

# *Efficient Knowledge Compilation Beyond Weighted Model Counting<sup>\*†</sup>*

RAFAEL KIESEL

TU Wien, Vienna, Austria  
(e-mail: [rafael.kiesel@web.de](mailto:rafael.kiesel@web.de))

PIETRO TOTIS and ANGELIKA KIMMIG

KU Leuven, Leuven, Belgium  
(e-mails: [pietro.totis@kuleuven.be](mailto:pietro.totis@kuleuven.be), [angelika.kimmig@kuleuven.be](mailto:angelika.kimmig@kuleuven.be))

submitted 24 May 2022; accepted 8 June 2022

---

## Abstract

Quantitative extensions of logic programming often require the solution of so called *second level* inference tasks, that is, problems that involve a third operation, such as maximization or normalization, on top of addition and multiplication, and thus go beyond the well-known weighted or algebraic model counting setting of probabilistic logic programming under the distribution semantics. We introduce Second Level Algebraic Model Counting (2AMC) as a generic framework for these kinds of problems. As 2AMC is to (algebraic) model counting what forall-exists-SAT is to propositional satisfiability, it is notoriously hard to solve. First level techniques based on Knowledge Compilation (KC) have been adapted for specific 2AMC instances by imposing variable order constraints on the resulting circuit. However, those constraints can severely increase the circuit size and thus decrease the efficiency of such approaches. We show that we can exploit the logical structure of a 2AMC problem to omit parts of these constraints, thus limiting the negative effect. Furthermore, we introduce and implement a strategy to generate a sufficient set of constraints statically, with a priori guarantees for the performance of KC. Our empirical evaluation on several benchmarks and tasks confirms that our theoretical results can translate into more efficient solving in practice.

**KEYWORDS:** design, analysis and implementation of languages, logic programming methodology and applications

---

## 1 Introduction

Especially in recent years, research on quantitative extensions in Logic Programming has flourished. ProbLog (De Raedt *et al.* 2007), PITA (Riguzzi and Swift 2011), SMProbLog (Totis *et al.* 2021) and others (Baral *et al.* 2009; Lee and Yang 2017) allow for probabilistic reasoning, DeepProbLog (Manhaeve *et al.* 2018), NeurASP (Yang

<sup>†</sup> Full proofs for all technical results are in the Supplementary Material for this paper in the TPLP archives.

<sup>\*</sup> This work has been supported by the Austrian Science Fund (FWF) Grant W1255-N23, the Fonds Wetenschappelijk Onderzoek (FWO) project N. G066818N, and TU Wien Bibliothek through its Open Access Funding Programme.

*et al.* 2020) and SLASH (Skryagin *et al.* 2021) additionally integrate neural networks into programs.

Algebraic Prolog (Kimmig *et al.* 2011), algebraic model counting (AMC) (Kimmig *et al.* 2017) and algebraic answer set counting (Eiter *et al.* 2021) define general frameworks based on semirings to express and solve quantitative *first level* problems, which compute *one* aggregate over all models, for example, counting the number of models, or summing or maximizing values associated with them, such as probabilities or utilities. Kimmig *et al.* (2017) showed that we can solve first level problems by compiling the logic program into a *tractable circuit representation*, on which evaluating an AMC task is in polynomial time if the semiring operations have constant cost.

However, many interesting tasks require two kinds of aggregation, and thus are *second level* problems that go beyond AMC. Examples include Maximum A Posteriori (MAP) inference in probabilistic programs, which involves maximizing over some variables while summing over others, inference in SLASH and SMProbLog, and optimization tasks in decision-theoretic or constrained probabilistic programming languages such as DTProbLog (Van den Broeck *et al.* 2010; Derkinderen and De Raedt 2020) and SC-ProbLog (Latour *et al.* 2017).

While second level problems stay hard on general tractable circuit representations, DTProbLog and SC-ProbLog are known to be polynomial time on  $\mathbf{X}$ -constrained SDDs. The key idea is to ensure that all variables of the *outer* aggregation appear before those of the *inner* one in the circuit so that we can perform both aggregations sequentially. This, however, comes at a high cost: circuits respecting this constraint may be exponentially larger than non-constrained ones, resulting in significantly slower inference. Additionally, certain optimization techniques used in knowledge compilation, such as unit propagation, may cause constraint violations.

In this paper, we generalize the AMC approach to second level problems, and show that we can often weaken the order constraint using *definability*: A variable  $Y$  is defined by a set of variables  $\mathbf{X}$  and a propositional theory  $\mathcal{T}$  if the value of  $Y$  is functionally determined by the assignment to  $\mathbf{X}$  in every satisfying assignment to  $\mathcal{T}$ . For example,  $a$  is defined by  $b$  in the theory  $\{a \leftrightarrow b\}$ . Informally, if a variable participating in the inner aggregation becomes defined by the variables of the outer aggregation at any point during compilation, we can move that variable to the outer aggregation. This can allow for exponentially smaller circuits and consequently faster evaluation, and additionally justifies the use of unit propagation.

Our main contributions are as follows:

- We introduce second level algebraic model counting (2AMC), a semiring-based unifying framework for second level quantitative problems, and show that MAP, DTProbLog and SMProbLog inference are 2AMC tasks.
- We weaken  $\mathbf{X}$ -*firstness*, the constraint that the variables in  $\mathbf{X}$  need to occur first, to  $\mathbf{X}$ -*firstness modulo definability* and show that this is sufficient for solving 2AMC tasks under weak additional restrictions.
- We lift methods for generating good variable orders statically from tree decompositions to the constrained setting.
- We implement our contributions in the algebraic answer set counter `aspmc` (Eiter *et al.* 2021) and the probabilistic reasoning engine ProbLog2 (Fierens *et al.* 2015).
- We evaluate our contributions on a range of benchmarks, demonstrating that drastically smaller circuits can be possible, and that our general tools are competitive

with state of the art implementations of specific second level tasks in logic programming.

## 2 Preliminaries

We consider propositional theories  $\mathcal{T}$  over variables  $\mathbf{X}$ . For a set of variables  $\mathbf{Y}$ , we denote by  $lit(\mathbf{Y})$  the set of literals over  $\mathbf{Y}$ , by  $int(\mathbf{Y})$  the set of assignments  $\mathbf{y}$  to  $\mathbf{Y}$  and by  $mod(\mathcal{T})$  those assignments that satisfy  $\mathcal{T}$ . Here, an assignment is a subset of  $lit(\mathbf{Y})$ , which contains exactly one of  $a$  and  $\neg a$  for each variable  $a \in \mathbf{Y}$ . Given a partial assignment  $\mathbf{y} \in int(\mathbf{Y})$  for a theory  $\mathcal{T}$  over  $\mathbf{X}$ , we denote by  $\mathcal{T} \upharpoonright_{\mathbf{y}}$  the theory over  $\mathbf{X} \setminus \mathbf{Y}$  obtained by conditioning  $\mathcal{T}$  on  $\mathbf{y}$ . We use  $\models$  for the usual entailment relation of propositional logic.

Algebraic Model Counting (AMC) is a general framework for quantitative reasoning over models that generalizes weighted model counting to the semiring setting (Kimmig et al. 2017).

*Definition 1 (Monoid, Semiring)*

A monoid  $\mathcal{M} = (M, \otimes, e_\otimes)$  consists of an associative binary operation  $\otimes$  with neutral element  $e_\otimes$  that operates on elements of  $M$ . A commutative semiring  $\mathcal{S} = (S, \oplus, \otimes, e_\oplus, e_\otimes)$  consists of two commutative monoids  $(S, \oplus, e_\oplus)$  and  $(S, \otimes, e_\otimes)$  such that  $\otimes$  right and left distributes over  $\oplus$  and  $e_\oplus$  annihilates  $S$ , that is  $\forall s \in S : s \otimes e_\oplus = e_\oplus = e_\oplus \otimes s$ .

