuits from data, we have the freedom to choose the md-vtree(s) in order to enable tractable inference:

Problem 3 (Backward Problem). *Given any query expressed as a pipeline of basic operations, derive input md-vtree(s) such that the pipeline is tractable.*

To this end, in Table 3 we show a set of input-output conditions called the *MD-calculus*. These are sufficient conditions on the input(s) to an operation to guarantee that the output is Q -deterministic. The MD-calculus forms a set of rules that we can apply backwards from deterministic operations (which require \mathbf{V} -determinism), in order to determine a sufficient set of marginal determinisms \mathbf{S} for each intermediate circuit. Finally, we can enforce those marginal determinisms on the input md-vtree(s) using Algorithm 1. We defer proofs of the MD-calculus rules to the Appendix.

Theorem 3 (MD-calculus). *The conditions in Table 3 hold.*

We show examples of compositional queries in the bottom half of Table 2, with the marginal determinism condition on the input circuit C derived using this approach. For MMAP, for the MAX operation, we require MARG($C; \mathbf{V} \setminus \mathbf{W}$) to be \mathbf{W} -deterministic. Using the MD-calculus rule for MARG, we can see that it is sufficient for C to also be \mathbf{W} -deterministic. For mutual information, applying a similar approach we obtain that C should be both \mathbf{X} -deterministic and \mathbf{Y} -deterministic, but we have seen in Proposition 1 that this is not possible without restricting support.

6 APPLICATION: CAUSAL INFERENCE

In this section, we use our md-vtree framework to analyse tractability conditions for *exact* causal inference for

Task	Computation	Operations	Condition
Cross-Entropy	$-\sum_{\mathbf{V}} p_{C^{(1)}}(\mathbf{V}) \log(p_{C^{(2)}}(\mathbf{V}))$	MARG, PROD, LOG	$C^{(1)}, C^{(2)}$ Cmp.; $C^{(2)}$ Det.
Var. Elim.	$\sum_{\mathbf{W}} p_{C^{(1)}}(\mathbf{V}) p_{C^{(2)}}(\mathbf{V})$	MARG, PROD	$C^{(1)}, C^{(2)}$ Cmp.
MMAP	$\max_w \sum_{\mathbf{V} \setminus \mathbf{W}} p_C(w, \mathbf{V} \setminus \mathbf{W})$	MARG, MAX	C Mdet. wrt. \mathbf{W}
Mut. Inf.	$p_C(\mathbf{X}, \mathbf{Y}) \log \frac{p_C(\mathbf{X}, \mathbf{Y})}{p_C(\mathbf{X}) p_C(\mathbf{Y})}$	MARG, PROD, LOG, POW	-
BD Adj.	$\sum_{\mathbf{Z}} p_C(\mathbf{Y} \mathbf{X}, \mathbf{Z}) p_C(\mathbf{Z})$	MARG, PROD, POW	C Mdet. wrt. $\mathbf{X} \cup \mathbf{Z}$; Str. Dec.

Table 2: Examples of complex inference tasks expressed as compositions of basic operations.

Operation	Requirement	Input Condition	Output Condition
MARG($C; \mathbf{W}$)	-	Q -det	Q -det
INST($C; w$)	-	$\exists \mathbf{W}' \subseteq \mathbf{W} : (Q \cup \mathbf{W}')\text{-det}$	Q -det
PROD($C^{(1)}, C^{(2)}$)	$C^{(1)}, C^{(2)}$ respect compatible vtrees	$\exists Q^{(1)}, Q^{(2)} : Q^{(1)}\text{-det}, Q^{(2)}\text{-det}, \text{ and:}$ <ul style="list-style-type: none"> • Either (a) $Q \subseteq V^{(1)} \cap V^{(2)}$ and $Q^{(1)} = Q^{(2)} = Q$; • Or (b) $Q^{(1)}, Q^{(2)} \supseteq V^{(1)} \cap V^{(2)}$ and $Q = Q^{(1)} \cup Q^{(2)}$ 	Q -det
POW($C; \alpha$)	Det	Q -det	Q -det
MAX(C)	Det	N/A	N/A (scalar output)
LOG(C)	Det	-	-

Table 3: MD-calculus: sufficient input-output conditions for each basic operation

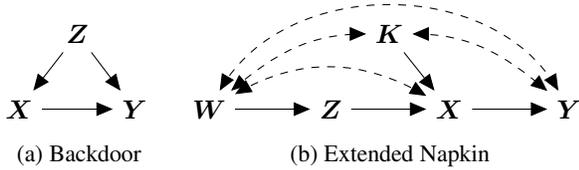


Figure 2: Examples of causal diagrams

PCs. The typical setup in causal inference is that we have access to an observed distribution $p(\mathbf{V})$, and some domain assumptions, often conveniently expressed using a *causal diagram* (Pearl, 2009), and we are interested in computing some *interventional distribution* $p_{\mathbf{X}}(\mathbf{Y})$, where $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ are disjoint subsets of the observed variables. $p_{\mathbf{X}}(\mathbf{Y})$ is said to be *identifiable* if the assumptions are sufficient for it to be uniquely determined; and if it is identifiable, a causal formula/estimand can be obtained using the *do-calculus* (Pearl, 1995; Shpitser and Pearl, 2006).

6.1 Hardness of Backdoor Adjustment on Circuits

We assume that the observed data distribution is modelled by a PC C (perhaps learned from data), and we wish to compute some causal query. One of the most common cases where the interventional distribution is identifiable is when there exists a valid *backdoor adjustment set* $\mathbf{Z} \subseteq \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})$ (also known as the conditional exchangeability/ignorability assumption), an example of which is illustrated in Figure 2a. Whenever such a set exists, the in-

terventional distribution $p_{C, \mathbf{X}}(\mathbf{Y})$ is given by the backdoor adjustment formula⁵:

$$\sum_{\mathbf{Z}} p_C(\mathbf{Z}) p_C(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}) \frac{p_C(\mathbf{Y}, \mathbf{X}, \mathbf{Z})}{p_C(\mathbf{X}, \mathbf{Z})}$$

Unfortunately, we now show that the backdoor adjustment is not tractable for existing classes of probabilistic circuits:

Theorem 4 (Hardness of Backdoor Query). *The backdoor query for decomposable and smooth PCs is #P-hard, even if the PC is structured decomposable and deterministic.*

We can also view this result as placing a theoretical barrier on interpreting probabilistic circuits as causal models (Zhao et al., 2015; Papantonis and Belle, 2020); namely that, even if such an interpretation exists, it is not possible to tractably perform causal inference with the causal PC. Thus, whether we interpret a PC as itself expressing causality, or merely a model of the observational probability distribution, new PC classes are required. In the following, we take the latter perspective, employing our md-vtrees.

6.2 MD-calculus for Causal Formulae

We now employ the MD-calculus to derive tractability conditions for the backdoor query, given $C(\mathbf{V})$ as input. Note that there is only one deterministic operation, namely the reciprocal $\text{POW}(\cdot; -1)$. The input to this operation is

⁵As commonly done in causal inference we assume positivity, i.e. $p_C(\mathbf{V}) \geq 0$ such that the conditional is well defined.

$C(\mathbf{X}, \mathbf{Z})$, which we thus require to be deterministic, that is, $(\mathbf{X} \cup \mathbf{Z})$ -deterministic. We then have that $C(\mathbf{X}, \mathbf{Z}) = \text{MARG}(C(\mathbf{V}); \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Z}))$, so, using the MD-calculus rule for MARG, it is sufficient for $C(\mathbf{V})$ to be $(\mathbf{X} \cup \mathbf{Z})$ -deterministic. Given a vtree, we can then use Algorithm 1 to derive an md-vtree for the input circuit $C(\mathbf{V})$.

It is worth noting that, unlike typical causal inference approaches which derive a scalar $p_{C,\mathbf{x}}(\mathbf{y})$ (or some causal effect) for a particular intervention $\mathbf{X} = \mathbf{x}$, the output of the backdoor query here is a *circuit* over variables \mathbf{X}, \mathbf{Y} encoding $p_{C,\mathbf{x}}(\mathbf{Y})$. This means that we can do further downstream reasoning over different values of \mathbf{X}, \mathbf{Y} .

As a second example, consider the *extended napkin* causal diagram in Figure 2b. We can use the do-calculus to derive the following expression for $p_{C,\mathbf{x}}(\mathbf{Y})$ in this case⁶:

$$\sum_{\mathbf{K}} f(C, \mathbf{K}) \frac{\sum_{\mathbf{W}} p_C(\mathbf{X}, \mathbf{Y} | \mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}{\sum_{\mathbf{W}} p_C(\mathbf{X} | \mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}$$

For this formula, we can try to apply the MD-calculus as usual. The denominator (input to the $\text{POW}(\cdot; -1)$ operation) is a function of $\{\mathbf{X}, \mathbf{K}\}$; thus, passing through the marginalization operation, we require $\text{PROD}(C(\mathbf{X} | \mathbf{K}, \mathbf{z}, \mathbf{W}), C(\mathbf{W}, \mathbf{K}))$ to be $(\mathbf{X} \cup \mathbf{K})$ -deterministic. However, if we then look at the conditions for the PROD operation, we see this cannot be achieved. In particular, $C(\mathbf{X} | \mathbf{K}, \mathbf{z}, \mathbf{W})$ and $C(\mathbf{W}, \mathbf{K})$ are circuits over $\mathbf{V}^{(1)} = \{\mathbf{X}, \mathbf{K}, \mathbf{W}\}$ and $\mathbf{V}^{(2)} = \{\mathbf{K}, \mathbf{W}\}$ respectively, and there are no possible $\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}$ such that condition (a) or (b) holds.

However, we *can* derive a tractable pipeline for $p_{C,\mathbf{x}}(\mathbf{Y})$ where we commit to a specific intervention $\mathbf{X} = \mathbf{x}$. In this case, the new denominator is a function of just \mathbf{K} , and so we require $\text{PROD}(C(\mathbf{x} | \mathbf{K}, \mathbf{z}, \mathbf{W}), C(\mathbf{W}, \mathbf{K}))$ to be $(\mathbf{X} \cup \mathbf{K}) \setminus (\mathbf{X}) = \mathbf{K}$ -deterministic. This can be achieved by choosing $\mathbf{Q}^{(1)} = \mathbf{Q}^{(2)} = \{\mathbf{K}\}$, which satisfies condition (a) for PROD. By analyzing the rest of the formula, we can derive a set \mathcal{S} of marginal determinisms for $C(\mathbf{V})$ that is sufficient for the extended napkin query, and apply Algorithm 1; we defer the full derivations (and analysis of the frontdoor formula) to the Appendix.

7 EMPIRICAL EVALUATION

In this section, we empirically evaluate our tractable algorithm for backdoor adjustment derived using MD-calculus. We generate datasets by sampling 1000 datapoints from the (discrete variable) Bayesian network (BN) models in the *bnrepository* (Scutari, 2022), and learn a MDNet over all variables \mathbf{V} from data. For each Bayesian network (causal graph), we select a single treatment variable X and single

⁶Here, \mathbf{z} can be any instantiation of \mathbf{Z} , and $f(C, \mathbf{K})$ is an expression over C, \mathbf{K} that we omit here for clarity.

Dataset	Z	Error		Time	
		MD	Counting	MD	Counting
Asia	4	0.269	0.0143	0.5	1.0
Sachs	4	0.180	0.0219	1.7	0.7
Child	13	0.0802	0.135	1.9	0.9
Win95pts	59	0.0044	0.0511	3.2	0.9
Andes	202	0.0382	0.0982	7.9	1.3

Table 4: Backdoor Estimation (averaged over 10 runs)

outcome variable Y , as well as a set of variables \mathbf{Z} forming a valid backdoor adjustment set for (X, Y) , and seek to estimate $\sum_{\mathbf{Z}} p(Y | X, \mathbf{Z}) p(\mathbf{Z})$. We manually select a vtree which splits the scope into $(X \cup \mathbf{Z})$ and Y at the root, and thereafter generates the rest of the vtree randomly. Given a vtree, we use Algorithm 1 to generate a labelling, i.e. regular md-vtree. The required tractability properties for backdoor adjustment are then enforced through the structure of the corresponding MDNet.

The results are shown in Table 4; for comparison, we show also results for the *counting* approach, where we estimate $p(Y | X, \mathbf{Z})$ as $\frac{N_{Y,X,Z}}{N_{X,Z}}$, where N refers to the number of datapoints with the subscripted assignment of variables (this is set to 0 if $N_{X,Z} = 0$). It can be seen that, while the counting approach is generally more robust in lower dimensions, the advantage in terms of learning a full model becomes apparent with the higher-dimensional adjustment sets, as shown by the lower error on the Win95pts and Andes datasets. Remarkably, as the size of the adjustment set \mathbf{Z} increases, the time taken for the algorithm based on the MD-calculus increases only approximately *linearly* in the dimension. This illustrates the attractiveness of tractable probabilistic modelling, in that we can systematically control the computational cost of exact inference by restricting the size of the PC model.

8 CONCLUSION

In summary, we introduced the md-vtree framework for support properties in structured decomposable circuits, and showed how it can be employed for reasoning about tractability of compositional inference queries using our MD-calculus rules. Our unifying framework naturally provides insight into the properties of previously proposed PC classes such as PSDDs, as well as inspiring our newly designed MDNet architecture. Nonetheless, there remain a number of interesting challenges for future work. For example, for more challenging datasets, how can we also learn the vtree and/or MDNet structure from data while maintaining tractability? For what other inference queries can we derive tractability conditions using the MD-calculus? We hope that our work will help lay the theoretical foundations for tackling these questions.

Acknowledgements

This project was funded by the ERC under the European Union’s Horizon 2020 research and innovation programme (FUN2MODEL, grant agreement No.834115).

References

- Butz, C., S. Oliveira, J., and Peharz, R. (2020). Sum-product network decompilation. In Jaeger, M. and Nielsen, T. D., editors, *Proceedings of the 10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, pages 53–64. PMLR.
- Chen, Y. and Darwiche, A. (2022). On the definition and computation of causal treewidth. In *Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence*.
- Choi, A. and Darwiche, A. (2017). On relaxing determinism in arithmetic circuits. In *Proceedings of the 34th International Conference on Machine Learning, ICML’17*, page 825–833.
- Choi, Y., Friedman, T., and Van den Broeck, G. (2022). Solving marginal map exactly by probabilistic circuit transformations. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- Choi, Y., Vergari, A., and Van den Broeck, G. (2020). Probabilistic circuits: A unifying framework for tractable probabilistic models. *arXiv preprint*.
- Dang, M., Vergari, A., and Van den Broeck, G. (2020). Strudel: Learning structured-decomposable probabilistic circuits. In Jaeger, M. and Nielsen, T. D., editors, *Proceedings of the 10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, pages 137–148. PMLR.
- Darwiche, A. (2001). On the tractable counting of theory models and its application to truth maintenance and belief revision. *Journal of Applied Non-Classical Logics*, 11(1-2):11–34.
- Darwiche, A. (2003). A differential approach to inference in bayesian networks. *Journal of the ACM (JACM)*, 50(3):280–305.
- Darwiche, A. (2011). Sdd: A new canonical representation of propositional knowledge bases. In *IJCAI*.
- Darwiche, A. (2021). Causal inference using tractable circuits. In *NeurIPS Workshop on Causal Inference and Machine Learning: Why Now? (WHY21)*.
- Darwiche, A. and Marquis, P. (2002). A knowledge compilation map. *J. Artif. Int. Res.*, 17(1):229–264.
- Di Mauro, N., Gala, G., Iannotta, M., and Basile, T. M. (2021). Random probabilistic circuits. In de Campos, C. and Maathuis, M. H., editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1682–1691. PMLR.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. volume 27.
- Huang, J., Chavira, M., and Darwiche, A. (2006). Solving map exactly by searching on compiled arithmetic circuits. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2, AAAI’06*, page 1143–1148. AAAI Press.
- Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Kisa, D., Van den Broeck, G., Choi, A., and Darwiche, A. (2014). Probabilistic sentential decision diagrams. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.
- Liang, Y., Bekker, J., and Van den Broeck, G. (2017). Learning the structure of probabilistic sentential decision diagrams. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Oztok, U., Choi, A., and Darwiche, A. (2016). Solving pppp-complete problems using knowledge compilation. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning, KR’16*, page 94–103. AAAI Press.
- Papantonis, I. and Belle, V. (2020). Interventions and counterfactuals in tractable probabilistic models: Limitations of contemporary transformations. *arXiv preprint*, abs/2001.10905.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.
- Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, UK, second edition.
- Peharz, R. (2015). *Foundations of sum-product networks for probabilistic modeling*. PhD thesis, Medical University of Graz.
- Peharz, R., Lang, S., Vergari, A., Stelzner, K., Molina, A., Trapp, M., Van den Broeck, G., Kersting, K., and Ghahramani, Z. (2020a). Einsum networks: Fast and scalable learning of tractable probabilistic circuits. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*.
- Peharz, R., Vergari, A., Stelzner, K., Molina, A., Shao, X., Trapp, M., Kersting, K., and Ghahramani, Z. (2020b). Random sum-product networks: A simple and effective approach to probabilistic deep learning. In Adams, R. P. and Gogate, V., editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115

- of *Proceedings of Machine Learning Research*, pages 334–344. PMLR.
- Pipatsrisawat, K. and Darwiche, A. (2008). New compilation languages based on structured decomposability. In *AAAI*, pages 517–522. AAAI Press.
- Pipatsrisawat, T. and Darwiche, A. (2010). A lower bound on the size of decomposable negation normal form. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Rezende, D. and Mohamed, S. (2015). Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR.
- Scutari, M. (2022). Bayesian network repository. <https://www.bnlearn.com/bnrepository/>.
- Shen, Y., Choi, A., and Darwiche, A. (2016). Tractable operations for arithmetic circuits of probabilistic models. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29.
- Shih, A., Van den Broeck, G., Beame, P., and Amarilli, A. (2019). Smoothing structured decomposable circuits. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32.
- Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-markovian causal models. *AAAI'06*, page 1219–1226. AAAI Press.
- Vergari, A., Choi, Y., Liu, A., Teso, S., and Van den Broeck, G. (2021). A compositional atlas of tractable circuit operations for probabilistic inference. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 13189–13201.
- Wang, B., Lyle, C., and Kwiatkowska, M. (2021). Provable guarantees on the robustness of decision rules to causal interventions. In *In Proc. International Joint Conference on Artificial Intelligence (IJCAI-21)*.
- Zecevic, M., Dhami, D. S., Karanam, A., Natarajan, S., and Kersting, K. (2021). Interventional sum-product networks: Causal inference with tractable probabilistic models. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*.
- Zhao, H., Melibari, M., and Poupart, P. (2015). On the relationship between sum-product networks and bayesian networks. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 116–124, Lille, France. PMLR.

A Md-vtree Proofs

In this section, we provide proofs of the results in Section 3 regarding md-vtrees.

A.1 Q-determinism Results

We begin with the results regarding (structured) marginal determinism.

Proposition 1 (Conflicting Q-Determinisms for Sum Nodes). *Let C be a PC, and let $Q, Q' \subseteq V$ such that neither is a subset of the other. Suppose that there exists a non-trivial sum node T in C that is Q-deterministic and Q'-deterministic, but not $(Q \cap Q')$ -deterministic. Then the circuit rooted at T , C_T , cannot have full support.*

Proof. Since the sum node T is non-trivial, it has at least two children. Let N_1, N_2 be two distinct children of T . Let I_1, I_2 be the sets of values q of Q such that $p_{N_1}(q) > 0, p_{N_2}(q) > 0$ respectively. Define I'_1, I'_2 similarly for Q' . Note that, by Q-determinism and Q'-determinism, I_1, I_2 are disjoint, and similarly I'_1, I'_2 are disjoint.

Now, we claim that there exists values $q \in I_1$ and $q' \in I'_2$ such that they agree over the intersection $Q \cap Q'$. If not, then p_{N_1} and p_{N_2} are non-zero for disjoint subsets of values of $(Q \cap Q')$, which implies $(Q \cap Q')$ -determinism, which is a contradiction of the assumption of the Proposition. Now consider the value $q \cup q'$ of $Q \cup Q'$. $p_{N_1}(q \cup q') = 0$ since $p_{N_1}(q') = 0$ (by the disjointness of I'_1, I'_2), and similarly $p_{N_2}(q \cup q') = 0$ since $p_{N_2}(q) = 0$. For any other child N_3 of T , we have that I_1 and I_3 are disjoint by Q-determinism, so $p_{N_3}(q) = 0$ and we get $p_{N_3}(q \cup q') = 0$. Putting it all together, $p_T(q \cup q') = 0$ and thus the circuit C_T does not have full support. \square

An important corollary of this result is that any circuit that is Q-deterministic and Q'-deterministic cannot have full support, as the root sum node R of the circuit must be Q-deterministic and Q'-deterministic.

Proposition 2 (Superset Q-Determinisms for Sum Nodes). *Suppose that a sum node T is Q-deterministic. Then it is also Q'-deterministic for any $Q \subseteq Q' \subseteq V$.*

Proof. By definition, a sum node T is Q-deterministic if for any instantiation q of Q , at most one of its children N_i evaluate to a nonzero output under q . If $Q' \supseteq Q$, then any instantiation q' of Q' will imply a specific instantiation of q , and so at most one of the children of T evaluate to a nonzero output under q' . More formally, $p_{N_i}(q) = \sum_{Q' \setminus Q} p_{N_i}(q, Q' \setminus Q) = 0 \implies p_{N_i}(q') = 0$. \square

Proposition 3 (Validity of Implied Q-Determinisms). *For any PC C respecting md-vtree w , both over V , and any $Q \subseteq V$ s.t. w implies Q-determinism, it follows that C is Q-deterministic.*

Proof. Since C respects w , every sum unit $T \in C$ has scope $\phi(T) = \phi(m)$ for some $m \in M$, and T is $\psi(m)$ -deterministic. Further, since w implies Q-determinism, we have that $\phi(m) \cap Q = \emptyset$, or else $Q \supseteq \psi(m)$. Combining these statements, we see that for all sum units $T \in C$, either $\phi(T) \cap Q = \emptyset$, or else T is $\psi(m)$ -deterministic and thus (by Proposition 2) Q-deterministic. This shows that C is Q-deterministic. \square

The following theorem justifies the intuition that having smaller labels $\psi(m)$ corresponds to a stronger restriction on the circuit, such that less circuits respect the md-vtree, but more marginal determinisms are implied:

Theorem 1 (Generality-Tractability Tradeoff). *Let $w = (v, \psi)$ and $w' = (v, \psi')$ be two md-vtrees, such that $\psi'(m) \supseteq \psi(m)$ for all $m \in M$. Then we have that $\mathcal{Q}(w) \supseteq \mathcal{Q}(w')$, and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$.*

Proof. For the first part, suppose $Q \in \mathcal{Q}(w')$. Then for all $m \in M$, it holds that $\phi(m) \cap Q = \emptyset$, or else $Q \supseteq \psi'(m)$. Since $\psi'(m) \supseteq \psi(m)$, it holds that $\phi(m) \cap Q = \emptyset$, or else $Q \supseteq \psi(m)$ also, so $Q \in \mathcal{Q}(w)$. This shows that $\mathcal{Q}(w) \supseteq \mathcal{Q}(w')$.

For the second part, suppose that $C \in \mathcal{C}_w$. Then, for any sum unit $T \in C$, there is an $m \in M$ that T is marginally deterministic w.r.t. $\psi(m)$. As $\psi'(m) \supseteq \psi(m)$, this means that T is also marginally deterministic w.r.t. $\psi'(m)$. Thus C respects w' also, i.e. $C \in \mathcal{C}_{w'}$. \square

A.2 Regular and Optimal md-vtrees

We now move to the results regarding regular and optimal md-vtrees. Recall that Algorithm 1 is designed to return a labelling of a given vtree that satisfies some desired set of marginal determinisms. We begin with a formal correctness proof of Algorithm 1, i.e. that it does indeed return an md-vtree which implies the given set \mathcal{S} of marginal determinisms:

Proposition 5 (Correctness of Algorithm 1). *For any input vtree $v = (M, E)$ and set of marginal determinisms \mathcal{S} , Algorithm 1 returns an md-vtree that implies \mathcal{Q} -determinism for all $\mathcal{Q} \in \mathcal{S}$.*

Proof. We need to show that, for each $m \in M$ and $\mathcal{Q} \in \mathcal{S}$, if $\phi(m) \cap \mathcal{Q} \neq \emptyset$, then $\mathcal{Q} \supseteq \psi(m)$.

This can be seen by inspection of the Algorithm. For every $m \in M$, and every $\mathcal{Q} \in \mathcal{S}$, if $\phi(m) \cap \mathcal{Q} \neq \emptyset$, then in line 5 of the Algorithm, we intersect the label with the \mathcal{Q} . At each iteration of the loop in lines 3-5, no elements can be added to $\psi(m)$ as we are taking an intersection, and in lines 6-7 we also take an intersection (with the scope). Thus, it follows that for all $\mathcal{Q} \in \mathcal{S}$ such that $\phi(m) \cap \mathcal{Q} \neq \emptyset$, the label at the end of the outer loop will satisfy $\mathcal{Q} \supseteq \psi(m)$. \square

Now, we move on to properties of Algorithm 1. We begin by showing that the output md-vtree is regular; that is, the labelling function satisfies a certain structure as defined in Definition 13 and reproduced below.

Definition 13 (Regular md-vtree). *We say that a md-vtree $w = (v, \psi)$ is regular if for every non-leaf node m , and its children m_1, m_2 , it holds that either $\psi(m) = \psi(m_1)$, $\psi(m) = \psi(m_2)$, or $\psi(m) = \psi(m_1) \cup \psi(m_2)$.*

Proposition 4 (Regularity of Algorithm 1). *The output of Algorithm 1 is a regular md-vtree.*

Proof. By examining the Algorithm, we can see that the label of each vtree node node is given by:

$$\psi(m) = \begin{cases} U & \text{if } \mathcal{Q} \cap \phi(m) = \emptyset \quad \forall \mathcal{Q} \in \mathcal{S} \\ \phi(m) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}: \mathcal{Q} \cap \phi(m) \neq \emptyset} \mathcal{Q}) & \text{otherwise} \end{cases} \quad (1)$$

For any non-leaf vtree-node m , let its children be m_1, m_2 , and note that by definition, $\phi(m) = \phi(m_1) \cup \phi(m_2)$ with $\phi(m_1), \phi(m_2)$ disjoint. Firstly, if $\psi(m) = U$, this means that no $\mathcal{Q} \in \mathcal{S}$ has non-empty intersection with $\phi(m)$, and since the scopes of the children are subsets of $\phi(m)$, we must also have $\psi(m_1) = \psi(m_2) = U$ and the regularity condition is satisfied.

Otherwise, we have that $\psi(m) = \phi(m) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q})$, where we have written $\mathcal{S}_m := \{\mathcal{Q} \in \mathcal{S} : \mathcal{Q} \cap \phi(m) \neq \emptyset\}$ to denote the subset of \mathcal{S} that has non-empty intersection of $\phi(m)$. Note that \mathcal{S}_m is non-empty. Since $\phi(m) = \phi(m_1) \cup \phi(m_2)$, it must be the case that every $\mathcal{Q} \in \mathcal{S}_m$ also intersects with at least one of $\phi(m_1), \phi(m_2)$. More formally, $\mathcal{S}_m = \mathcal{S}_{m_1} \cup \mathcal{S}_{m_2}$. We consider three cases separately:

- If \mathcal{S}_{m_2} is empty, then we must have $\mathcal{S}_m = \mathcal{S}_{m_1}$. Thus \mathcal{S}_{m_1} is non-empty, meaning that $\psi(m_1) = \phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_1}} \mathcal{Q})$ in the definition. We can then derive that $\psi(m_1) = \phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_1}} \mathcal{Q}) = \phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q}) = \phi(m) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q}) = \psi(m)$, where the second-to-last equality follows since all $\mathcal{Q} \in \mathcal{S}_m$ have empty intersection with $\phi(m_2)$. This satisfies the regularity condition.
- If \mathcal{S}_{m_1} is empty, by similar reasoning we have that $\psi(m_2) = \psi(m)$.
- If neither $\mathcal{S}_{m_1}, \mathcal{S}_{m_2}$ is empty, then we have that $\psi(m) = \phi(m) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q}) = (\phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q})) \cup (\phi(m_2) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_m} \mathcal{Q})) = (\phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_1}} \mathcal{Q})) \cup (\phi(m_2) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_2}} \mathcal{Q}))$. Since both $\mathcal{S}_{m_1}, \mathcal{S}_{m_2}$ are non-empty, we have that $\psi(m_1) = \phi(m_1) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_1}} \mathcal{Q})$ and $\psi(m_2) = \phi(m_2) \cap (\bigcap_{\mathcal{Q} \in \mathcal{S}_{m_2}} \mathcal{Q})$ in the definition. Combining, we have that $\psi(m) = \psi(m_1) \cup \psi(m_2)$, which satisfies the regularity condition.

Thus we have shown that the output md-vtree is regular. \square

In the following theorem, we show that the significance of regularity, in that regular md-vtrees are ‘‘optimally expressive’’ among all md-vtrees with the same vtree.

Theorem 2 (Optimal md-vtrees). *Let $w = (v, \psi)$ be any md-vtree. Then there exists a regular md-vtree $w' = (v, \psi')$ such that $\mathcal{Q}(w) = \mathcal{Q}(w')$, and $\mathcal{C}_{w'} \supseteq \mathcal{C}_w$.*

We will prove this Theorem explicitly by constructing a regular md-vtree with these properties. To do this, we prove two Lemmas which define operations which do not change $\mathcal{Q}(w)$, while keeping the same or increasing the set \mathcal{C}_w ; the result of iterative application of the two operations being a regular md-vtree.