Some examples of well-known commutative semirings are

- $\mathcal{P} = ([0, 1], +, \cdot, 0, 1)$ , the probability semiring,
- $\mathcal{S}_{\max, \cdot} = (\mathbb{R}_{\geq 0}, \max, \cdot, 0, 1)$ , the max-times semiring,
- $\mathcal{S}_{\max, +} = (\mathbb{R} \cup \{-\infty\}, \max, +, -\infty, 0)$ , the max-plus semiring,
- $\text{EU} = (\{(p, eu) \mid p \in [0, 1], eu \in \mathbb{R}\}, +, \otimes, (0, 0), (1, 0))$ , the expected utility semiring, where addition is coordinate-wise and  $(a_1, b_1) \otimes (a_2, b_2) = (a_1 \cdot a_2, a_2 \cdot b_1 + a_1 \cdot b_2)$ .

An AMC instance  $A = (\mathcal{T}, \mathcal{S}, \alpha)$  consists of a theory  $\mathcal{T}$  over variables  $\mathbf{X}$ , a commutative semiring  $\mathcal{S}$  and a labeling function  $\alpha : lit(\mathbf{X}) \rightarrow S$ . The value of  $A$  is

$$AMC(A) = \bigoplus_{M \in mod(\mathcal{T})} \bigotimes_{l \in M} \alpha(l).$$

A prominent example of AMC is inference in probabilistic logic programming, which uses the probability semiring  $\mathcal{P}$  and a theory for a probabilistic logic program (PLP). A PLP  $\mathcal{L} = \mathbf{F} \cup R$  is a set  $\mathbf{F}$  of probabilistic facts  $p :: f$  and a set  $R$  of rules  $h \leftarrow b_1, \dots, b_n, not\ c_1, \dots, not\ c_m$  such that each assignment (world)  $\mathbf{f}$  to  $\mathbf{F}$  leads to exactly one model. By abuse of notation, we use  $\mathcal{L}$  also for a propositional theory with the same models. We denote the Herbrand base, that is, the set of all ground atoms in  $\mathcal{L}$ , by  $\mathbf{H}$ . The success probability of a query  $q$  is  $SUCC(q) = \sum_{\mathbf{h} \in mod(\mathcal{L}|_q)} P(\mathbf{h})$ , the sum of the probabilities  $P(\mathbf{h}) = \prod_{l \in \mathbf{h}} \alpha(l)$  of the models  $\mathbf{h}$  where  $q$  succeeds. Here,  $\alpha(f) = p, \alpha(\neg f) = 1 - p$  for  $p :: f \in \mathbf{F}$  and  $\alpha(l) = 1$ , otherwise.

*Example 1 (Running Example)*

We use the following probabilistic logic program  $\mathcal{L}_{ex}$  throughout the paper.

$$0.4 :: a \qquad c \leftarrow a \qquad 0.6 :: b \qquad d \leftarrow b$$

Its four worlds are  $\mathbf{f}_1 = \{a, b\}, \mathbf{f}_2 = \{a, \neg b\}, \mathbf{f}_3 = \{\neg a, b\}, \mathbf{f}_4 = \{\neg a, \neg b\}$ , and the probabilities of their models  $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4$  are  $P(\mathbf{h}_1) = 0.24, P(\mathbf{h}_2) = 0.16, P(\mathbf{h}_3) = 0.36$

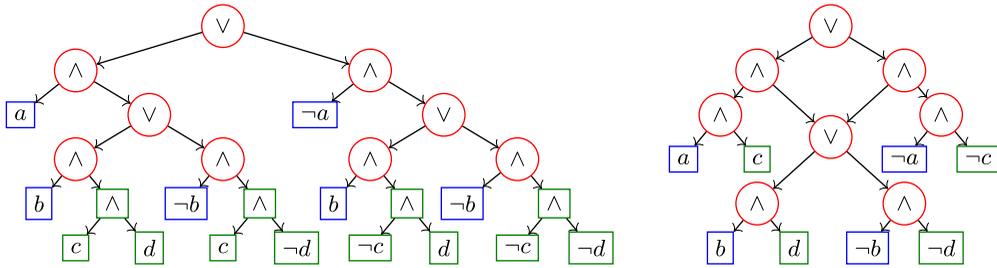


Fig. 1. Two sd-DNNFs for  $\mathcal{L}_{ex}$ . Mixed nodes for partition  $\{a, b\}, \{c, d\}$  are circled red and pure nodes are boxed blue or boxed green, when they are from  $\{a, b\}$  or  $\{c, d\}$ , respectively.

and  $P(\mathbf{h}_4) = 0.24$ . The query  $c$  succeeds in the models  $\mathbf{h}_1$  and  $\mathbf{h}_2$ , thus  $SUCC(c) = P(\mathbf{h}_1) + P(\mathbf{h}_2) = 0.4$ .

Kimmig et al. (2017) showed that Knowledge Compilation (KC) to sd-DNNFs solves any AMC problem. sd-DNNFs are special negation normal forms (NNFs). An NNF (Darwiche 2004) is a rooted directed acyclic graph in which each leaf node is labeled with a literal, true or false, and each internal node is labeled with a conjunction  $\wedge$  or disjunction  $\vee$ . For any node  $n$  in an NNF graph,  $Vars(n)$  denotes all variables in the subgraph rooted at  $n$ . By abuse of notation, we use  $n$  also to refer to the formula represented by the graph  $n$ . sd-DNNFs are NNFs that satisfy the following additional properties:

**Decomposability (D):**  $Vars(n_i) \cap Vars(n_j) = \emptyset$  for any two children  $n_i$  and  $n_j$  of an and-node.

**Determinism (d):**  $n_i \wedge n_j$  is logically inconsistent for any two children  $n_i$  and  $n_j$  of an or-node.

**Smoothness (s):**  $Vars(n_i) = Vars(n_j)$  for any two children  $n_i$  and  $n_j$  of an or-node.

In order to solve second level problems, we need Constrained KC (CKC), that is, an additional property on NNFs, apart from **s**, **d** and **D**, that restricts the order in which variables occur.

*Definition 2 (X-Firstness)*

Given an NNF  $n$  on variables partitioned into  $\mathbf{X}, \mathbf{Y}$ , we say an internal node  $n_i$  of  $n$  is *pure* if  $Vars(n_i) \subseteq \mathbf{X}$  or  $Vars(n_i) \subseteq \mathbf{Y}$  and *mixed* otherwise.  $n$  is an *X-first* NNF, if for each of its and-nodes  $n_i$  either all children of  $n_i$  are pure nodes, or one child is mixed and all other children  $n_j$  of  $n_i$  are pure with  $Vars(n_j) \subseteq \mathbf{X}$ .

Note that **X**-first NNFs contain a node  $n_{\mathbf{x}}$  equivalent to  $n \mid_{\mathbf{x}}$  for each  $\mathbf{x} \in int(\mathbf{X})$ .

*Example 2 (cont.)*

The NNFs in Figure 1 are both sd-DNNFs and equivalent to  $\mathcal{L}_{ex}$ . The left sd-DNNF is furthermore an  $\{a, b\}$ -first sd-DNNF, whereas the right is not.

### 3 Second Level Algebraic Model Counting

We now introduce Second Level Algebraic Model Counting (2AMC), a generalization of AMC that provides a unified view on second level problems.