Definition 14 (Expand Child Labels). *Given a md-vtree $w = (v, \psi)$, and any vtree nodes m_{pa^*}, m_{ch^*} such that m_{ch^*} is a child of m_{pa^*} , the operation $ECL(w, m_{pa^*}, m_{ch^*})$ returns a new md-vtree $w' = (v, \psi')$, defined as follows:*

$$\psi'(m) = \begin{cases} \psi(m_{ch^*}) \cup (\psi(m_{pa^*}) \cap \phi(m_{ch^*})) & \text{if } m = m_{ch} \\ \psi(m) & \text{otherwise} \end{cases} \quad (2)$$

Lemma 1. *The output $w' = ECL(w, m_{pa^*}, m_{ch^*})$ satisfies $\mathcal{Q}(w') = \mathcal{Q}(w)$, and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$.*

Proof. The only difference between w and w' is the label of m_{ch^*} . Suppose that $\mathbf{Q} \in \mathcal{Q}(w)$, then we have that either $\phi(m_{ch^*}) \cap \mathbf{Q} = \emptyset$, or else $\mathbf{Q} \supseteq \psi(m_{ch^*})$. In the former case, since the vtrees and thus scopes are the same between w, w' , it follows that $\mathbf{Q} \in \mathcal{Q}(w')$ also. In the latter case, since the scope of the parent $\phi(m_{pa^*}) \supseteq \phi(m_{ch^*})$, \mathbf{Q} overlaps with the parent scope as well, implying that $\mathbf{Q} \supseteq \psi(m_{pa^*})$. Thus we have that $\mathbf{Q} \supseteq \psi(m_{ch^*}) \cup \psi(m_{pa^*}) \supseteq \psi(m_{ch^*}) \cup (\psi(m_{pa^*}) \cap \phi(m_{ch^*})) = \psi'(m_{ch^*})$. Thus, $\mathbf{Q} \in \mathcal{Q}(w')$ also. That is, $\mathcal{Q}(w) \subseteq \mathcal{Q}(w')$.

To complete the result, note that $\psi(m) \subseteq \psi'(m)$ for all vtree nodes m . Thus by Theorem 1, it follows that $\mathcal{Q}(w) \supseteq \mathcal{Q}(w')$ and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$. Combining with the paragraph above we have shown that $\mathcal{Q}(w) = \mathcal{Q}(w')$. \square

Intuitively, this operation "pushes down" elements of $\psi(m_{pa^*})$ to its children. If we apply this operation to all pairs of parent/child vtree nodes (m_{pa}, m_{ch}) , then it can be seen that the new labels will have the property that *all elements of the parent label that are contained in the scope of a child, will be in the label of that child*. More formally, $\psi'(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \psi'(m_{ch})$. This is the starting point for the next operation:

Definition 15 (Expand Parent Labels). *Let $w = (v, \psi)$ be a md-vtree such that $\psi(m_{pa}) \cap \phi(m_{ch}) = \psi(m_{pa}) \cap \psi(m_{ch})$ holds for all pairs of parents m_{pa} and children m_{ch} . Then, given any non-leaf vtree node m_{pa^*} , the operation $EPL(w, m_{pa^*})$ returns a new md-vtree $w' = (v, \psi')$, defined as follows:*

$$\psi'(m) = \begin{cases} \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*}) & \text{if } m = m_{pa} \\ \psi(m) & \text{otherwise} \end{cases} \quad (3)$$

where we define $M_{active} = \{m_{ch^*} \mid m_{ch^*} \in M_{ch^*}, \psi(m_{ch^*}) \cap \psi(m_{pa^*}) \neq \emptyset\}$ to be the set of all children whose labellings have non-empty intersection with the labelling of the parent.

Lemma 2. *The output $w' = EPL(w, m_{pa^*})$ satisfies $\mathcal{Q}(w') = \mathcal{Q}(w)$, and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$.*

Further, the property that $\psi'(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \psi'(m_{ch})$ holds for all pairs of parents m_{pa} and children m_{ch} in w' (i.e. is maintained in w').

Proof. Firstly, we show that $\psi(m_{pa^*}) \subseteq \psi'(m_{pa^*})$. This follows by taking a union over children of both sides of the assumption $\psi(m_{pa^*}) \cap \phi(m_{ch^*}) = \psi(m_{pa^*}) \cap \psi(m_{ch^*})$, where the LHS becomes $\bigcup_{m_{ch^*} \in M_{ch^*}} (\psi(m_{pa^*}) \cap \phi(m_{ch^*})) = \psi(m_{pa^*}) \cap \bigcup_{m_{ch^*} \in M_{ch^*}} \phi(m_{ch^*}) = \psi(m_{pa^*}) \cap \phi(m_{pa^*}) = \psi(m_{pa^*})$, and the RHS becomes $\bigcup_{m_{ch^*} \in M_{ch^*}} (\psi(m_{pa^*}) \cap \psi(m_{ch^*})) = \bigcup_{m_{ch^*} \in M_{active}} (\psi(m_{pa^*}) \cap \psi(m_{ch^*})) = \psi(m_{pa^*}) \cap \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*}) \subseteq \psi'(m_{pa^*})$. Thus $\psi(m) \subseteq \psi'(m)$ for all vtree nodes m , and by Theorem 1, it follows that $\mathcal{Q}(w) \supseteq \mathcal{Q}(w')$ and $\mathcal{C}_w \subseteq \mathcal{C}_{w'}$.

Now suppose $\mathbf{Q} \in \mathcal{Q}(w)$. We consider two cases. Firstly, if $\phi(m_{pa^*}) \cap \mathbf{Q} = \emptyset$, then since vtrees and scopes are the same between w, w' , we have $\mathbf{Q} \in \mathcal{Q}(w')$. Otherwise, we have $\phi(m_{pa^*}) \cap \mathbf{Q} \neq \emptyset$ and $\mathbf{Q} \supseteq \psi(m_{pa^*})$. Now, for those children in M_{active} , we have that $\psi(m_{ch^*}) \cap \psi(m_{pa^*}) \neq \emptyset$ and so since $\mathbf{Q} \supseteq \psi(m_{pa^*})$ and $\phi(m_{ch^*}) \supseteq \psi(m_{ch^*})$, we have $\mathbf{Q} \cap \phi(m_{ch^*}) \neq \emptyset$. The marginal determinism property on these children then implies that $\mathbf{Q} \supseteq \psi(m_{ch^*})$ for all $m_{ch^*} \in M_{active}$; and so $\mathbf{Q} \supseteq \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*}) = \psi'(m_{pa^*})$. This shows that $\mathbf{Q} \in \mathcal{Q}(w')$ also. This gives $\mathcal{Q}(w) \subseteq \mathcal{Q}(w')$, and combined with the previous result, $\mathcal{Q}(w) = \mathcal{Q}(w')$.

Finally, we show that the property that $\psi'(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \psi'(m_{ch})$ holds for all pairs of parents m_{pa} and children m_{ch} in w' . The only label which has changed is that of m_{pa} , so we need only consider the pairs (m_{pa}, m_{ch}) with either (a) $m_{pa} = m_{pa^*}$ and m_{ch} is a child of m_{pa^*} or (b) $m_{ch} = m_{pa^*}$ and m_{pa} is the parent of m_{pa^*} .

Operation	Requirement	Input Condition	Output Condition
MARG($C; W$)	-	Q -det	Q -det
INST($C; w$)	-	$\exists W' \subseteq W : (Q \cup W')$ -det	Q -det
PROD($C^{(1)}, C^{(2)}$)	$C^{(1)}, C^{(2)}$ respect compatible vtrees	$\exists Q^{(1)}, Q^{(2)} : Q^{(1)}$ -det, $Q^{(2)}$ -det, and: • Either (a) $Q \subseteq V^{(1)} \cap V^{(2)}$ and $Q^{(1)} = Q^{(2)} = Q$; • Or (b) $Q^{(1)}, Q^{(2)} \supseteq V^{(1)} \cap V^{(2)}$ and $Q = Q^{(1)} \cup Q^{(2)}$	Q -det
POW($C; \alpha$)	Det	Q -det	Q -det
MAX(C)	Det	N/A	N/A (scalar output)
LOG(C)	Det	-	-

Table 5: MD-calculus: sufficient input-output conditions for each basic operation

- In case (a), by definition we have that $\psi'(m_{pa}) = \psi'(m_{pa^*}) = \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*})$, and $\psi'(m_{ch}) = \psi'(m_{ch})$. If m_{ch} is an active child of m_{pa^*} , then we have that the LHS of the property $\psi'(m_{pa}) \cap \phi(m_{ch}) = \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*}) \cap \phi(m_{ch}) = \psi(m_{ch})$, and the RHS of the property $\psi'(m_{pa}) \cap \psi'(m_{ch}) = \bigcup_{m_{ch^*} \in M_{active}} \psi(m_{ch^*}) \cap \psi'(m_{ch}) = \psi(m_{ch})$. If m_{ch} is not an active child of m_{pa^*} , then both sides of the property correspond to the empty set.
- In case (b), by definition we have $\psi'(m_{pa}) = \psi(m_{pa})$, and $\psi'(m_{ch}) = \psi'(m_{pa^*})$. We have shown above that $\psi(m_{pa^*}) \subseteq \psi'(m_{pa^*})$, so $\psi(m_{ch}) \subseteq \psi'(m_{ch})$. By the precondition for applying the EPL operation, we have that $\psi(m_{pa}) \cap \phi(m_{ch}) = \psi(m_{pa}) \cap \psi(m_{ch})$. Substituting, we get $\psi'(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \psi(m_{ch}) \subseteq \psi'(m_{pa}) \cap \psi'(m_{ch})$. The other direction $\psi'(m_{pa}) \cap \phi(m_{ch}) \supseteq \psi'(m_{pa}) \cap \psi'(m_{ch})$ is immediate as the label of a node is contained in its scope.

Thus, we have shown that $\psi'(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \psi'(m_{ch})$ holds for all pairs of parents m_{pa} and children m_{ch} in w' , concluding the proof. \square

Intuitively, this operation "pulls up" elements $\psi(m_{ch})$ of the *active* children to the parent. After applying this operation to all nodes, we obtain a regular md-vtree, which has the same set Q of marginal determinisms, and is at least as expressive. More formally:

Proof. (of Theorem) Starting from w , apply the ECL operation to each pair of parent and child nodes, in a topological order starting from the root. For each m_{pa}, m_{ch} pair, we have that $\psi'(m_{pa}) = \psi(m_{pa})$ and $\psi'(m_{ch}) = \psi(m_{ch}) \cup (\psi(m_{pa}) \cap \phi(m_{ch}))$ by definition of the operation. Then $\psi'(m_{pa}) \cap \psi'(m_{ch}) = \psi(m_{pa}) \cap (\psi(m_{ch}) \cup (\psi(m_{pa}) \cap \phi(m_{ch}))) = (\psi(m_{pa}) \cap \psi(m_{ch})) \cup (\psi(m_{pa}) \cap \phi(m_{ch})) = \psi(m_{pa}) \cap \phi(m_{ch}) = \psi'(m_{pa}) \cap \phi(m_{ch})$, which is the required property for applying the EPL operation. As we proceed in a topological order, and the operation only modifies the label of the child, it follows that the property $\psi'(m_{pa}) \cap \psi'(m_{ch}) = \psi'(m_{pa}) \cap \phi(m_{ch})$ holds for all parent/children pairs at the end.

This allows us to apply the EPL operation. We apply this operation to every non-leaf node, in a reverse topological order from the leaves to the root. The precondition for applying the operation holds at all points due to the result of Lemma 2. This operation only modifies the label of the parent, and so after we have modified all the labels, we have the property that $\psi'(m_{pa}) = \bigcup_{m_{ch} \in M_{active}} \psi(m_{ch})$ for every non-leaf node m_{pa} . That is, it satisfies the conditions to be a regular md-vtree, i.e. $\psi'(m_{pa}) = \psi'(m_1), \psi'(m_2),$ or $\psi'(m_1) \cup \psi'(m_2)$, where m_1, m_2 are the children of m_{pa} . \square

B Operations and MD-Calculus

In this section, we provide further details on the results in Section 5 regarding inference on md-vtrees. First, for the *forward* problem, we describe algorithms for soundly propagating the md-vtree forward under each basic circuit operation, as mentioned in Section 5.2. Then, for the *backward* problem, we analyze these algorithms to prove the MD-calculus for propagating marginal determinisms backwards through operations.

B.1 Algorithms and the Forward Problem

For each of the basic operations, there exist efficient (polynomial time) algorithms for computing them on probabilistic circuits satisfying the requirement column in Table 5 (Choi et al., 2020; Vergari et al., 2021). In this section, we will also describe, for each basic operation, an algorithm for computing the operation on md-vtrees w , that is a sound abstraction of the corresponding algorithm on circuits. By *sound*, we mean that, given an input md-vtree w and the output of the md-vtree algorithm w' , it is guaranteed for any input PC respecting w , the output of the corresponding PC algorithm will respect w' .

The construction of these md-vtrees algorithms is based upon the corresponding PC algorithm. Thus, we present the algorithms as applying to both the md-vtree and PC. For convenience, we assume that the PC satisfies the following condition, which we call *exactly respecting* a md-vtree:

Definition 16 (PC exactly respecting md-vtree). *A PC C exactly respects a md-vtree w if (1) it respects w and (2) the children of any sum node T corresponding to a non-leaf vtree node m are all product nodes P , where P has two children which are sum nodes, each corresponding to a child of m .*

PC architectures are typically designed with these alternating sum and product nodes, where the product nodes are binary; for example, both MDNets and PSSDs satisfy this property. Further, any PC which respects a md-vtree can be transformed into an equivalent PC which exactly respects the md-vtree, as follows. For every sum node T which has a sum node child T' , we can directly attach the children of the T' to T (with the appropriate combination of weights). Then, for every product node P which has a product node child P' , we can replace P' with a new single-child sum node T' , which has P' as its child. The resulting circuit still encodes the same function, and has the same marginal determinisms as the original circuit.

Given that a PC exactly respects a md-vtree, for each non-leaf vtree node m , we can represent the corresponding PC layer simply as a vector of sum nodes \mathbf{T}_m with length $K_m := |\mathbf{T}_m|$, and a weight/parameter matrix θ_m with shape (K_m, K_{m_1}, K_{m_2}) , with the semantics that $p_{T_{m,i}}(\mathbf{V}) = \sum_{jk} \theta_{m,ijk} p_{T_{m_1,j}}(\mathbf{V}) p_{T_{m_2,k}}(\mathbf{V})$ (where m_1, m_2 are the children of m). Note that for any pair of sum nodes $T_{m_1,j}, T_{m_2,k}$ for which there isn't a product in the PC connected to $T_{m,i}$, we can simply set the weight $\theta_{m,ijk}$ to zero⁷. For leaf vtree nodes, the corresponding layer can consist of both sum and leaf nodes (with the sum nodes being mixtures over the leaf nodes, e.g. $0.7\mathbb{1}_{X=0} + 0.3\mathbb{1}_{X=1}$). In this case, we represent the sum and leaf nodes as a vector \mathbf{T}_m with length $K_m := |\mathbf{T}_m|$, and θ_m , and a weight matrix θ_m , with $\theta_{m,ij} > 0$ iff $T_{m,j}$ is a leaf node that is a child of sum node $T_{m,i}$.