Definition 3 (Second Level Algebraic Model Counting (2AMC))

A 2AMC instance is a tuple  $A = (\mathcal{T}, \mathbf{X}_I, \mathbf{X}_O, \alpha_I, \alpha_O, \mathcal{S}_I, \mathcal{S}_O, t)$ , where

- $\mathcal{T}$  is a propositional theory,
- $(\mathbf{X}_I, \mathbf{X}_O)$  is a partition of the variables in  $\mathcal{T}$ ,
- $\mathcal{S}_j = (S_j, \oplus^j, \otimes^j, e_{\oplus^j}, e_{\otimes^j})$  for  $j \in \{I, O\}$  is a commutative semiring,
- $\alpha_j : lit(\mathbf{X}_j) \rightarrow S_j$  for  $j \in \{I, O\}$  is a labeling function for literals, and
- $t : S_I \rightarrow S_O$  is a weight transformation function that respects  $t(e_{\oplus^I}) = e_{\oplus^O}$ .

The value of  $A$ , denoted by  $2AMC(A)$ , is defined as

$$2AMC(A) = \bigoplus_{\mathbf{x}_O \in int(\mathbf{X}_O)}^O \bigotimes_{x \in \mathbf{x}_O}^O \alpha_O(x) \otimes^O t \left( \bigoplus_{\mathbf{x}_I \in mod(\mathcal{T}|_{\mathbf{x}_O})}^I \bigotimes_{y \in \mathbf{x}_I}^I \alpha_I(y) \right).$$

An AMC instance is a 2AMC instance, where  $\mathbf{X}_O = \emptyset$  and  $t$  is the identity function, meaning we only sum up weights over  $\mathcal{S}_I$ . Intuitively, the idea behind 2AMC is that we solve an *inner* AMC instance over the variables in  $\mathbf{X}_I$  for each assignment to  $\mathbf{X}_O$ . Then we apply the transformation function to the result, thus replacing the inner summation by a corresponding element from the outer semiring, and solve a second *outer* AMC instance over the variables in  $\mathbf{X}_O$ .

Example 3 (cont.)

Consider the question whether it is more likely for  $c$  to be true or false in  $\mathcal{L}_{ex}$  from Example 1, that is, we want to find  $\arg \max_{\mathbf{c} \in int(c)} SUCC(\mathbf{c})$ . To keep notation simple, we consider max rather than  $\arg \max$  (see the discussion below Definition 5 for full details). Denoting the label of literal  $l$  by  $\alpha(l)$  as in the definition of  $SUCC$ , the task corresponds to

$$\max_{\mathbf{c} \in int(c)} \alpha(\mathbf{c}) \sum_{\mathbf{h} \in mod(\mathcal{L}_{ex} | \mathbf{c})} \prod_{l \in \mathbf{h}} \alpha(l).$$

We thus have a 2AMC task with outer variables  $\{c\}$ , inner variables  $\{a, b, d\}$  and the probability semiring and max-times semiring as inner and outer semiring, respectively, and both kinds of labels given by  $\alpha$ . The formal definition of this 2AMC instance is  $A_{ex} = (\mathcal{L}_{ex}, \{a, b, d\}, \{c\}, \alpha_I, \alpha_O, \mathcal{P}, \mathcal{S}_{max}, id)$ , where  $\alpha_I(l)$  is the probability of  $l$  if  $l \in lit(\{a, b\})$  and 1 if  $l \in lit(\{d\})$ , and  $\alpha_O(l) = 1, l \in lit(\{c\})$ . We can further evaluate the value as follows:

$$\begin{aligned} 2AMC(A_{ex}) &= \max_{\mathbf{c} \in int(c)} 1 \cdot \sum_{\substack{\mathbf{a} \in int(\{a\}), \mathbf{b} \in int(b), \mathbf{d} \in int(d) \\ \mathbf{a}=\mathbf{c}, \mathbf{b}=\mathbf{d}}} 1 \cdot \alpha_I(\mathbf{a}) \cdot \alpha_I(\mathbf{b}) \\ &= \max\{\alpha_I(a)\alpha_I(b) + \alpha_I(a)\alpha_I(-b), \alpha_I(-a)\alpha_I(b) + \alpha_I(-a)\alpha_I(-b)\} \\ &= \max\{0.4 \cdot 0.6 + 0.4 \cdot 0.4, 0.6 \cdot 0.6 + 0.6 \cdot 0.4\} = \max\{0.4, 0.6\} = 0.6 \end{aligned}$$

that is, the most likely value is 0.6 and corresponds to  $\neg c$ .

Before we further illustrate this formalism with tasks from the literature, we prove that 2AMC can be solved in polynomial time on  $\mathbf{X}_O$ -first sd-DNNFs. A similar result is already known for DTProbLog (Derkinderen and De Raedt 2020) and SC-ProbLog (Latour et al. 2017).

*Theorem 4 (Tractable 2AMC with  $\mathbf{X}_O$ -first sd-DNNFs)*

Let  $A = (\mathcal{T}, \mathbf{X}_I, \mathbf{X}_O, \alpha_I, \alpha_O, \mathcal{S}_I, \mathcal{S}_O, t)$  be a 2AMC instance, where  $\mathcal{T}$  is an  $\mathbf{X}_O$ -first sd-DNNF. Then, we can compute  $2AMC(A)$  in polynomial time in the size of  $\mathcal{T}$  assuming constant time semiring operations.

*Proof (Sketch)*

Consider a subgraph  $n$  of  $\mathcal{T}$  with exactly one outgoing edge for each or-node and all outgoing edges for each and-node. As  $\mathcal{T}$  is  $\mathbf{X}_O$ -first and smooth, there is a node  $n'$  in  $n$  such that  $Vars(n') = X_I$ , that is, exactly the outer variables occur above  $n'$  (see also the lowest and-nodes of the left NNF in Figure 1). Thus,  $n'$  is equivalent to  $\mathcal{T} \upharpoonright_{\mathbf{x}_O}$  for some assignment  $\mathbf{x}_O$  to the outer variables, for which  $n'$  computes the value of the inner AMC instance. As evaluation sums over all these subgraphs, it obtains the correct result.  $\square$

We illustrate 2AMC with three tasks from quantitative logic programming.

*Maximum a Posteriori Inference* A typical second level probabilistic inference task is *maximum a posteriori* inference, which involves maximizing over one set of variables while summing over another, as in Example 3.

*Definition 5 (The MAP task)*

**Given** a probabilistic logic program  $\mathcal{L}$ , a conjunction  $\mathbf{e}$  of observed literals for the set of evidence atoms  $\mathbf{E}$ , and a set of ground query atoms  $\mathbf{Q}$

**Find** the most probable assignment  $\mathbf{q}$  to  $\mathbf{Q}$  given the evidence  $\mathbf{e}$ , with  $\mathbf{R} = \mathbf{H} \setminus (\mathbf{Q} \cup \mathbf{E})$ :

$$MAP(\mathbf{Q}|\mathbf{e}) = \arg \max_{\mathbf{q}} P(\mathbf{Q} = \mathbf{q}|\mathbf{e}) = \arg \max_{\mathbf{q}} \sum_{\mathbf{r}} P(\mathbf{Q} = \mathbf{q}, \mathbf{e}, \mathbf{R} = \mathbf{r})$$

*MAP as 2AMC.* Solving MAP requires (1) summing probabilities over the truth values of the atoms in  $\mathbf{R}$  (with fixed truth values for atoms in  $\mathbf{E} \cup \mathbf{Q}$ ), and (2) determining truth values of the atoms in  $\mathbf{Q}$  that maximize this inner sum. Thus, we have  $\mathbf{X}_O = \mathbf{Q}$  and  $\mathbf{X}_I = \mathbf{E} \cup \mathbf{R}$ .

The inner problem corresponds to usual probabilistic inference, that is,  $\mathcal{S}_I = \mathcal{P}$  and  $\alpha_I$  assigns 1 to the literals in  $\mathbf{e}$  and 0 to their negations,  $p$  and  $1 - p$  to the positive and negative literals for probabilistic facts  $p::f$  that are not part of  $\mathbf{E}$ , and 1 to both literals for all other variables in  $\mathbf{X}_I$ .

We choose  $\mathcal{S}_O = (\mathbb{R}_{\geq 0} \times 2^{lit(\mathbf{Q})}, \oplus, \otimes, (0, \emptyset), (1, \emptyset))$  as the max-times semiring combined with the subsets of the query literals  $lit(\mathbf{Q})$  to remember the assignment that was used. Here,  $(r_1, S_1) \oplus (r_2, S_2)$  is  $(r_1, S_1)$  if  $r_1 > r_2$  and  $(r_2, S_2)$  if  $r_1 < r_2$ . To ensure commutativity, if  $r_1 = r_2$ , the sum is  $(r_1, \min_{>}(S_1, S_2))$  where  $>$  is some arbitrary but fixed total order on  $2^{lit(\mathbf{Q})}$ . The product  $(r_1, S_1) \otimes (r_2, S_2)$  is defined as  $(r_1 \cdot r_2, S_1 \cup S_2)$ . The weight function is given by  $\alpha_O(l) = (p, \{l\})$  if  $l = f$  for probabilistic fact  $p::f$ ,  $\alpha_O(l) = (1-p, \{l\})$  if  $l = \neg f$  for probabilistic fact  $p::f$ , and  $\alpha_O(l) = (1, \{l\})$  otherwise. The transformation function is the function  $t(p) = (p, \emptyset)$ .

*Maximizing Expected Utility* Another second level probabilistic task is *maximum expected utility* (Van den Broeck et al. 2010; Derkinderen and De Raedt 2020), which introduces an additional set of variables  $\mathbf{D}$  whose truth value can be arbitrarily chosen by a strategy  $\sigma(\mathbf{D}) = \mathbf{d}, \mathbf{d} \in int(\mathbf{D})$ , and is neither governed by probability nor logical rules. A *utility* function  $u$  maps each literal  $l$  to a reward  $u(l) \in \mathbb{R}$  for  $l$  being true.

*Definition 6 (Maximum Expected Utility (MEU) Task)*

**Given** A program  $\mathcal{L} = \mathbf{F} \cup R \cup \mathbf{D}$  with a utility function  $u$

**Find** a strategy  $\sigma^*$  that maximizes the expected utility:

$$\sigma^* = \arg \max_{\sigma} \sum_{M \in \text{mod}(\mathbf{F} \cup R \cup \sigma(\mathbf{D}))} \left( \prod_{l \in M} \alpha(l) \right) \left( \sum_{l \in M} u(l) \right),$$

where the label  $\alpha(l)$  of literal  $l$  is as defined for *SUCC*.

*Example 4 (cont.)*

Consider the program  $\mathcal{L}_{EU}$  obtained from  $\mathcal{L}_{ex}$  by replacing  $0.4::a$  by a decision variable  $?::a$ , with  $u(c) = 40$ ,  $u(\neg d) = 20$  and  $u(l) = 0$  for all other literals. Setting  $a$  to true, we have models  $\{a, b, c, d\}$  with probability 0.6 and utility 40 and  $\{a, c\}$  with probability 0.4 and utility 60, and thus expected utility  $0.6 \cdot 40 + 0.4 \cdot 60 = 48$ . Similarly, we have expected utility  $0.6 \cdot 0 + 0.4 \cdot 20 = 8$  for setting  $a$  to false. Thus, the MEU strategy sets  $a$  true.

*MEU as 2AMC.* Solving MEU involves (1) summing expected utilities of models over the non-decision variables (with fixed truth values for  $\mathbf{D}$ ), and (2) determining truth values of the atoms in  $\mathbf{D}$  that maximize this inner sum. Thus, we have  $\mathbf{X}_O = \mathbf{D}$  and  $\mathbf{X}_I = \mathbf{H} \setminus \mathbf{D}$ .

The inner problem is solved by the expected utility semiring  $\mathcal{S}_I = EU$ , with  $\alpha_I$  mapping literal  $l$  to  $(p_l, p_l \cdot u(l))$  if  $l$  is a probabilistic literal with probability  $p_l$ , and to  $(1, u(l))$  otherwise.

The basis for solving the outer problem is the max-plus semiring  $\mathcal{S}_{\max,+}$ , with  $\alpha_O$  the utility function  $u$ , and the transformation function  $t((p, pu)) = pu$ , if  $p \neq 0$  and  $t((0, pu)) = -\infty$ . This is extended to argmax using the same idea as for MAP.

*Probabilistic Inference with Stable Models* A more recent second level probabilistic task is probabilistic inference with stable model semantics (Totis et al. 2021; Skryagin et al. 2021). Inference of success probabilities reduces to a variant of weighted model counting, where the weight of a (stable) model of a program  $\mathcal{L} = R \cup \mathbf{F}$  is normalized with the number of models sharing the same assignment  $\mathbf{f}$  to the probabilistic facts  $\mathbf{F}$ :

$$SUCC^{sm}(q) = \sum_{\mathbf{h} \in \text{mod}(\mathcal{L} |_{\mathbf{f}}), \mathbf{f} \in \text{int}(\mathbf{F})} \frac{|\text{mod}(\mathcal{L} |_{\mathbf{f} \cup \{q\}})|}{|\text{mod}(\mathcal{L} |_{\mathbf{f}})|} \cdot \prod_{l \in \mathbf{h}} \alpha(l),$$

where the label  $\alpha(l)$  of literal  $l$  is as defined for *SUCC*.

*Example 5 (cont.)*

For  $\mathcal{L}_{ex}$ , each assignment  $\mathbf{f}$  introduces exactly one stable model, that is  $SUCC^{sm}$  equals *SUCC*. In the extended program  $\mathcal{L}_{sm} = \mathcal{L}_{ex} \cup \{e \leftarrow \text{not } f, f \leftarrow \text{not } e\}$ , however, all assignments have two stable models, one where  $e$  is added and one where  $f$  is added. Therefore, for each assignment  $\mathbf{F} = \mathbf{f}$  it holds that  $|\text{mod}(\mathcal{L}_{sm} |_{\mathbf{f}})| = 2$  but  $|\text{mod}(\mathcal{L}_{sm} |_{\mathbf{f} \cup \{e\}})| = 1$ . Thus  $SUCC^{sm}(e) = 0.5$ .

*SUCC<sup>sm</sup> as 2AMC.* Computing  $SUCC^{sm}$  requires (1) counting the number of models for a given total choice and those that satisfy the query  $q$ , and (2) summing the normalized probabilities. Thus, we have  $\mathbf{X}_O = \mathbf{F}$  and  $\mathbf{X}_I = \mathbf{H} \setminus \mathbf{F}$ .  $\mathcal{S}_I$  is the semiring over pairs of natural numbers,  $\mathcal{S}_I = (\mathbb{N}^2, +, \cdot, (0, 0), (1, 1))$ , where operations are component-wise.  $\alpha_I$

maps  $\neg q$  to  $(0, 1)$  and all other literals to  $(1, 1)$ . The first component thus only counts the models where  $q$  is true ( $|\text{mod}(\mathcal{L} \upharpoonright_{\mathbf{f} \cup \{q\}})|$ ), whereas the second component counts all models ( $|\text{mod}(\mathcal{L} \upharpoonright_{\mathbf{f}})|$ ). The transformation function is given by  $t((n_1, n_2)) = \frac{n_1}{n_2}$ . The outer problem then corresponds to usual probabilistic inference, that is,  $\mathcal{S}_O = \mathcal{P}$ , and  $\alpha_O$  assigns  $p$  and  $1 - p$  to the positive and negative literals for probabilistic facts  $p :: f$ , respectively, and 1 to all other literals.