This characterization of a PC as a pair $\rho(m) := (\mathbf{T}_m, \theta_m)$ for each vtree node, which we call the *parameter function*, allows us to efficiently describe algorithms for the basic operations. Thus, in the algorithms below we will represent C exactly respecting some md-vtree using the triple $C = (v, \psi, \rho)$, where v is the vtree, ψ is the labelling function, and ρ the parameter function.

MARG($\cdot; W$) The marginalization algorithm is depicted in Algorithm 2. For the marginalization operation, we can take advantage of the fact that marginalization commutes with both product and sum nodes in a decomposable and smooth PC (which is the basis of tractable marginal inference).

$$\sum_{\mathbf{W}} p_P(\phi(P)) = \sum_{\mathbf{W}} p_{N_1}(\phi(N_1)) p_{N_2}(\phi(N_2)) = \left(\sum_{\mathbf{W}} p_{N_1}(\phi(N_1)) \right) \left(\sum_{\mathbf{W}} p_{N_2}(\phi(N_2)) \right) \quad (4)$$

$$\sum_{\mathbf{W}} p_T(\phi(T)) = \sum_{\mathbf{W}} \sum_{N_i \in \text{ch}(T)} \theta_i p_{N_i}(\phi(N_i)) = \sum_{N_i \in \text{ch}(T)} \theta_i \left(\sum_{\mathbf{W}} p_{N_i}(\phi(N_i)) \right) \quad (5)$$

where the last equality on the first line holds because $\phi(N_1) \cap \phi(N_2) = \emptyset$ by decomposability. This means that, in order to marginalize a circuit, we simply need to marginalize the leaf nodes. In Algorithm 2, we show the (recursive) procedure of marginalizing a circuit represented as (v, ψ, ρ) . In lines 3-5 we marginalize out \mathbf{W} from the leaf nodes in the PC, and in lines 6-12, we handle non-leaf vtree nodes simply by copying the existing circuit. To update the md-vtree, in line 13, we update the scope of the vtree node, removing the marginalized variables, and in lines 14-17, we assign a label to the new vtree node m' .

⁷While this is sufficient to represent any PC exactly respecting an md-vtree, it may be inefficient to represent $\theta_{m,ijk}$ as a dense matrix if the connections in the PC are sparse, i.e. $\theta_{m,ijk} = 0$ for many i, j, k . In the evaluation of a sum vector as a function of its child sum vectors, we only require the sum $\sum_{jk} \theta_{m,ijk} p_{T_{m_1,j}} p_{T_{m_2,k}}$ to be computed, so this can be implemented in a sparse manner if that is more appropriate. Similar reasoning applies to the product algorithm.

Algorithm 2: MARG(C, \mathbf{W})

Input: Input circuit $C = (v = (M, E, \phi), \psi, \rho)$; set of variables to be marginalized \mathbf{W}
Result: Output circuit $C' = (v', \psi', \rho')$

```

1  $m \leftarrow \text{root}(v)$ ;
2  $m' \leftarrow \text{newnode}()$ ;
3 if  $m$  is leaf then // Update vtree structure and parameter function (leaf)
4    $v' \leftarrow \text{createvtree}(m')$ ; // create vtree with single node
5    $\rho'(m') \leftarrow (\text{MARG}(L; \mathbf{W}) \text{ for } L \in T_m, \theta_m)$ ; // marginalize leaf PC nodes
6 else // Update vtree structure and parameter function (non-leaf)
7    $m_l, m_r \leftarrow \text{children}(m)$ ;
8    $v'_l, \psi'_l, \rho'_l \leftarrow \text{MARG}((v_{m_l}, \psi, \rho), \mathbf{W})$ ;
9    $v'_r, \psi'_r, \rho'_r \leftarrow \text{MARG}((v_{m_r}, \psi, \rho), \mathbf{W})$ ;
10   $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r$ ; // combine the vtrees/labelling fn/param fn
11   $v' \leftarrow \text{addnode}(v'; m')$ ;  $v' \leftarrow \text{addchildren}(v'; m', \text{root}(v'_l), \text{root}(v'_r))$ ;
12   $\rho'(m') \leftarrow \rho(m)$ ;
13  $\phi'(m') \leftarrow \phi(m) \setminus \mathbf{W}$ ; // Update scope function
14 if  $\psi(m) \cap \mathbf{W} = \emptyset$  then // Update labelling function
15    $\psi'(m') \leftarrow \psi(m)$ ;
16 else
17    $\psi'(m') \leftarrow U$ ;
18 Return  $(v', \psi', \rho')$ 

```

The new label is justified as follows. Suppose we have a sum node $T \in T_m$, with children N_1, \dots, N_n ; by definition, T is marginally deterministic with respect to $\psi(m)$. After marginalization, the function encoded by each child N'_i satisfies $p_{N'_i}(\mathbf{q}) = \sum_{\mathbf{W}} p_{N_i}(\mathbf{q})$ for any value \mathbf{q} of $\psi(m)$ by definition. Now:

- If $\psi(m) \cap \mathbf{W} = \emptyset$, then this is just proportional to $p_{N_i}(\mathbf{q})$ and so the marginalized support will remain the same for each child, and T' will maintain $\psi(m)$ -determinism.
- On the other hand, if $\psi(m) \cap \mathbf{W} \neq \emptyset$, then we do not have any such guarantee; in fact, we cannot be sure that T' will be Q -deterministic for any Q , so we assign the universal set.

INST($\cdot; \mathbf{w}$) For the instantiation operation, we have Algorithm 3. At first glance, this seems to be very similar to the marginalization operation; it changes the scope in the same way, and the changes to the circuit can be implemented through the leaf nodes. However, the crucial difference is in the label function.

The new label of $\psi'(m) = \psi(m) \setminus \mathbf{W}$ is justified as follows. Suppose that we have a sum node $T \in T_m$, with children N_1, \dots, N_n , with T marginally deterministic with respect to $\psi(m)$. After instantiation (of \mathbf{W} with the value \mathbf{w}), the function encoded by each child N'_i satisfies $p_{N'_i}(\mathbf{q} \setminus \mathbf{W}) = p_{N_i}(\mathbf{w}, \mathbf{q} \setminus \mathbf{W})$, for any value \mathbf{q} of $\psi(m)$ by definition⁸.

Now, we claim that N'_i is $(\psi(m) \setminus \mathbf{W})$ -deterministic, i.e. N'_i, N'_j have distinct marginalized support $\text{supp}_{\psi(m) \setminus \mathbf{W}}(N'_i)$, $\text{supp}_{\psi(m) \setminus \mathbf{W}}(N'_j)$ for $i \neq j$. Suppose for contradiction there exists a value $\mathbf{q}^* \setminus \mathbf{W}$ of $(\psi(m) \setminus \mathbf{W})$ such that $p_{N'_i}(\mathbf{q}^* \setminus \mathbf{W}) > 0$ and $p_{N'_j}(\mathbf{q}^* \setminus \mathbf{W}) > 0$. Then we have

$$p_{N'_i}(\mathbf{q}^* \setminus \mathbf{W}) > 0 \text{ and } p_{N'_j}(\mathbf{q}^* \setminus \mathbf{W}) > 0 \quad (6)$$

$$p_{N_i}(\mathbf{w}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \text{ and } p_{N_j}(\mathbf{w}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \quad (7)$$

$$\sum_{\mathbf{W} \setminus \mathbf{Q}} p_{N_i}(\mathbf{W} \setminus \mathbf{Q}, \mathbf{w} \cap \mathbf{Q}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \text{ and } \sum_{\mathbf{W} \setminus \mathbf{Q}} p_{N_j}(\mathbf{W} \setminus \mathbf{Q}, \mathbf{w} \cap \mathbf{Q}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \quad (8)$$

$$p_{N_i}(\mathbf{w} \cap \mathbf{Q}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \text{ and } p_{N_j}(\mathbf{w} \cap \mathbf{Q}, \mathbf{q}^* \setminus \mathbf{W}) > 0 \quad (9)$$

The second line follows by definition of the instantiation algorithm, the third line is a sum of non-negative terms including a positive term from the previous line (when $\mathbf{W} \setminus \mathbf{Q} = \mathbf{w} \setminus \mathbf{Q}$), and the fourth line rewrites the sum. Now we have

⁸Note that we write $\mathbf{q} \setminus \mathbf{W}$ to represent the value of $\psi(m) \setminus \mathbf{W}$ given by \mathbf{q} restricted to this variable set.

Algorithm 3: INST(C, \mathbf{w})

Input: Input circuit $C = (v = (M, E, \phi), \psi, \rho)$; instantiation \mathbf{w} of some subset of variables \mathbf{W}
Result: Output circuit $C' = (v', \psi', \rho')$

```

1  $m \leftarrow \text{root}(v)$ ;
2  $m' \leftarrow \text{newnode}()$ ;
3 if  $m$  is leaf then // Update vtree structure and parameter function (leaf)
4    $v' \leftarrow \text{createvtree}(m')$ ; // create vtree with single node
5    $\rho'(m') \leftarrow (\text{INST}(L; \mathbf{w}) \text{ for } L \in \mathbf{T}_m, \theta_m)$ ; // instantiate leaf PC nodes
6 else // Update vtree structure and parameter function (non-leaf)
7    $m_l, m_r \leftarrow \text{children}(m)$ ;
8    $v'_l, \psi'_l, \rho'_l \leftarrow \text{INST}((v_{m_l}, \psi, \rho), \mathbf{w})$ ;
9    $v'_r, \psi'_r, \rho'_r \leftarrow \text{INST}((v_{m_r}, \psi, \rho), \mathbf{w})$ ;
10   $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r$ ; // combine the vtrees/labelling fn/param fn
11   $v' \leftarrow \text{addnode}(v'; m')$ ;  $v' \leftarrow \text{addchildren}(v'; m', \text{root}(v'_l), \text{root}(v'_r))$ ;
12   $\rho'(m') \leftarrow \rho(m)$ ;
13  $\phi'(m') \leftarrow \phi(m) \setminus \mathbf{W}$ ; // Update vtree scope function
14  $\psi'(m') \leftarrow \psi(m) \setminus \mathbf{W}$ ; // Update labelling function
15 Return  $(v', \psi', \rho')$ 

```

a value $\mathbf{q} := (\mathbf{w} \cap \mathbf{Q}, \mathbf{q}^* \setminus \mathbf{W})$ of $\psi(m)$, such that $p_{N_i}(\mathbf{q}) > 0$ and $p_{N_j}(\mathbf{q}) > 0$, which is a contradiction as T is $\psi(m)$ -deterministic.

PROD(\cdot, \cdot) Now, we consider the product of two circuits exactly respecting *compatible* vtrees.

Definition 17 (Vtree Compatibility). Let $v^{(1)} = (M^{(1)}, E^{(1)}, \phi^{(1)})$ and $v^{(2)} = (M^{(2)}, E^{(2)}, \phi^{(2)})$ be two vtrees, with root nodes $m^{(1)}, m^{(2)}$ respectively. Define $\mathbf{C} := \phi^{(1)}(m^{(1)}) \cup \phi^{(2)}(m^{(2)})$ to be the common variables. Then we say that $v^{(1)}, v^{(2)}$ are compatible if any of the following hold:

1. There are no common variables, $\mathbf{C} = \emptyset$;
2. Both of $m^{(1)}, m^{(2)}$ are leaf vtree nodes;
3. One of the root nodes has the same restricted scope on \mathbf{C} as one of the children of the other root node, **and** the vtrees rooted at these nodes are compatible. For example, $\phi_{\mathbf{C}}^{(1)}(m^{(2)}) = \phi_{\mathbf{C}}^{(1)}(m_r^{(1)})$, and $v_{m^{(2)}}^{(2)}$ and $v_{m_r^{(1)}}^{(1)}$ are compatible.
4. The children of the root nodes have matching restricted scopes, **and** are compatible. For example, $\phi_{\mathbf{C}}^{(1)}(m_l^{(1)}) = \phi_{\mathbf{C}}^{(2)}(m_l^{(2)})$ and $v_{m_l^{(1)}}^{(1)}, v_{m_l^{(2)}}^{(2)}$ are compatible, and $\phi_{\mathbf{C}}^{(1)}(m_r^{(1)}) = \phi_{\mathbf{C}}^{(2)}(m_r^{(2)})$ and $v_{m_r^{(1)}}^{(1)}, v_{m_r^{(2)}}^{(2)}$ are compatible.

This recursive definition allows for products of circuits not necessarily respecting the same vtree, but merely vtrees which “essentially have the same structure” over the shared variables. Intuitively, there are four cases that allow us to maintain (structured) decomposability in the output circuit, illustrated in Figure 3. The first two are base cases where the product is directly tractable: namely, when the root vtree nodes have disjoint scopes, or when they are both leaves. Note, in particular, that the product of a leaf region and a non-leaf region that have overlapping variables is considered intractable here (unless condition 3. holds). The last two are cases where we can recursively call the product algorithm on the children of the root vtree nodes. Each of the four cases above correspond to a slightly different algorithm for computing the product of the corresponding PC sum nodes, which we will explain next.⁹

The product algorithm, depicted in Algorithm 4, (recursively) constructs a circuit $C = (v', \psi', \rho')$ that is the product of the input circuits $C^{(1)} = (v^{(1)}, \psi^{(1)}, \rho^{(1)})$ and $C^{(2)} = (v^{(2)}, \psi^{(2)}, \rho^{(2)})$ respectively. In particular, at each recursive step, it computes the root node m' of the new vtree, its label $\psi'(m')$, and the parameter function value $\rho'(m') = (\mathbf{T}_{m'}, \theta_{m'})$ of that node¹⁰. We consider each of the four compatibility cases separately:

⁹The notion of compatibility between vtrees is somewhat similar to the notion of compatibility between circuits (Vergari et al., 2021), but acts at the level of groups of circuit nodes with the same scope (i.e. a vtree node) rather than individual circuit nodes.

¹⁰Consider the root nodes of the input vtrees $m^{(1)}, m^{(2)}$, and their parameter function values $\rho^{(1)}(m^{(1)}) =$

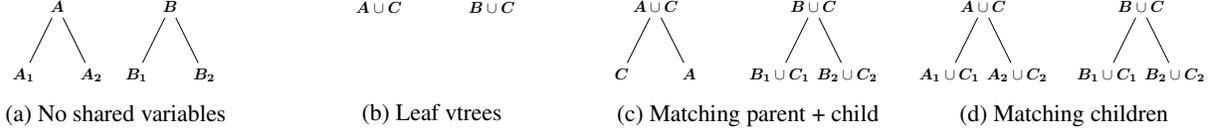


Figure 3: Examples of (possibly) compatible vtrees, where $A \cap B = \emptyset$, and C are the shared variables

1. Firstly, if there are no common variables, i.e. $C = \emptyset$, then we can simply introduce product nodes for each pair of sum nodes, while maintaining decomposability, as in Figure 4a.

- *Vtree node*: We create a vtree node m' with $(m'_l = m^{(1)}, m'_r = m^{(2)})$ as children.
- *Parameter function*: The parameter function value for the new node $\rho'(m') = (\mathbf{T}_{m'}, \theta_{m'})$ is given as follows. For each pair of sum nodes $T_{m^{(1)},j} \in \mathbf{T}_{m^{(1)}}, T_{m^{(2)},k} \in \mathbf{T}_{m^{(2)}}$, we create a sum node $T_{m',i^{(1)}i^{(2)}}$, representing the product of $T_{m^{(1)},j}, T_{m^{(2)},k}$. To achieve this, the weights $\theta_{m',i^{(1)}i^{(2)}jk}$ are defined to be 1 if $i^{(1)} = j$ and $i^{(2)} = k$, and 0 otherwise. Note that j, k only have a single index as the children m'_l, m'_r correspond to the input circuits, which only have a single dimension.
- *Md-vtree labelling*: The label is set to be $\psi'(m') := \emptyset$; this is since all sum nodes only effectively have a single child, so they are trivially \mathcal{Q} -deterministic for any \mathcal{Q} . An example can be seen in Figure 4a.

2. Secondly, if both $m^{(1)}, m^{(2)}$ are leaves, then the PC nodes corresponding to these vtree nodes are also either leaves, or simple mixtures (sum nodes) of leaves. To compute the product of two sum nodes, we expand all combinations of the children of the sum nodes, as shown in Figure 4c.

- *Vtree node*: We create a leaf vtree node m' .
- *Parameter function*: The parameter function value for the new node $\rho'(m') = (\mathbf{T}_{m'}, \theta_{m'})$ is given as follows. For each pair of nodes $N_{m^{(1)},j} \in \mathbf{T}_{m^{(1)}}, N_{m^{(2)},k} \in \mathbf{T}_{m^{(2)}}$, we create a sum/leaf node $N_{m',i^{(1)}i^{(2)}}$, representing the product of $N_{m^{(1)},j}, N_{m^{(2)},k}$. The weights are defined as $\theta_{m',i^{(1)}i^{(2)},j^{(1)}j^{(2)}} := \theta_{m',i^{(1)},j^{(1)}} \theta_{m',i^{(2)},j^{(2)}}$.
- *Md-vtree labelling*: The label is set to be $\psi(m') := \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$. This is best seen with an example; in Figure 4c, we see an example of the product of two sum nodes with leaf node children, where one node is A -deterministic and the other is B -deterministic. The resulting sum node in the output circuit has children corresponding to (the product of) each combination of the children of the original two sum nodes; as a result, each child of the output sum node corresponds to a different value of (A, B) , and so the output sum node is $\{A, B\}$ -deterministic.

3. Thirdly, if one of the nodes has the same restricted scope as a child of the other node, e.g. $\phi_C^{(1)}(m^{(2)}) = \phi_C^{(1)}(m_r^{(1)})$, we “defer” the product as shown in Figure 4b.

- *Vtree node*: We create a vtree node m' with $(m'_l = m_l^{(1)}, m'_r = \text{PROD}(v_{m_r^{(1)}}^{(1)}, v_{m^{(2)}}^{(2)}))$ as children.
- *Parameter function*: The parameter function value for the new node $\rho'(m') = (\mathbf{T}_{m'}, \theta_{m'})$ is given as follows. For every pair of sum nodes $T_{m^{(1)},i^{(1)}} \in \mathbf{T}_{m^{(1)}}, T_{m^{(2)},i^{(2)}} \in \mathbf{T}_{m^{(2)}}$, we create a sum node $T_{m',i^{(1)}i^{(2)}}$. The weights are defined as $\theta_{m',i^{(1)}i^{(2)}j^{(1)}k^{(1)}k^{(2)}} := \theta_{m^{(1)},i^{(1)}j^{(1)}k^{(1)}} \mathbb{1}_{i^{(2)}=k^{(2)}}$. Note that the index j only has a single index as the left child $m'_l = m_l^{(1)}$, and so the sum nodes $\mathbf{T}_{m'_l}$ are copies of the sum nodes from $\mathbf{T}_{m_l^{(1)}}$.
- *Md-vtree labelling*: The label is set to be $\psi'(m') := \psi^{(1)}(m^{(1)})$, as the marginalized support of the children of the output sum nodes is a subset of the marginalized support of the corresponding sum node from $m^{(1)}$, as can be seen in Figure 4b.

4. Finally, in any other case, the children of $m^{(1)}, m^{(2)}$ have matching restricted scopes, e.g. $\phi_{C_{12}}^{(1)}(m_l^{(1)}) = \phi_{C_{12}}^{(2)}(m_l^{(2)})$ and $\phi_C^{(1)}(m_r^{(1)}) = \phi_C^{(2)}(m_r^{(2)})$, we expand all combinations:

- *Vtree node*: We create a vtree node m' with $(m'_l = \text{PROD}(m_l^{(1)}, m_l^{(2)}), m'_r = \text{PROD}(m_r^{(1)}, m_r^{(2)}))$ as children.

$(\mathbf{T}_{m^{(1)}}, \theta_{m^{(1)}}), \rho^{(2)}(m^{(2)}) = (\mathbf{T}_{m^{(2)}}, \theta_{m^{(2)}})$. In every case, $\mathbf{T}_{m'}$ will contain one node $T_{m',i^{(1)}i^{(2)}}$ corresponding to every pair of nodes $T_{m^{(1)},i^{(1)}} \in \mathbf{T}_{m^{(1)}}, T_{m^{(2)},i^{(2)}} \in \mathbf{T}_{m^{(2)}}$; we thus notate it with two dimensions. Correspondingly, in general, $\theta_{m'}$ will contain one weight $\theta_{m',i^{(1)}i^{(2)}j^{(1)}j^{(2)}k^{(1)}k^{(2)}}$ for every combination of nodes $T_{m',i^{(1)}i^{(2)}} \in \mathbf{T}_{m'}, T_{m'_l,j^{(1)}j^{(2)}} \in \mathbf{T}_{m'_l}, T_{m'_r,k^{(1)}k^{(2)}} \in \mathbf{T}_{m'_r}$, where m'_l and m'_r are the left and right children of m' in the new md-vtree.

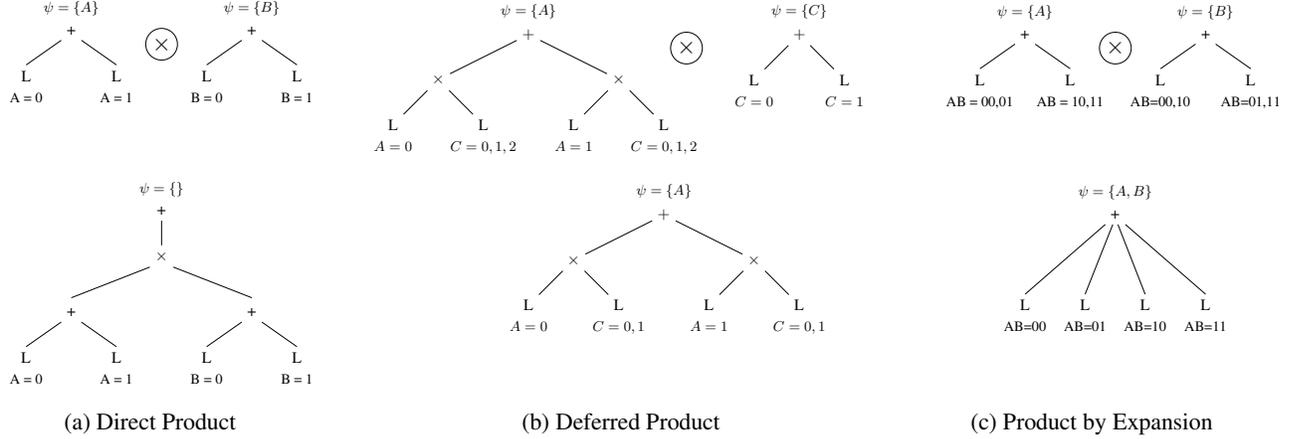


Figure 4: Examples of product of the two sum nodes on the top half, with the result shown in the bottom half. The root sum node is labelled with the corresponding vtree node label, while the leaves are labelled with their support.

- *Parameter function*: The parameter function value for the new node $\rho'(m') = (T_{m'}, \theta_{m'})$ is given as follows. For every pair of sum nodes $T_{m^{(1)}, i^{(1)}} \in T_{m^{(1)}}$, $T_{m^{(2)}, i^{(2)}} \in T_{m^{(2)}}$, we create a sum node $T_{m', i^{(1)} i^{(2)}}$. The weights are defined as $\theta_{m', i^{(1)} i^{(2)} j^{(1)} j^{(2)} k^{(1)} k^{(2)}} := \theta_{m^{(1)}, i^{(1)} j^{(1)} k^{(1)}} \theta_{m^{(2)}, i^{(2)} j^{(2)} k^{(2)}}$.
- *Md-vtree labelling*: The label is set to be $\psi(m') := \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$, for similar reasons to the product of two leaf vtree nodes above.

With this, we have shown how to each recursive step of the product algorithm. Now, taking a step back, we consider the entire run of the recursive algorithm. Starting from md-vtrees $v^{(1)}, v^{(2)}$ over variables $V^{(1)}, V^{(2)}$, with common variables $C_{\text{global}} := V^{(1)} \cap V^{(2)}$, in each recursive call, we reduce the common variables, until either we reach two vtree nodes that do not have common variables, or we reach leaf vtree nodes. One property of the algorithm, which will be important for the proof of the MD-calculus rule below, is that at each recursive step of the product algorithm, the two input vtree nodes have the same restricted scope over C_{global} .

Proposition 6. *At each recursive step of Algorithm 4, we have that $\phi_{C_{\text{global}}}^{(1)}(m^{(1)}) = \phi_{C_{\text{global}}}^{(2)}(m^{(2)})$.*

Proof. Proof is by inspection; in each recursive case (3, 4), we have that $C = \phi_{C_{\text{global}}}^{(1)}(m^{(1)}) = \phi_{C_{\text{global}}}^{(2)}(m^{(2)})$, and match up the common variables among the recursive call(s). \square

POW($\cdot; \alpha$) For the power operation, we have Algorithm 5. This algorithm simply inverts all the weights/parameters of the circuit, as well as replacing the leaves with their reciprocals. Provided that the input circuit is deterministic, the output circuit faithfully represents the reciprocal of the input circuit (Vergari et al., 2021). As the transformation is simply numerical (i.e. not affecting the support of any node), the labels of all nodes remain the same.

MAX($\cdot; \alpha$) This operation returns a scalar.

LOG(\cdot) This operation returns a circuit which respects the same vtree, but does not have (marginal) determinism (Vergari et al., 2021).

B.2 MD-calculus and the Backward Problem

The algorithms for each of the basic operations above allow us to derive a md-vtree for the output circuit, given the md-vtree that the input circuit respects. The MD-calculus in Table 3 (repeated for convenience in Table 5) turns these results into a series of straightforward rules that can easily be applied to derive sufficient (but possibly not necessary) conditions for tractability of compositions of operations.

Theorem 3 (MD-calculus). *The conditions in Table 3 hold.*

Proof. To state the result more formally, we claim that if the input circuit(s) respect md-vtrees(s) implying the input condition, then the result of the operation applied to the input circuit(s) will respect a md-vtree implying the output condition.

MARG($\cdot; W$) For the marginalization operation, the output md-vtree is over variables $V \setminus W$. Thus, let Q be any subset of $V \setminus W$.

- *Input Condition:* The input condition requires that the input md-vtree w implies Q -determinism; that is, for every vtree node m , either $\phi_Q(m) = \emptyset$, or else $Q \supseteq \psi(m)$.
- *Algorithm:* In Algorithm 2, every vtree node m' in the output md-vtree corresponds to a vtree node m in the input md-vtree, such that $\phi'(m') = \phi(m) \setminus W$, and $\psi'(m') = \psi(m)$ if $\psi(m) \cap W = \emptyset$, or $\psi'(m') = U$ otherwise.
- *Proof for Output Condition:* For each vtree node m' , if $\phi'_Q(m') \neq \emptyset$, then, we have that:

$$\begin{aligned}
 & \phi_Q(m) \setminus W \neq \emptyset && \text{(by effect of algorithm)} \\
 \implies & \phi_Q(m) \neq \emptyset && \text{(weaker statement)} \\
 \implies & \psi(m) \subseteq Q && \text{(by input condition)} \\
 \implies & \psi'(m') \subseteq Q
 \end{aligned}$$

The last line follows since $Q \cap W = \emptyset$, so $\psi(m) \cap W = \emptyset$, and so we are in the algorithm case where the label is "copied". Thus, we have shown that the output md-vtree implies Q -determinism, as required.

INST($\cdot; w$) For the instantiation operation, the output md-vtree is over variables $V \setminus W$. Thus, let Q be any subset of $V \setminus W$.

- *Input Condition:* The input condition requires that the input md-vtree w implies $(Q \cup W')$ -determinism for some $W' \subseteq W$; that is, for every vtree node m , either $\phi_{Q \cup W'}(m) = \emptyset$, or else $Q \cup W' \supseteq \psi(m)$.
- *Algorithm:* In Algorithm 3, every vtree node m' in the output md-vtree corresponds to a vtree node m in the input md-vtree, such that $\phi'(m') = \phi(m) \setminus W$, and $\psi'(m') = \psi(m) \setminus W$.
- *Proof for Output Condition:* For each vtree node m' , if $\phi'_Q(m') \neq \emptyset$, then, we have that:

$$\begin{aligned}
 & \phi_Q(m) \setminus W \neq \emptyset && \text{(by effect of algorithm)} \\
 \implies & \phi_Q(m) \neq \emptyset && \text{(weaker statement)} \\
 \implies & \phi_{Q \cup W'}(m) \neq \emptyset && \text{(weaker statement)} \\
 \implies & \psi(m) \subseteq Q \cup W' && \text{(by input condition)} \\
 \implies & \psi'(m') \subseteq Q
 \end{aligned}$$

Here, the last line follows since the new label $\psi'(m') = \psi(m) \setminus W$ removes all elements of W , and thus W' , from $\psi(m)$. Thus, we have shown that the output md-vtree implies Q -determinism, as required.