#### 4 Weakening X-firstness

While any 2AMC problem can be solved in polynomial time on an  $\mathbf{X}_O$ -first sd-DNNF representing the logical theory  $\mathcal{T}$ , such an sd-DNNF can be much bigger than the smallest (ordinary) sd-DNNF for  $\mathcal{T}$ , as the  $\mathbf{X}$ -firstness may severely restrict the order in which variables are decided. In the following, we show that for a wide class of transformation functions, we can exploit the logical structure of the theory to relax the  $\mathbf{X}_O$ -first property.

Recall that a 2AMC task includes an AMC task for every assignment  $\mathbf{x}_O$  to the outer variables, which sums over all assignments  $\mathbf{x}_I$  to the inner variables that extend  $\mathbf{x}_O$  to a model of the theory. Consider the CNF  $\mathcal{T} = \{y \vee \neg x, \neg y \vee x\}$  and let  $\mathbf{X}_O = \{y\}$  and  $\mathbf{X}_I = \{x\}$ . The value of the outer variable  $y$  already determines the value of the inner variable  $x$ . Distributivity allows us to pull  $x$  out of every inner sum, as each such sum only involves one of the literals for  $x$ . If it does not matter whether we first apply the transformation function and then multiply or the other way around, we have a choice between keeping  $x$  in the inner semiring, or pushing its transformed version to the outer semiring. Thus, we can decide between an  $\mathbf{X}_O$ -first or an  $\mathbf{X}_O \cup \{x\}$ -first sd-DNNF. Naturally, the more such variables we have, the more freedom we gain.

This situation might also occur after we have already decided some of the variables. Consider the CNF  $\mathcal{T}' = \{z \vee y \vee \neg x, z \vee \neg y \vee x\}$  and let  $\mathbf{X}_O = \{z, y\}$  and  $\mathbf{X}_I = \{x\}$ . If we set  $z$  to true, both  $y$  and  $x$  can take any value, therefore the value of  $x$  is not determined by  $z$  and  $y$ . However, if we set  $z$  to false, we are in the same situation as above and can move  $x$  to  $\mathbf{X}_O$  on a local level.

We formalize this, starting with definability to capture when a variable is determined by others.

*Definition 7 (Definability Lagniez et al. 2016)*

A variable  $a$  is *defined* by a set of variables  $\mathbf{X}$  with respect to a theory  $\mathcal{T}$  if for every assignment  $\mathbf{x}$  of  $\mathbf{X}$  it holds that  $\mathbf{x} \cup \mathcal{T} \models a$  or  $\mathbf{x} \cup \mathcal{T} \models \neg a$ . We denote the set of variables that are not in  $\mathbf{X}$  and defined by  $\mathbf{X}$  with respect to  $\mathcal{T}$  by  $\mathbf{D}(\mathcal{T}, \mathbf{X})$ .

*Example 6 (cont.)*

In  $\mathcal{L}_{ex}$  the atoms  $c$  and  $d$  are defined by  $\{a, b\}$  since  $c$  holds if  $a$  holds, and  $d$  holds if  $b$  holds.

*Definition 8 (X-Firstness Modulo Definability)*

Given an NNF  $n$  on variables partitioned into  $\mathbf{X}, \mathbf{Y}$ , we say an internal node  $n_i$  of  $n$  is *pure modulo definability* if  $\text{Vars}(n_i) \subseteq \mathbf{X} \cup \mathbf{D}(n_i, \mathbf{X})$  or  $\text{Vars}(n_i) \subseteq \mathbf{Y}$  and *mixed modulo definability*, otherwise.  $n$  is an  $\mathbf{X}$ -first NNF *modulo definability*,  $\mathbf{X}/\mathbf{D}$ -first NNF for short, if for each of its and-nodes  $n_i$  either all children of  $n_i$  are pure modulo definability, or one

child of  $n_i$  is mixed modulo definability and all other children  $n_j$  of  $n_i$  are pure modulo definability and  $\text{Vars}(n_j) \subseteq \mathbf{X} \cup \mathbf{D}(n_i, \mathbf{X})$ .

The intuition here is that we can decide variables from  $\mathbf{Y}$  earlier, if they are defined by the variables in  $\mathbf{X}$  in terms of the theory  $\mathcal{T}$  conditioned on the decisions we have already made. Thus, we only decide the variables in  $\mathbf{X}$ -first *modulo definability*.

*Example 7 (cont.)*

In Figure 1, the left NNF is an  $\{a, b\}$ -first NNF and therefore also an  $\{a, b\}/\mathbf{D}$ -first NNF. The right NNF is not an  $\{a, b\}$ -first NNF but an  $\{a, b\}/\mathbf{D}$ -first NNF since  $\mathbf{D}(\mathcal{L}_{ex}, \{a, b\})$  contains  $c$ .

The following lemma generalizes this example from 2 to  $n$  pairs of equivalent variables.

*Lemma 9 (Exponential Separation)*

Let  $\mathcal{T} = \bigwedge_{i=1}^n X_i \leftrightarrow Y_i$ ,  $\mathbf{X} = \{X_1, \dots, X_n\}$  and  $\mathbf{D} = \{Y_1, \dots, Y_n\}$ , then the size of the smallest  $\mathbf{X}$ -first sd-DNNF for  $\mathcal{T}$  is exponential in  $n$  and the size of the smallest  $\mathbf{X}/\mathbf{D}$ -first sd-DNNF for  $\mathcal{T}$  is linear in  $n$ .

*Proof (Sketch)*

Since  $\mathbf{D}(\mathcal{T}, \mathbf{X}) = \mathbf{Y}$ , every sd-DNNF for  $\mathcal{T}$  is an  $\mathbf{X}/\mathbf{D}$ -first sd-DNNF. As  $\mathcal{T}$  has treewidth 2, there exists an sd-DNNF of linear size. On the other hand, an  $\mathbf{X}$ -first sd-DNNF must contain a node that is equivalent to  $\mathcal{T} \upharpoonright_{\mathbf{x}}$  for each of the  $2^{|\mathbf{X}|}$  assignments  $\mathbf{x} \in \text{int}(\mathbf{X})$ .  $\square$

We see that  $\mathbf{X}/\mathbf{D}$ -first sd-DNNFs can be much smaller than  $\mathbf{X}$ -first sd-DNNFs, even on very simple propositional theories. It remains to show that we maintain tractability. As in the beginning of this section, we want to regard defined inner variables as outer variables. For this to work it must not matter whether we first multiply and then apply the transform  $t$  or the other way around, that is,  $t$  must be a *homomorphism* for the multiplications of the semirings.

*Definition 10 (Monoid Homomorphism, Generated Monoid)*

Let  $\mathcal{M}_i = (M_i, \odot^i, e_{\odot^i})$  for  $i = 1, 2$  be monoids. Then a *monoid homomorphism* from  $\mathcal{M}_1$  to  $\mathcal{M}_2$  is a function  $f : M_1 \rightarrow M_2$  such that

$$\text{for all } m, m' \in M_1 \quad f(m \odot^1 m') = f(m) \odot^2 f(m') \quad \text{and} \quad f(e_{\odot^1}) = e_{\odot^2}.$$

Furthermore, for a subset  $M' \subseteq M$  of a monoid  $\mathcal{M} = (M, \odot, e_{\odot})$  the *monoid generated by  $M'$* , denoted  $\langle M' \rangle_{\mathcal{M}}$ , is  $\mathcal{M}^* = (M^*, \odot, e_{\odot})$ , where  $M' \subseteq M^*$  and  $M^*$  is the subset minimal set such that  $\mathcal{M}^*$  is a monoid.