PROD(\cdot, \cdot) For the product operation, the output md-vtree is over variables $V^{(1)} \cup V^{(2)}$. Thus, let Q be any subset of $V^{(1)} \cup V^{(2)}$.

- *Input Condition:* The input condition requires that the first input md-vtree $w^{(1)}$ implies $Q^{(1)}$ -determinism, and the second input md-vtree $w^{(2)}$ implies $Q^{(2)}$ -determinism, where *one* of the following holds:
 - $Q \subseteq V^{(1)} \cap V^{(2)}$ and $Q^{(1)} = Q^{(2)} = Q$;
 - $Q^{(1)}, Q^{(2)} \supseteq V^{(1)} \cap V^{(2)}$ and $Q = Q^{(1)} \cup Q^{(2)}$
- *Algorithm:* In Algorithm 4, every vtree node m' in the output md-vtree corresponds to a pair $m^{(1)}, m^{(2)}$ in the input md-vtrees respectively, such that $\phi'(m') = \phi^{(1)}(m^{(1)}) \cup \phi^{(2)}(m^{(2)})$. There are four cases of the algorithm to consider, in which the label is:
 1. $\psi'(m') = \emptyset$.

2. $\psi'(m') = \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$
3. $\psi'(m') = \psi^{(1)}(m^{(1)})$
4. $\psi'(m') = \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$

- *Proof for Output Condition:* We need to show that for each case 1-3 of the algorithm, and for either input condition (a), (b), that the condition for implied Q -determinism holds on m' ; that is, if $\phi'_Q(m') \neq \emptyset$, then $\psi'(m') \subseteq Q$. Assuming that $\phi'_Q(m') \neq \emptyset$, we have that

$$\begin{aligned}
 & \phi'_Q(m') \neq \emptyset \\
 \implies & \phi'(m') \cap Q \neq \emptyset && \text{(by definition of restricted scope)} \\
 \implies & (\phi^{(1)}(m^{(1)}) \cup \phi^{(2)}(m^{(2)})) \cap Q \neq \emptyset && \text{(by effect of algorithm)} \\
 \implies & \phi_{Q^{(1)}}^{(1)}(m^{(1)}) \cup \phi_{Q^{(2)}}^{(2)}(m^{(2)}) \neq \emptyset && \text{(rewriting)}
 \end{aligned}$$

However, this does not in general imply that $\phi_{Q^{(1)}}^{(1)}(m^{(1)}) \neq \emptyset$ or $\phi_{Q^{(2)}}^{(2)}(m^{(2)}) \neq \emptyset$. Thus, we look at the special cases defined by (a) and (b), and the algorithm variations 1, 2, 3, 4.

(a1, a2, a3, a4) In case (a), we have $Q^{(1)} = Q^{(2)} = Q \subseteq C_{\text{global}}$.

$$\begin{aligned}
 & \phi_{C_{\text{global}}}^{(1)}(m^{(1)}) = \phi_{C_{\text{global}}}^{(2)}(m^{(2)}) && \text{(by Proposition 6)} \\
 \implies & \phi_{C_{\text{global}}}^{(1)}(m^{(1)}) \cap Q = \phi_{C_{\text{global}}}^{(2)}(m^{(2)}) \cap Q \\
 \implies & \phi_{C_{\text{global}} \cap Q}^{(1)}(m^{(1)}) = \phi_{C_{\text{global}} \cap Q}^{(2)}(m^{(2)}) \\
 \implies & \phi_Q^{(1)}(m^{(1)}) = \phi_Q^{(2)}(m^{(2)}) && \text{(as } Q \subseteq C_{\text{global}}) \\
 \implies & \phi_Q^{(1)}(m^{(1)}) \neq \emptyset, \phi_Q^{(2)}(m^{(2)}) \neq \emptyset && \text{(as } \phi_Q^{(1)}(m^{(1)}) \cup \phi_Q^{(2)}(m^{(2)}) \neq \emptyset) \\
 \implies & \phi_{Q^{(1)}}^{(1)}(m^{(1)}) \neq \emptyset, \phi_{Q^{(2)}}^{(2)}(m^{(2)}) \neq \emptyset && \text{(as } Q^{(1)} = Q^{(2)} = Q)
 \end{aligned}$$

Thus, we have that $Q \supseteq \psi^{(1)}(m^{(1)})$ and $Q \supseteq \psi^{(2)}(m^{(2)})$, and so $Q \supseteq \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$. Finally, in each of the cases 1-4, we have $Q \supseteq \psi'(m')$, so the output md-vtree implies Q -determinism as required.

(b1) In case (b), we need to consider the cases of the algorithm separately. In case 1, $\psi'(m') \subseteq Q$ holds trivially as $\psi'(m') = \emptyset$, so we are done.

(b2, b3, b4) We have that $Q^{(1)}, Q^{(2)} \supseteq C_{\text{global}}$ and $Q = Q^{(1)} \cup Q^{(2)}$. The key observation is that, as we are not in case 1 of the Algorithm, $C = \phi^{(1)}(m^{(1)}) \cap \phi^{(2)}(m^{(2)})$ must be non-empty.

$$\begin{aligned}
 & C \neq \emptyset \\
 \implies & \phi_C^{(1)}(m^{(1)}) \neq \emptyset, \phi_C^{(2)}(m^{(2)}) \neq \emptyset && \text{(by definition of } C) \\
 \implies & \phi_{Q^{(1)}}^{(1)}(m^{(1)}) \neq \emptyset, \phi_{Q^{(2)}}^{(2)}(m^{(2)}) \neq \emptyset && \text{(as } Q^{(1)}, Q^{(2)} \supseteq C_{\text{global}} \supseteq C)
 \end{aligned}$$

Thus, we have that $Q^{(1)} \supseteq \psi^{(1)}(m^{(1)})$ and $Q^{(2)} \supseteq \psi^{(2)}(m^{(2)})$, and so $Q = Q^{(1)} \cup Q^{(2)} \supseteq \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)})$. In each of the cases 2-4, we have $Q \supseteq \psi'(m')$, so the output md-vtree implies Q -determinism as required.

POW Since the reciprocal algorithm retains the same labelling function in the output as the input, it follows that a Q -deterministic input circuit will result in a Q -deterministic output circuit.

MAX This operation returns a scalar.

LOG This operation does not have any marginal determinism conditions.

□

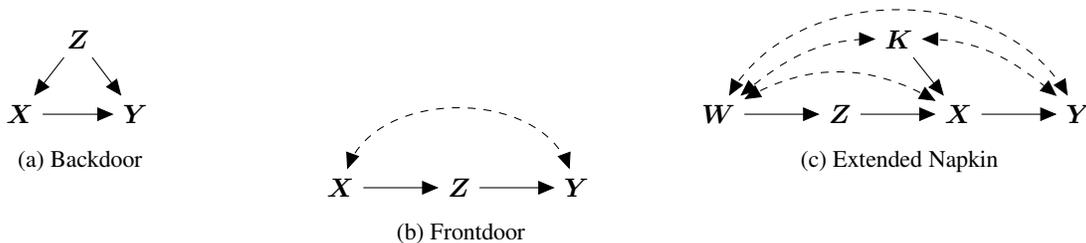


Figure 5: Examples of causal diagrams

C Causal Inference

In this section, we provide further background on causal inference for interested readers, the proof of our hardness result for backdoor adjustment, and then provide the full derivations of our tractability results for the backdoor, frontdoor and extended napkin formulae.

C.1 Background on Causal Inference

We use the framework of structural causal models (SCMs) (Pearl, 2009) to define the task of causal inference. SCMs provide a formal model of the underlying reality of a data-generating system over variables \mathbf{V} . In particular, each variable has an associated causal mechanism, which is a deterministic function of other variables in \mathbf{V} and a set of *exogenous variables* \mathbf{U} . The exogenous variables represent the “external state of the world”, or in other words, the source of randomness in generated data.

Definition 18 (Structural Causal Model). *A SCM \mathcal{M} is a tuple $(\mathbf{U}, \mathbf{V}, \mathbf{F}, p(\mathbf{U}))$ where:*

- \mathbf{U} is a set of exogenous (i.e. outside the model) random variables, which are typically unobserved;
- $\mathbf{V} = \{V_1, \dots, V_d\}$ is a set of endogenous (i.e. inside the model) random variables;
- $\mathbf{F} = \{F_1, \dots, F_d\}$ is a set of causal mechanisms (functions). In particular, each endogenous variable V_i has an associated function F_i which is a mapping from (the domain of) some subset $pa(V_i) \subseteq \mathbf{U} \cup (\mathbf{V} \setminus \{V_i\})$ to (the domain of) V_i . Members of $pa(V_i)$ are referred to as parents of V_i , and can be split into exogenous parents \mathbf{U}_i and endogenous parents Pa_i .
- $p(\mathbf{U})$ is a probability distribution over the exogenous variables.

Definition 19 (Semi-Markovian SCM). *A SCM is said to be semi-Markovian if the causal diagram induced by the SCM does not contain directed cycles. In such a case, for each variable $V \in \mathbf{V}$ we write $V(\mathbf{u})$ to denote the unique value of V given a particular value \mathbf{u} of the exogenous variables \mathbf{U} .*

For semi-Markovian SCMs, we define a causal graph/diagram G over the endogenous variables \mathbf{V} , as follows. For each variable $V \in \mathbf{V}$, we add a directed edge to this variable from any endogenous parents of V . Then, for each exogenous variable which is a parent of two endogenous variables V_1, V_2 , we add a bidirected edge between the variables. Intuitively, directed edges represent direct/functional relationships between observed variables, while bidirected arrows represent unobserved confounding. Examples of causal diagrams are shown in Figure 5.

A semi-Markovian SCM naturally induces a distribution $p(\mathbf{V})$ on the observed variables, through the distribution on exogenous variables $p(\mathbf{U})$. However, we can also use it to define the semantics of interventions, or actions on a system:

Definition 20 (Intervention). *Given a SCM \mathcal{M} , and any subset $\mathbf{X} \subseteq \mathbf{V}$, and an instantiation \mathbf{x} of \mathbf{X} , we can define an intervention on \mathcal{M} as producing a submodel $\mathcal{M}_{\mathbf{x}}$. $\mathcal{M}_{\mathbf{x}}$ differs from \mathcal{M} in that the functional relationships $F_{\mathbf{X}}$ are replaced with setting each variable to the fixed value \mathbf{x} , i.e. $F_{\mathbf{X}} = \mathbf{x}$.*

The interventional distribution $p_{\mathbf{x}}(\mathbf{V})$ (also written $p(\mathbf{V}|\text{do}(\mathbf{x}))$) is defined to be the distribution of \mathbf{V} in the intervened model $\mathcal{M}_{\mathbf{x}}$.

While SCMs are powerful functional models, we rarely have access to the true underlying SCM for a system. Thus, we often make the much weaker assumption of knowledge of the causal graph encoding the *qualitative* functional dependencies between the endogenous variables. In particular, an important question is, given just a causal graph G , and observational data $p(\mathbf{V})$ on the variables \mathbf{V} , under what circumstances can we deduce (properties of) interventional distributions $p_x(\mathbf{V})$? More formally, we say that an interventional distribution (query) $p_x(\mathbf{Y})$ (for some $\mathbf{Y} \subseteq \mathbf{V}$) is *identifiable* with respect to the causal graph G , if $p_x(\mathbf{Y})$ is uniquely computable from $p(\mathbf{V})$ for any SCM \mathcal{M} that induces G .

The problem of causal identification can be solved constructively using the do-calculus (Pearl, 1995), which comprises a set of rules for transforming a given interventional expression $p_x(\mathbf{Y})$ into a function the observational distribution $p(\mathbf{V})$, given the causal graph G . In particular, it was later shown that the do-calculus is *complete* (Shpitser and Pearl, 2006), i.e., if a query is identifiable, then the do-calculus can derive a formula (estimand), and there exists a polynomial time algorithm for finding such a formula (Shpitser and Pearl, 2008). For example, for the backdoor causal graph (Figure 2a), we can obtain a formula $p_x(\mathbf{Y}) = \sum_{\mathbf{Z}} p(\mathbf{Z})p(\mathbf{Y}|\mathbf{x}, \mathbf{Z})$.

C.2 Hardness of Causal Inference for Probabilistic Circuit Models

Now suppose we have a model $p(\mathbf{V})$ of the observational distribution, perhaps learned from data, and we would like to compute $p_x(\mathbf{Y})$ in the identifiable backdoor case. Unfortunately, in high dimensions, naïve computation of the do-calculus formula $p_x(\mathbf{Y}) = \sum_{\mathbf{Z}} p(\mathbf{Z})p(\mathbf{Y}|\mathbf{x}, \mathbf{Z})$ is computationally intractable, as it involves a summation that is exponential in the dimension $|\mathbf{Z}|$. The natural question is then whether there exist probabilistic models in which we can compute the backdoor query more efficiently. Unfortunately, despite the tractability of PCs for most probabilistic inference tasks, we show that, if the observed data distribution is modelled by a probabilistic circuit, current structural and support properties are not sufficient for exact causal inference:

Theorem 4 (Hardness of Backdoor Query). *The backdoor query for decomposable and smooth PCs is #P-hard, even if the PC is structured decomposable and deterministic.*

Proof. We prove this in the case of binary variables for brevity of presentation, though the proof can be easily extended to non-binary discrete variables. Our proof is based on a reduction from the problem of computing the expectation of a logistic regression model, which was defined and shown to be #P-hard in Van den Broeck et al. (2022) and which we refer to as the EXPLR problem. In particular, for any EXPLR problem over variables \mathbf{Z} , with input size $n_{\mathbf{Z}} = |\mathbf{Z}|$, we construct a circuit in time and with size linear in \mathbf{Z} and where computing the backdoor query $p_x(\mathbf{y}) = \sum_{\mathbf{Z}} p(\mathbf{Z})p(\mathbf{y}|\mathbf{x}, \mathbf{Z})$ is equivalent to solving the EXPLR problem.

The EXPLR problem is defined as computing the following quantity (where $w_i \in \mathbb{R}$):

$$\text{EXPLR}(\mathbf{w}) = \sum_{\mathbf{z}} \frac{1}{1 + e^{-(w_0 + \sum_i w_i z_i)}} \quad (10)$$

We will construct a circuit over variables $\mathbf{V} = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$, where the sets $\mathbf{X} = \{X\}$ and $\mathbf{Y} = \{Y\}$ each consist of a single variable, and consider the backdoor query for instantiations x, y of X, Y . We begin by defining a number of auxiliary circuits/nodes for \mathbf{X}, \mathbf{Y} and \mathbf{Z} individually, all structured decomposable, smooth and deterministic, which will be part of the construction of the main circuit.

First, for \mathbf{Y} we define the leaf nodes $\mathbb{1}_y, \mathbb{1}_{\neg y}$ to encode the functions $p_{\mathbb{1}_y}(Y) := \mathbb{1}(Y = y), p_{\mathbb{1}_{\neg y}}(Y) := \mathbb{1}(Y = \neg y)$ respectively. For \mathbf{X} , we define $\mathbb{1}_x, \mathbb{1}_{\neg x}$ to encode $p_{\mathbb{1}_x}(X) := \mathbb{1}(X = x), p_{\mathbb{1}_{\neg x}}(X) := \mathbb{1}(X = \neg x)$ (respectively) in a similar manner. Finally, for \mathbf{Z} , we define two circuits, $\mathbb{1}_{\mathbf{Z}}$ and $C_{\mathbf{Z}}$, as follows. Let $\mathbf{Z} := \{Z_1, \dots, Z_{n_{\mathbf{Z}}}\}$ be an arbitrary ordering of the variables in \mathbf{Z} , and let $\mathbf{Z}_{\geq i}$ denote $\{X_i, \dots, X_{n_{\mathbf{Z}}}\}$ for any $1 \leq i \leq n_{\mathbf{Z}}$. Then we define the circuit $\mathbb{1}_{\mathbf{Z}}$ recursively as follows, where $\mathbb{1}_{\mathbf{Z}} := \mathbb{1}_{\mathbf{Z}_{\geq 1}}$ (where \times indicates a product node with its arguments as children):

$$\mathbb{1}_{\mathbf{Z}_{\geq i}} := \begin{cases} \mathbb{1}_{Z_i} \times \mathbb{1}_{\mathbf{Z}_{\geq i+1}} & 1 \leq i < n_{\mathbf{Z}} \\ \mathbb{1}_{Z_i} & i = n_{\mathbf{Z}} \end{cases} \quad (11)$$

This circuit consists of a series of product units, and leaf units $\mathbb{1}_{Z_i}$ for each $Z_i \in \mathbf{Z}$ which we define to encode the function $p_{\mathbb{1}_{Z_i}}(Z_i) \equiv 1$ (for all values of Z_i). Thus, the circuit as a whole encodes $p_{\mathbb{1}_{\mathbf{Z}}}(\mathbf{Z}) \equiv 1$ for all values of \mathbf{Z} . In terms of structural and support properties, the circuit is trivially deterministic and smooth as it does not contain any sum nodes, and is clearly also structured decomposable. Finally, it can also be seen that the size $|\mathbb{1}_{\mathbf{Z}}|$ (number of edges) of $\mathbb{1}_{\mathbf{Z}}$ is $O(n_{\mathbf{Z}})$.

We now design a circuit C_Z to encode the function $e^{-(w_0 + \sum_i w_i z_i)}$ as follows, where $C_Z := C_{Z_{\geq 1}}$:

$$C_{Z_{\geq i}} := \begin{cases} C_{Z_i} \times C_{Z_{\geq i+1}} & 1 \leq i < n_Z \\ C_{Z_i} & i = n_Z \end{cases} \quad (12)$$

where we define leaf nodes C_{Z_i} to encode $p_{C_{Z_i}}(Z_i) := e^{-w_i Z_i}$ for $1 \leq i < n_Z$ and $p_{C_{Z_i}}(Z_i) := e^{-(w_0 + w_i Z_i)}$ for $i = n_Z$. By recursion it can be seen that this circuit does indeed encode $p_{C_Z}(Z) = e^{-(w_0 + \sum_i w_i z_i)}$. This circuit is deterministic and smooth, and also decomposes in the same way as $\mathbb{1}_Z$, i.e. they are structured decomposable with the same vtree. It can also be seen that the size $|C_Z|$ of C_Z is $O(n_Z)$.

Now, consider the following probabilistic circuit over $V = X \cup Y \cup Z$ (where $\times, +$ represent product, sum nodes respectively):

$$C := \mathbb{1}_y \times (\mathbb{1}_x \times \mathbb{1}_Z + \mathbb{1}_{\neg x} \times \mathbb{1}_Z) + \mathbb{1}_{\neg y} \times (\mathbb{1}_x \times C_Z + \mathbb{1}_{\neg x} \times \mathbb{1}_Z) \quad (13)$$

C is structured decomposable as all of the product units with the same scope in the equation above decompose in the same way, and we have seen that $\mathbb{1}_Z$ and C_Z are structured decomposable with respect to the same vtree. It is also smooth and deterministic as the individual circuits $\mathbb{1}_Z$ and C_Z are smooth and deterministic, and the sum nodes in the equation satisfy determinism by the fact that $(\mathbb{1}_y, \mathbb{1}_{\neg y})$ and $(\mathbb{1}_x, \mathbb{1}_{\neg x})$ have disjoint support. Finally, as the sizes of $\mathbb{1}_Z$ and C_Z are $O(n_Z)$, $|C(V)|$ is also $O(n_Z)$.

Now, we show that the backdoor query on C is equivalent to solving the corresponding EXPLR problem. First, we derive expressions for all of the individual components of the backdoor formula on the circuit C , by evaluating according to Equation 13:

$$\begin{aligned} p_C(x, y, z) &= p_{\mathbb{1}_y}(y) \times (p_{\mathbb{1}_x}(x) \times p_{\mathbb{1}_Z}(z) + p_{\mathbb{1}_{\neg x}}(x) \times p_{\mathbb{1}_Z}(z)) + p_{\mathbb{1}_{\neg y}}(y) \times (p_{\mathbb{1}_x}(x) \times p_{C_Z}(z) + p_{\mathbb{1}_{\neg x}}(x) \times p_{\mathbb{1}_Z}(z)) \\ &= 1 \times (1 \times 1 + 0 \times 1) + 0 \times (1 \times 1 + 0 \times 1) \\ &= 1 \end{aligned}$$

$$\begin{aligned} p_C(x, z) &= p_C(x, y, z) + p_C(x, \neg y, z) \\ &= 1 + C_Z(z) \end{aligned}$$

$$\begin{aligned} p_C(z) &= p_C(x, z) + p_C(\neg x, y, z) + p_C(\neg x, \neg y, z) \\ &= p_C(x, z) + 1 + 1 \\ &= p_C(x, z) + 2 \end{aligned}$$

The backdoor query for C can then be expressed as

$$\begin{aligned} p_x(\mathbf{y}) &= \sum_z p_C(z) p_C(y|x, z) = \sum_z (p_C(x, z) + 2) \frac{p_C(x, y, z)}{p_C(x, z)} = \sum_z \left[1 + \frac{2}{1 + p_{C_Z}(z)} \right] \\ &= 2^{n_Z} + 2 \sum_z \frac{1}{1 + e^{-(w_0 + \sum_i w_i z_i)}} \end{aligned}$$

Thus, if we can compute the backdoor query for C , then we can compute the given EXPLR problem, completing the reduction. \square

The significance of this result is that it implies hardness of causal inference for structured decomposable and deterministic PCs whenever there is a *valid backdoor adjustment* (whether we use the backdoor formula or not), one of the simplest and most common cases where the causal effect is identifiable.

Corollary 1. *For any interventional query $p_x(\mathbf{y})$ and causal diagram G such that the query is identifiable through a backdoor adjustment, and the observational distribution $p(V)$ given as a decomposable and smooth circuit C encoding p , computing $p_x(\mathbf{y})$ is #P-hard, even if the circuit is structured decomposable and deterministic.*

Proof. By the identifiability condition, we have that $p_x(\mathbf{y}) = \sum_z p(\mathbf{y}|x, z) p(z) = \sum_z C(z) C(\mathbf{y}|x, z)$, which is the backdoor query. Hardness of computing the causal effect then follows from hardness of backdoor queries for the probabilistic circuit. \square

C.3 MD-calculus and Causal Inference

As sketched in the main paper, MD-calculus provides the tools for us to analyze what properties we need (to add to structured decomposability and determinism) to enable tractable computation of causal queries. We now provide the full derivations for the backdoor, frontdoor, and extended napkin cases. For convenience, in the following we will refer to the intermediate circuits in a computation by the functions they encode, e.g. $C(p_C(\mathbf{X}, \mathbf{Z}))$ for the circuit obtained from applying the $\text{MARG}(\cdot; \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Z}))$ operation to C .

C.3.1 Backdoor

We begin with the backdoor query. In cases where there is a valid backdoor adjustment set, such as in Figure 2a, we have the following formula for the interventional distribution:

$$p_{C,\mathbf{X}}(\mathbf{Y}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}) p_C(\mathbf{Y}|\mathbf{X}, \mathbf{Z}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}) \frac{p_C(\mathbf{Y}, \mathbf{X}, \mathbf{Z})}{p_C(\mathbf{X}, \mathbf{Z})}$$

Before continuing, it is worth noting that the expression above is valid for any value of \mathbf{X}, \mathbf{Y} , though in causal inference it is more typical that we are interested in evaluating $p_{C,\mathbf{X}}(\mathbf{Y})$ for specific values \mathbf{x} of \mathbf{X} , i.e. a specific intervention, or a small set of interventions. This distinction is important as we will see, interestingly, that instantiating \mathbf{X} makes the query more tractable in the sense that the marginal determinism requirements for the circuit C are more relaxed.

To apply the MD-calculus, we first identify the deterministic operations in the pipeline. For the backdoor query, the only deterministic operation is the $\text{POW}(\cdot; -1)$ operation, applied to $C(p_C(\mathbf{X}, \mathbf{Z}))$ which requires a deterministic input circuit. Then, we can work *backwards* through the pipeline from POW in order to derive tractability conditions on C .

1. **Requirement:** $C(p_C(\mathbf{X}, \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic.
2. $\text{MARG}(\cdot; \mathbf{Y})$: $C(p_C(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic $\implies C(p_C(\mathbf{X}, \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic
3. $\text{MARG}(\cdot; \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}))$: $C(p_C(\mathbf{V}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic $\implies C(p_C(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic
4. **Sufficient Condition:** $C = C(p_C(\mathbf{V}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic.

This simple derivation shows that it suffices for C to be $(\mathbf{X} \cup \mathbf{Z})$ -deterministic to compute the backdoor query. Now, let us consider the case in which we instantiate \mathbf{X} with a value \mathbf{x} :

$$p_{C,\mathbf{x}}(\mathbf{Y}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}) p_C(\mathbf{Y}|\mathbf{x}, \mathbf{Z}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}) \frac{p_C(\mathbf{Y}, \mathbf{x}, \mathbf{Z})}{p_C(\mathbf{x}, \mathbf{Z})}$$

Here, employing the MD-calculus gives us the following:

1. **Requirement:** $C(p_C(\mathbf{x}, \mathbf{Z}))$ is \mathbf{Z} -deterministic.
2. $\text{MARG}(\cdot; \mathbf{Y})$: $C(p_C(\mathbf{x}, \mathbf{Y}, \mathbf{Z}))$ is \mathbf{Z} -deterministic $\implies C(p_C(\mathbf{x}, \mathbf{Z}))$ is \mathbf{Z} -deterministic
3. $\text{MARG}(\cdot; \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}))$: $C(p_C(\mathbf{x}, \mathbf{V} \setminus \mathbf{X}))$ is \mathbf{Z} -deterministic $\implies C(p_C(\mathbf{x}, \mathbf{Y}, \mathbf{Z}))$ is \mathbf{Z} -deterministic
4. $\text{INST}(\cdot; \mathbf{x})$: $C(p_C(\mathbf{V}))$ is $(\mathbf{X}' \cup \mathbf{Z})$ -deterministic for some $\mathbf{X}' \subseteq \mathbf{X}$ $\implies C(p_C(\mathbf{x}, \mathbf{V} \setminus \mathbf{X}))$ is \mathbf{Z} -deterministic
5. **Sufficient Condition:** C is $(\mathbf{X}' \cup \mathbf{Z})$ -deterministic for some $\mathbf{X}' \subseteq \mathbf{X}$.

Notice that in the requirement, due to the instantiation, the input to the POW operation $C(p_C(\mathbf{x}, \mathbf{Z}))$ has scope \mathbf{Z} , meaning that we require it to be \mathbf{Z} -deterministic rather than $(\mathbf{X} \cup \mathbf{Z})$ -deterministic. In the final step, we use the MD-calculus rule for instantiation. This shows that the instantiated backdoor adjustment is tractable for a wider range of circuits than if we insisted on a circuit encoding $p_{C,\mathbf{X}}(\mathbf{Y})$ as a function of \mathbf{X} (and \mathbf{Y}).

C.3.2 Frontdoor

Another common case where the causal effect is identifiable is the frontdoor causal diagram, shown in Figure 5b. Unlike the backdoor case, there is unobserved confounding of \mathbf{X} and \mathbf{Y} , represented by the dashed bidirectional arrow. However, the existence of the observed mediator \mathbf{Z} nonetheless allows for identifiability, via the formula

$$p_{C,\mathbf{x}}(\mathbf{Y}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}|\mathbf{X}) \sum_{\mathbf{X}'} p_C(\mathbf{X}') p_C(\mathbf{Y}|\mathbf{X}', \mathbf{Z}) \quad (14)$$

We now employ the MD-calculus to derive tractability conditions. This time, looking at the conditionals, there are two deterministic (POW) operations, as well as an auxiliary variable \mathbf{X}' (that is summed out in the end), which has the same joint distribution with $\mathbf{V} \setminus \mathbf{X}$ as \mathbf{X} .

1. **Requirement:** $C(p_C(\mathbf{X}))$ is \mathbf{X} -deterministic and $C(p_C(\mathbf{X}, \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic.
2. $\text{MARG}(\cdot; \mathbf{V} \setminus \mathbf{X})$: $C(p_C(\mathbf{V}))$ is \mathbf{X} -deterministic $\implies C(p_C(\mathbf{X}))$ is \mathbf{X} -deterministic.
3. $\text{MARG}(\cdot; \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Z}))$: $C(p_C(\mathbf{V}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic $\implies C(p_C(\mathbf{X} \cup \mathbf{Z}))$ is $(\mathbf{X} \cup \mathbf{Z})$ -deterministic.
4. **Sufficient Condition:** C is \mathbf{X} -deterministic and $(\mathbf{X} \cup \mathbf{Z})$ -deterministic.

Now, let us consider instantiating \mathbf{X} with value \mathbf{x} . In this case, we have:

$$p_{C,\mathbf{x}}(\mathbf{Y}) = \sum_{\mathbf{Z}} p_C(\mathbf{Z}|\mathbf{x}) \sum_{\mathbf{X}'} p_C(\mathbf{X}') p_C(\mathbf{Y}|\mathbf{X}', \mathbf{Z}) \quad (15)$$

Note that in this case, the conditional $p_C(\mathbf{Z}|\mathbf{x})$ does not impose any determinism requirements, since the input to the POW operation is a scalar $C(p_C(\mathbf{x}))$. However, the requirements for the other conditional remain the same, as \mathbf{X}' is an auxiliary variable that is not tied to the intervention value \mathbf{x} . Overall, we can conclude that C being $(\mathbf{X} \cup \mathbf{Z})$ -deterministic is sufficient for the instantiated frontdoor formula, which is again weaker than in the non-instantiated case.