*Example 8 (cont.)*

Consider again the 2AMC instance  $A_{ex}$  from Example 3. Since  $a$  is defined in terms of  $c$ , we want to argue that the following equality holds, allowing us to see  $a$  as an outer variable:

$$\begin{aligned} & \max_{\mathbf{x}_O \in \text{int}(\{c\})} \alpha_O(\mathbf{x}_O) \cdot \text{id} \left( \sum_{\mathbf{x}_I \in \text{mod}(\mathcal{L}_{ex} \upharpoonright_{\mathbf{x}_O})} \prod_{y \in \mathbf{x}_I} \alpha_I(y) \right) \\ &= \max_{\mathbf{x}_O \in \text{int}(\{c\}), \mathbf{x}_O' \in \text{int}(\{a\})} \alpha_O(\mathbf{x}_O) \cdot \text{id}(\alpha_I(\mathbf{x}_O')) \cdot \text{id} \left( \sum_{\mathbf{x}_I \in \text{mod}(\mathcal{L}_{ex} \upharpoonright_{\mathbf{x}_O \cup \mathbf{x}_O'})} \prod_{y \in \mathbf{x}_I} \alpha_I(y) \right) \end{aligned}$$

Here, this is easy to see since  $id$  is a homomorphism between the monoid  $([0, 1], \cdot)$  of the inner probability semiring and the monoid  $(\mathbb{R}_{\geq 0}, \cdot)$  of the outer max-times semiring, as  $id : [0, 1] \rightarrow \mathbb{R}_{\geq 0}$ , for any  $p, q \in [0, 1]$ ,  $id(p \cdot q) = p \cdot q = id(p) \cdot id(q)$ , and  $id(1) = 1$ .

In general, instead of applying the transform to a sum of products of literal labels for a set of variables, we want to apply it independently to (1) the literal labels of a defined variable and (2) the inner sum restricted to the remaining variables. It is therefore sufficient if the equality is valid for the monoid generated by the values we encounter in these situations, rather than for all values from the inner monoid’s domain. As we will illustrate for MEU at the end of this section, the transform of some 2AMC tasks only satisfies this weaker but sufficient condition. The following definition captures this idea, where the two subsets of  $O(A)$  correspond to the values observed in cases (1) and (2), respectively:

*Definition 11 (Observable Values)*

Let  $A = (\mathcal{T}, \mathbf{X}_I, \mathbf{X}_O, \alpha_I, \alpha_O, \mathcal{S}_I, \mathcal{S}_O, t)$  be a 2AMC instance. Then the set of *observable values* of  $A$ , denoted  $O(A)$  is

$$\{\alpha_I(l) \mid \mathbf{x}_O \in \text{int}(\mathbf{X}_O), y \in \mathbf{D}(\mathcal{T} \cup \mathbf{x}_O, \mathbf{X}_O), l \in \{y, \neg y\}\} \\ \cup \{\bigoplus_{\mathbf{x}_I \in \text{mod}(\mathcal{T} \upharpoonright_{\mathbf{x}_O})}^I \bigotimes_{y \in \mathbf{x}_I}^I \alpha_I(y) \mid \mathbf{x}_O \in \text{int}(\mathbf{X}_O \cup \mathbf{D}^*), \mathbf{D}^* \subseteq \mathbf{D}(\mathcal{T} \cup \mathbf{x}_O, \mathbf{X}_O)\}.$$

With this in mind, we can state our main result.

*Theorem 12 (Tractable AMC with X/D-first sd-DNNFs)*

The value of a 2AMC instance  $A = (\mathcal{T}, \mathbf{X}_I, \mathbf{X}_O, \alpha_I, \alpha_O, \mathcal{S}_I, \mathcal{S}_O, t)$  can be computed in polynomial time, assuming constant time semiring operations, under the following conditions:

- $\mathcal{T}$  is an  $\mathbf{X}/\mathbf{D}$ -first sd-DNNF
- $t$  is a homomorphism from the monoid  $\langle O(A) \rangle_{(R_I, \otimes^I, e_{\otimes^I})}$  generated by the observable values to  $(R_O, \otimes^O, e_{\otimes^O})$ .

*Proof (Sketch)*

The proof of this theorem exploits (a) distributivity and the form of the transformation function to move defined variables to the outer semiring as outlined at the start of this section, and (b) the fact that when we decide an outer variable, then we get two new 2AMC instances, where the theory  $\mathcal{T}$  is conditioned on the truth of the decided variable. On these new instances, we can also use definability in the same fashion as before.  $\square$

Despite the fact that checking which variables are defined for each partial assignment  $\mathbf{x}$  is not feasible, as checking definability is co-NP-complete and there are more than  $2^{|\mathbf{X}_O|}$  partial assignments, this result does have implications for constrained KC in practice.

Firstly, we can check which variables are defined by  $\mathbf{X}_O$  in terms of the whole theory  $\mathcal{T}$ , and use this to generate a variable order for compilation that leads to an  $\mathbf{X}/\mathbf{D}$ -first sd-DNNF with a priori guarantees on its size. We discuss this in Section 5.

Secondly, as entailment is a special case of definability, Theorem 12 justifies the use of unit propagation during compilation, which dynamically adapts the variable order when variables are entailed by the already decided ones, and thus may violate  $\mathbf{X}$ -firstness.

We still need to verify that our three example tasks satisfy the preconditions of Theorem 12, that is, their transformation functions are monoid homomorphisms on the observable values. For MAP and  $\text{SUCC}^{\text{sm}}$  this is easy to prove, as the transformation is already a homomorphism on all values. For MEU, however, the restriction to the monoid generated by the observable values is crucial, as the transformation function is only a homomorphism for tuples  $(p, pu)$  with  $p \in \{0, 1\}$ , which may not be the case in general. In the DTProbLog setting, however, assignments to decision facts and probabilistic facts are independent, and every assignment to both sets extends to a single model of the theory. Together with the  $(1, u)$ -labels of the remaining atoms, this ensures that the result of the inner sum is of the form  $(1, x)$ , and MEU thus meets the criteria.

### 5 Implementation

The general pipeline of PLP solvers takes a program, grounds it and optionally simplifies it or breaks its cycles. Using standard KC tools, this program is then compiled into a tractable circuit either directly or via conversion to CNF. We extended this pipeline in *aspmc* (Eiter et al. 2021) and *smProbLog* (Totis et al. 2021) to compile programs, via CNF, into  $\mathbf{X}/\mathbf{D}$ -first circuits.

To obtain  $\mathbf{X}/\mathbf{D}$ -first circuits we need to specify a variable order in which all variables in  $\mathbf{X}$  are decided first modulo definedness. Preferably, the chosen order should also result in efficient compilation and thus a small circuit. Korhonen and Järvisalo (2021) have shown how to generate variable orders from *tree decompositions* of the primal graph of the CNF that result in (unconstrained) sd-DNNFs whose size is bounded by the *width* of the decomposition. We adapt this result to our constrained setting, where we need to ensure that the tree decomposition satisfies an additional property that allows compilation to essentially consider the non-defined inner variables and the outer variables independently. We first define the necessary concepts.

*Definition 13 (Primal Graph, Tree Decomposition)*

Let  $\mathcal{T}$  be a CNF over variables  $\mathbf{X}$ . The *primal graph* of  $\mathcal{T}$ , denoted by  $\text{PRIM}(\mathcal{T})$ , is defined as  $V(\text{PRIM}(\mathcal{T})) = \mathbf{X}$  and  $(v_1, v_2) \in E(\text{PRIM}(\mathcal{T}))$  if  $v_1, v_2$  occur together in some clause of  $\mathcal{T}$ .

A *tree decomposition (TD)* for a graph  $G$  is a pair  $(T, \chi)$ , where  $T$  is a tree and  $\chi$  is a labeling of  $V(T)$  by subsets of  $V(G)$  s.t. (i) for all nodes  $v \in V(G)$  there is  $t \in V(T)$  s.t.  $v \in \chi(t)$ ; (ii) for every  $(v_1, v_2) \in E(G)$  there exists  $t \in V(T)$  s.t.  $v_1, v_2 \in \chi(t)$  and (iii) for all nodes  $v \in V(G)$  the set of nodes  $\{t \in V(T) \mid v \in \chi(t)\}$  forms a connected subtree of  $T$ . The width of  $(T, \chi)$  is  $\max_{t \in V'} |\chi(t)| - 1$ .

Next, we show that restricted TDs allow  $\mathbf{X}/\mathbf{D}$ -first compilation with performance guarantees:

*Lemma 14*

Let  $\mathcal{T}$  be a CNF over variables  $\mathbf{Y}$  and  $(T, \chi)$  a TD of  $\text{PRIM}(\mathcal{T})$  of width  $k$ . Furthermore, let  $\mathbf{X} \subseteq \mathbf{Y}$  and  $\mathbf{D} = \mathbf{D}(\mathcal{T}, \mathbf{X})$ . If there exists  $t^* \in V(T)$  such that (1)  $\chi(t^*) \subseteq \mathbf{X} \cup \mathbf{D}$  and (2)  $\chi(t^*)$  is a *separator* of  $\mathbf{X}$  and  $\mathbf{Y} \setminus (\mathbf{X} \cup \mathbf{D})$ , that is, every path from  $\mathbf{X}$  to  $\mathbf{Y} \setminus (\mathbf{X} \cup \mathbf{D})$  in  $\text{PRIM}(\mathcal{T})$  uses a vertex from  $\chi(t^*)$ , then we can compile  $\mathcal{T}$  into an  $\mathbf{X}/\mathbf{D}$ -first sd-DNNF in time  $\mathcal{O}(2^k \cdot \text{poly}(|\mathcal{T}|))$ .

*Proof (Sketch).*

The performance guarantee is due to Korhonen and Järvisalo (2021) and holds when we decide the variables in the order they occur in the TD starting from the root.  $\mathbf{X}/\mathbf{D}$ -firstness can be guaranteed by taking  $t^*$  as the root of the TD and, thus, first deciding all variables in  $\chi(t^*)$ . From condition (2) it follows that afterwards the CNF has decomposed into separate components, which either only use variables from  $\mathbf{X} \cup \mathbf{D}$  or use no variables from  $\mathbf{X}$ . Thus, their compilation only leads to pure NNFs.  $\square$

To find a TD of small width for which the lemma applies, we proceed as follows. Given a CNF  $\mathcal{T}$  and partition  $\mathbf{X}_I, \mathbf{X}_O$  of the variables in inner and outer variables, we first compute  $\mathbf{D}(\mathcal{T}, \mathbf{X}_O)$  by using an NP-oracle. Lagniez et al. (2016) showed that this is possible. As the width of a suitable TD is at least the size of the separator minus one, we first approximate a minimum size separator  $\mathbf{S} \subseteq \mathbf{X} \cup \mathbf{D}(\mathcal{T}, \mathbf{X})$  using clingo (Gebser et al. 2014) with a timeout of 30 s. To ensure that the TD contains a node  $t'$  with  $\mathbf{S} \subseteq \chi(t')$ , we add the clique over the vertices in  $\mathbf{S}$  to the primal graph, that is, we generate a tree decomposition  $(T, \chi)$  of  $\text{PRIM}(\mathcal{T}) \cup \text{Clique}(\mathbf{S})$ . For this, we use flow-cutter (Dell et al. 2017) with a timeout of 5 s. The resulting TD either already contains a node with  $\mathbf{S} = \chi(t')$  and thus satisfies the preconditions of Lemma 14, or can be modified locally to one that does by splitting the node with  $\mathbf{S} \subset \chi(t')$ .

*aspmc:*<sup>1</sup> We use the above strategy to generate a variable order, which is then given to c2d (Darwiche 2004) or miniC2D (Oztok and Darwiche 2015) together with the CNF to compile an  $\mathbf{X}/\mathbf{D}$ -first circuit, on which we evaluate the 2AMC task. We stress that c2d – contrary to miniC2D – always uses unit propagation during compilation, and thus can only be used due to Theorem 12.

Currently, *aspmc* supports DTProbLog, MAP and SMProbLog programs as inputs.

*SMProbLog:*<sup>2</sup> The SMProbLog implementation of Totis et al. (2021) uses DSHARP (Aziz et al. 2015) to immediately compile the logic program to a d-DNNF, which prevents a direct application of our new techniques. Our adapted version obtains a CNF  $\mathcal{T}$  and variable ordering as in *aspmc*, which it then compiles to SDD (Darwiche 2011) using the *PySDD* library<sup>3</sup> as in standard ProbLog. SDDs can be seen as a special case of sd-DNNFs. The main difference is that for each branch the variables are decided in the same order.

## 6 Experimental evaluation

Our experimental evaluation addresses the following questions:

- Q1: How does exploiting definedness influence the efficiency of 2AMC solving?
- Q2: How does *aspmc* compare to task-specific solvers from the PLP literature?
- Q3: How does our second level approach compare to the first level approach when definedness reduces 2AMC to AMC?

<sup>1</sup> *aspmc* is open source and available at [github.com/raki123/aspmc](https://github.com/raki123/aspmc).

<sup>2</sup> SMProbLog is open source and available at [github.com/PietroTotis/smProbLog](https://github.com/PietroTotis/smProbLog).

<sup>3</sup> [github.com/wannesm/PySDD](https://github.com/wannesm/PySDD).

## 6.1 General setup

To answer these questions, we consider logic programs from the literature on the three example 2AMC tasks. To eliminate differences on how different solvers handle n-ary random choices (known as annotated disjunctions) and 0/1-probabilities, we normalize probabilistic programs to contain only probabilistic facts and normal clauses, and replace all probabilities by values chosen uniformly at random from 0.1, 0.2, ..., 0.8, 0.9.

For MAP, we use the growing head, growing negated body, blood and graph examples of Bellodi *et al.* (2020), with a subset of the probabilistic facts of uniformly random size as query atoms and the given evidence. For MEU, we use the Bayesian networks provided by Derkinderen and De Raedt (2020) as well as the viral marketing example from Van den Broeck *et al.* (2010) on randomly generated power law graphs that are known to resemble social networks (Barabási and Bonabeau 2003). For  $\text{SUCC}^{\text{sm}}$ , we use an example from Totis *et al.* (2021) that introduces non-probabilistic choices into the well-known smokers example (Fierens *et al.* 2015).

Besides this basic set, we also use the original smokers example, where  $\text{SUCC}^{\text{sm}}$  reduces to SUCC, for Q3. For Q1, we use the grid graphs of Fierens *et al.* (2015) as an additional MAP benchmark. Here, we control definedness by choosing the MAP queries as the probabilistic facts for the edges that are reachable in  $k$  steps from the top left corner, for all possible values of  $k$ , and use the existence of a path from the top left to the bottom right corner as evidence.