C.3.3 Extended Napkin

The extended napkin causal graph in Figure 2b is an extension of the so-called *napkin* causal graph (Pearl, 2009), which is obtained by removing \mathbf{K} from the graph. For this diagram, the do-calculus gives us the following formula for the interventional distribution:

$$p_{C,\mathbf{X}}(\mathbf{Y}) = \sum_{\mathbf{K}} \left(\sum_{\mathbf{W}, \mathbf{X}', \mathbf{Y}'} p_C(\mathbf{X}', \mathbf{Y}'|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K}) \right) \frac{\sum_{\mathbf{W}} p_C(\mathbf{X}, \mathbf{Y}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}{\sum_{\mathbf{W}} p_C(\mathbf{X}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}$$

As described in the main paper, this is a case where we need to instantiate \mathbf{X} for tractability, leading to the following formula:

$$p_{C,\mathbf{x}}(\mathbf{Y}) = \sum_{\mathbf{K}} \left(\sum_{\mathbf{W}, \mathbf{X}', \mathbf{Y}'} p_C(\mathbf{X}', \mathbf{Y}'|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K}) \right) \frac{\sum_{\mathbf{W}} p_C(\mathbf{x}, \mathbf{Y}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}{\sum_{\mathbf{W}} p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K})}$$

In this formula, there are four deterministic operations. Three of them relate to the conditionals in the formula, while the final one is the POW($\cdot; -1$) operation applied to $C(\sum_{\mathbf{W}} p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K}))$. To derive a sufficient tractability condition on C , we work backward through the computation from each of these operations.

We first consider the conditionals in the formula. Through similar reasoning to the instantiated backdoor formula, all of these can be computed efficiently as long as C is $(\mathbf{K} \cup \mathbf{W} \cup \mathbf{Z}')$ -deterministic, for some $\mathbf{Z}' \subseteq \mathbf{Z}$.

For the final deterministic operation POW($\cdot; -1$) applied to $C(\sum_{\mathbf{W}} p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K}))$, note that the input circuit has scope \mathbf{K} , so the requirement is that it is \mathbf{K} -deterministic. At this point, we can apply the rules for MARG and PROD:

1. **Requirement:** $C(\sum_{\mathbf{W}} p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}) p_C(\mathbf{W}, \mathbf{K}))$ is \mathbf{K} -deterministic.

2. MARG: $C(p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W})p_C(\mathbf{W}, \mathbf{K}))$ is \mathbf{K} -deterministic $\implies C(\sum_{\mathbf{W}} p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W})p_C(\mathbf{W}, \mathbf{K}))$ is \mathbf{K} -deterministic.
3. PROD: $C(p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}))$, $C(p_C(\mathbf{W}, \mathbf{K}))$ both \mathbf{K} -deterministic $\implies C(p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W})p_C(\mathbf{W}, \mathbf{K}))$ is \mathbf{K} -deterministic, by rule (a), where $\mathbf{Q} = \mathbf{K}$.
4. MARG: $C(p_C(\mathbf{V}))$ is \mathbf{K} -deterministic $\implies C(p_C(\mathbf{W}, \mathbf{K}))$ is \mathbf{K} -deterministic
5. PROD: $C(p_C(\mathbf{x}, \mathbf{K}, \mathbf{z}, \mathbf{W}))$, $C(p_C(\mathbf{K}, \mathbf{z}, \mathbf{W})^{-1})$ both \mathbf{K} -deterministic $\implies C(p_C(\mathbf{x}|\mathbf{K}, \mathbf{z}, \mathbf{W}))$ is \mathbf{K} -deterministic, by rule (a), where $\mathbf{Q} = \mathbf{K}$.
6. MARG: $C(p_C(\mathbf{x}, \mathbf{z}, \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Z})))$ is \mathbf{K} -deterministic $\implies C(p_C(\mathbf{x}, \mathbf{K}, \mathbf{z}, \mathbf{W}))$ is \mathbf{K} -deterministic
7. INST: $C(p_C(\mathbf{V}))$ is $(\mathbf{K} \cup \mathbf{X}' \cup \mathbf{Z}')$ -deterministic for some $\mathbf{X}' \subseteq \mathbf{X}$, $\mathbf{Z}' \subseteq \mathbf{Z}$ $\implies C(p_C(\mathbf{x}, \mathbf{z}, \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Z})))$ is \mathbf{K} -deterministic
8. POW: $C(p_C(\mathbf{K}, \mathbf{z}, \mathbf{W}))$ is \mathbf{K} -deterministic $\implies C(p_C(\mathbf{K}, \mathbf{z}, \mathbf{W})^{-1})$ is \mathbf{K} -deterministic
9. MARG: $C(p_C(\mathbf{z}, \mathbf{V} \setminus \mathbf{Z}))$ is \mathbf{K} -deterministic $\implies C(p_C(\mathbf{K}, \mathbf{z}, \mathbf{W}))$ is \mathbf{K} -deterministic
10. INST: $C(p_C(\mathbf{V}))$ is $(\mathbf{K} \cup \mathbf{Z}')$ -deterministic for some $\mathbf{Z}' \subseteq \mathbf{Z}$ $\implies C(p_C(\mathbf{z}, \mathbf{V} \setminus \mathbf{Z}))$ is \mathbf{K} -deterministic
11. **Sufficient Condition:** C is \mathbf{K} -deterministic

We have underlined the individual conditions on C that have been derived. For the reciprocal to be tractable, we need C to satisfy all of these. However, it can be seen that the first condition implies the other two (by taking $\mathbf{X}' = \emptyset$, $\mathbf{Z}' = \emptyset$), giving the condition at the bottom of the derivation. Now, combining with the previous conditions due to the conditional distributions, the overall sufficient condition for tractability of the (instantiated)extended napkin query is that C is \mathbf{K} -deterministic and $(\mathbf{K} \cup \mathbf{W} \cup \mathbf{Z}')$ -deterministic, for some $\mathbf{Z}' \subseteq \mathbf{Z}$.

References

- Choi, Y., Vergari, A., and Van den Broeck, G. (2020). Probabilistic circuits: A unifying framework for tractable probabilistic models. *arXiv preprint*.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.
- Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, UK, second edition.
- Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-markovian causal models. AAAI'06, page 1219–1226. AAAI Press.
- Shpitser, I. and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9:1941–1979.
- Van den Broeck, G., Lykov, A., Schleich, M., and Suciu, D. (2022). On the tractability of shap explanations. *J. Artif. Int. Res.*, 74.
- Vergari, A., Choi, Y., Liu, A., Teso, S., and Van den Broeck, G. (2021). A compositional atlas of tractable circuit operations for probabilistic inference. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 13189–13201.

Algorithm 4: $\text{PROD}(C^{(1)}, C^{(2)})$

Input: Input circuits $C^{(1)} = (v^{(1)} = (M^{(1)}, E^{(1)}, \phi^{(1)}, \psi^{(1)}, \rho^{(1)}), C^{(2)} = (v^{(2)} = (M^{(2)}, E^{(2)}, \phi^{(2)}, \psi^{(2)}, \rho^{(2)})$

Result: Output circuit $C' = (v', \psi', \rho')$

- 1 $m^{(1)}, m^{(2)} \leftarrow \text{root}(v^{(1)}), \text{root}(v^{(2)});$
- 2 $(m_l^{(1)}, m_r^{(1)}), (m_l^{(2)}, m_r^{(2)}) \leftarrow \text{children}(m^{(1)}), \text{children}(m^{(2)});$ // null if $m^{(1)}/m^{(2)}$ are leaf
- 3 $m' \leftarrow \text{newvtree}(\text{node});$
- 4 $C \leftarrow \phi^{(1)}(m^{(1)}) \cap \phi^{(2)}(m^{(2)});$
- 5 **if** $C = \emptyset$ **then**
- 6 | $v'_l, \psi'_l, \rho'_l \leftarrow v^{(1)}, \psi^{(1)}, \rho^{(1)};$
- 7 | $v'_r, \psi'_r, \rho'_r \leftarrow v^{(2)}, \psi^{(2)}, \rho^{(2)};$
- 8 | $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r;$ // combine the vtrees/labelling fn/param fn
- 9 | $\psi'(m') \leftarrow \emptyset;$ // Update label function
- 10 | $\mathbf{T}_{m'} \leftarrow (\mathbf{T}_{m', i^{(1)} i^{(2)}} \text{ for } i^{(1)} = 1, \dots, |\mathbf{T}_{m^{(1)}}|, i^{(2)} = 1, \dots, |\mathbf{T}_{m^{(2)}}|);$
- 11 | $\theta_{m', i^{(1)} i^{(2)} j k} \leftarrow \mathbb{1}_{i^{(1)}=j, i^{(2)}=k};$
- 12 | $\rho'(m') \leftarrow (\mathbf{T}_{m'}, \theta_{m'})$ // Update parameter function
- 13 **else if** $m^{(1)}$ and $m^{(2)}$ are leaves **then**
- 14 | $v', \psi', \rho' \leftarrow \text{createemptyvtree}();$
- 15 | $\psi'(m') = \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)});$
- 16 | $\mathbf{T}_{m'} \leftarrow (\mathbf{T}_{m', i^{(1)} i^{(2)}} \text{ for } i^{(1)} = 1, \dots, |\mathbf{T}_{m^{(1)}}|, i^{(2)} = 1, \dots, |\mathbf{T}_{m^{(2)}}|);$
- 17 | $\theta_{m', i^{(1)} i^{(2)} j^{(1)} j^{(2)} k^{(1)} k^{(2)}} \leftarrow \theta_{m^{(1)}, i^{(1)} j^{(1)} k^{(1)}} \theta_{m^{(2)}, i^{(2)} j^{(2)} k^{(2)}};$
- 18 | $\rho'(m') \leftarrow (\mathbf{T}_{m'}, \theta_{m'})$ // Update parameter function
- 19 **else if** $\phi_C^{(1)}(m^{(2)}) = \phi_C^{(1)}(m_r^{(1)})$ **then**
- 20 | $v'_l, \psi'_l, \rho'_l \leftarrow v_{m_l^{(1)}}^{(1)}, \psi^{(1)}, \rho^{(1)};$
- 21 | $v'_r, \psi'_r, \rho'_r \leftarrow \text{PROD}(m_r^{(1)}, m^{(2)});$
- 22 | $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r;$ // combine the vtrees/labelling fn/param fn
- 23 | $\psi'(m') = \psi^{(1)}(m^{(1)});$ // Update label function
- 24 | $\mathbf{T}_{m'} \leftarrow (\mathbf{T}_{m', i^{(1)} i^{(2)}} \text{ for } i^{(1)} = 1, \dots, |\mathbf{T}_{m^{(1)}}|, i^{(2)} = 1, \dots, |\mathbf{T}_{m^{(2)}}|);$
- 25 | $\theta_{m', i^{(1)} i^{(2)} j^{(1)} k^{(1)} k^{(2)}} \leftarrow \theta_{m^{(1)}, i^{(1)} j^{(1)} k^{(1)}} \mathbb{1}_{i^{(2)}=k^{(2)}};$
- 26 | $\rho'(m') \leftarrow (\mathbf{T}_{m'}, \theta_{m'})$ // Update parameter function
- 27 **else if** $\phi_C^{(1)}(m_l^{(1)}) = \phi_C^{(2)}(m_l^{(2)})$ and $\phi_C^{(1)}(m_r^{(2)}) = \phi_C^{(1)}(m_r^{(2)})$ **then**
- 28 | $v'_l, \psi'_l, \rho'_l \leftarrow \text{PROD}(m_l^{(1)}, m_l^{(2)});$
- 29 | $v'_r, \psi'_r, \rho'_r \leftarrow \text{PROD}(m_r^{(1)}, m_r^{(2)});$
- 30 | $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r;$ // combine the vtrees/labelling fn/param fn
- 31 | $\psi'(m') = \psi^{(1)}(m^{(1)}) \cup \psi^{(2)}(m^{(2)});$
- 32 | $\mathbf{T}_{m'} \leftarrow (\mathbf{T}_{m', i^{(1)} i^{(2)}} \text{ for } i^{(1)} = 1, \dots, |\mathbf{T}_{m^{(1)}}|, i^{(2)} = 1, \dots, |\mathbf{T}_{m^{(2)}}|);$
- 33 | $\theta_{m', i^{(1)} i^{(2)} j^{(1)} j^{(2)} k^{(1)} k^{(2)}} \leftarrow \theta_{m^{(1)}, i^{(1)} j^{(1)} k^{(1)}} \theta_{m^{(2)}, i^{(2)} j^{(2)} k^{(2)}};$
- 34 | $\rho'(m') \leftarrow (\mathbf{T}_{m'}, \theta_{m'})$ // Update parameter function
- 35 **else**
- 36 | **Return fail (not compatible)**
- 37 $\phi'(m') \leftarrow \phi^{(1)}(m^{(1)}) \cup \phi^{(2)}(m^{(2)});$ // Update scope function
- 38 $v' \leftarrow \text{addnode}(v'; m');$
- 39 $v' \leftarrow \text{addchildren}(v'; m', \text{root}(v'_l), \text{root}(v'_r));$
- 40 **Return** (v', ψ', ρ')

Algorithm 5: POW(C, α)

Input: Input circuit $C = (v = (M, E, \phi), \psi, \rho)$; power α

Result: Output circuit $C' = (v', \psi', \rho')$

```

1  $m \leftarrow \text{root}(v)$ ;
2  $m' \leftarrow \text{newnode}()$ ;
3 if  $m$  is leaf then // Update vtree structure and parameter function (leaf)
4    $v' \leftarrow \text{createvtree}(m')$ ; // create vtree with single node
5    $\rho'(m') \leftarrow (\text{POW}(L; \alpha) \text{ for } L \in \mathbf{T}_m, \theta_m)$ ; // apply power to leaf PC nodes
6 else // Update vtree structure and parameter function (non-leaf)
7    $m_l, m_r \leftarrow \text{children}(m)$ ;
8    $v'_l, \psi'_l, \rho'_l \leftarrow \text{POW}((v_{m_l}, \psi, \rho), \alpha)$ ;
9    $v'_r, \psi'_r, \rho'_r \leftarrow \text{POW}((v_{m_r}, \psi, \rho), \alpha)$ ;
10   $v', \psi', \rho' \leftarrow v'_l \cup v'_r, \psi'_l \cup \psi'_r, \rho'_l \cup \rho'_r$ ; // combine the vtrees/labelling fn/param fn
11   $v' \leftarrow \text{addnode}(v'; m')$ ;  $v' \leftarrow \text{addchildren}(v'; m', \text{root}(v'_l), \text{root}(v'_r))$ ;
12   $\rho'(m') \leftarrow \rho(m)$ ;
13  $\phi'(m') \leftarrow \phi(m)$ ; // Update scope function
14  $\psi'(m') \leftarrow \psi(m)$ ; // Update labelling function
15 Return  $(v', \psi', \rho')$ 

```