We compare the following systems:

**aspmc** with c2d as default knowledge compiler

**ProbLog** in different versions: ProbLog2 (version 2.1.0.42) in MAP and default (SUCC) mode, the implementation of Derkinderen and De Raedt (2020) for MEU, and our implementation of SMProbLog for  $\text{SUCC}^{\text{sm}}$

**PITA** (Bellodi *et al.* 2020, version 4.5, included in SWI-Prolog) for MAP and MEU

**clingo** (Gebser *et al.* 2014, version 5.5.0.post3) as an indicator of how an enumeration based approach not using knowledge compilation might perform. Note that this approach does not actually compute the 2AMC value of the formula, but only enumerates models.

We limit each individual run of a system to 300 s and 4Gb of memory. When plotting running time per instance for a specific solver, we always sort instances by ascending time for that solver, using 300 s for any instance that did not finish successfully within these bounds.

The instances, results and benchmarking scripts are available at [github.com/raki123/CC](https://github.com/raki123/CC).

## 6.2 Results

**Q1:** *How does exploiting definedness influence the efficiency of 2AMC solving?* To answer the first question, we use all proper 2AMC benchmarks, and focus on the 2AMC task itself, that is, we start from the labeled CNF corresponding to the instance. We consider four different settings obtained by independently varying two dimensions: constraining compilation to either **X**-first or **X/D**-first, and compiling to either sd-DNNF using c2d or to SDD using miniC2D.

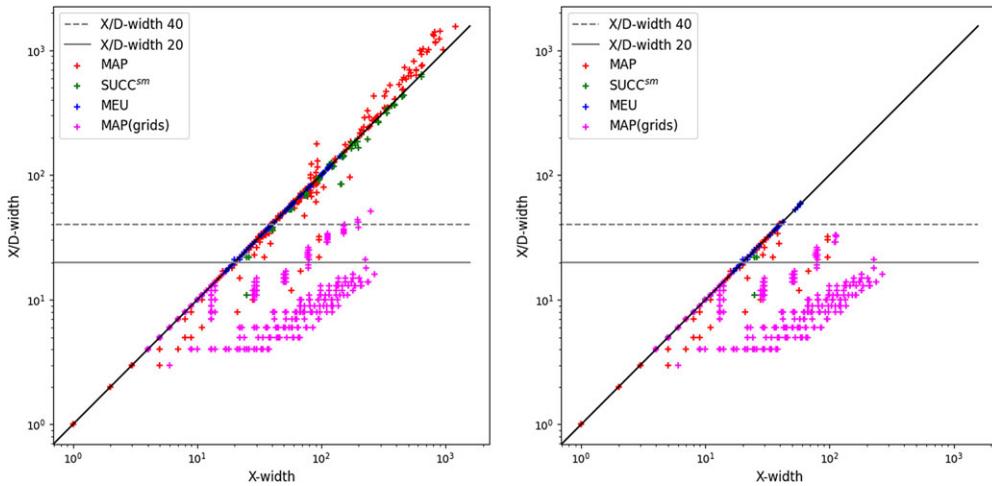


Fig. 2. Q1: Comparison of tree decomposition width for  $\mathbf{X}$ -first and  $\mathbf{X}/\mathbf{D}$ -first variable order, across all 2AMC instances (left) and for solved 2AMC instances only (right). We solve 822 of the 825 instances with  $\mathbf{X}/\mathbf{D}$ -width at most 20, 118 of the 219 instances with  $\mathbf{X}/\mathbf{D}$ -width between 21 and 40, and 7 of the 353 instances with  $\mathbf{X}/\mathbf{D}$ -width above 40.

On the left of Figure 2, we plot the width of the tree decompositions the solver uses to determine the variable order in the  $\mathbf{X}$ -first or  $\mathbf{X}/\mathbf{D}$ -first case, respectively. Recall from Lemma 14 that this width appears in the exponent of the compilation time bound. The *optimal* tree decomposition's width in the  $\mathbf{X}/\mathbf{D}$ -first case is at most that of the  $\mathbf{X}$ -first case, and would thus result in points on or below the black diagonal only. In practice, we observe many points close to the diagonal, with two notable exceptions. MAP instances with high width tend to be slightly above the diagonal, whereas MAP grids are mostly clearly below the diagonal. These results can be explained by the shape of the problems and the fact that we only approximate the optimal decomposition, as this is a hard task. We note that for many of the benchmarks, the amount of variables defined in terms of the outer variables (decision variables for MEU, query variables for MAP, probabilistic facts for  $\text{SUCC}^{SM}$ ) is limited. The exception are the MAP grids, where the choice of queries entails definedness.

We plot the same data restricted to the instances solved within the time limit on the right of Figure 2, along with summary statistics on the number of instances solved for three ranges of  $\mathbf{X}/\mathbf{D}$ -width in the caption. We observe that almost no instances with  $\mathbf{X}/\mathbf{D}$ -width above 40 are solved. At the same time, almost all instances with  $\mathbf{X}/\mathbf{D}$ -width below 20 are solved, including many cases with  $\mathbf{X}$ -width above 40, where we thus see a clear benefit from exploiting definedness.

In Figure 3, we plot the running times per instance for the different solvers. Given the width results, we distinguish between MAP grids and the remaining cases. On the grids, taking into account definedness results in clear performance gains when compiling SDDs (using miniC2D). On the other hand, compiling sd-DNNFs with c2d shows only a marginal difference between  $\mathbf{X}$  and  $\mathbf{X}/\mathbf{D}$  variable orders. The reason is that c2d implicitly exploits definedness even when given the  $\mathbf{X}$ -first order through its use of unit propagation. SDD compilation, on the other hand, cannot deviate from the given order, and only

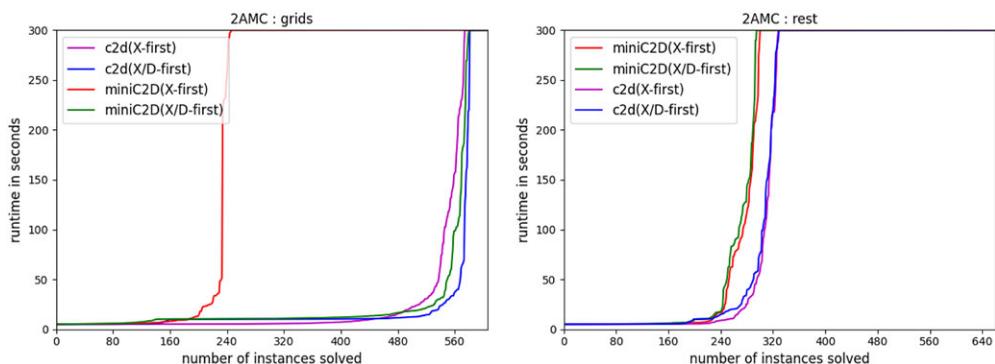


Fig. 3. Q1: Running times per instance for different configurations on MAP grids (left) and all other 2AMC instances (right).

benefits from definedness if it is reflected in the variable order. On the other benchmarks, with fewer defined variables, the variable order has little effect within the same circuit class, but sd-DNNFs outperform SDDs, likely because unit propagation can also exploit context-dependent definedness. In the following we thus use *c2d*.

**Q2:** *How does *aspmc* compare to task-specific solvers from the PLP literature?* We first consider the efficiency of the whole pipeline from instance to solution on the MAP and MEU tasks, which are addressed by both ProbLog and PITA. From the plots of running times in Figure 4, we observe that all solvers outperform clingo's model enumeration. ProbLog is slower than both *aspmc* and PITA except on the MAP graphs. Among *aspmc* and PITA, *aspmc* outperforms PITA on MEU, and vice versa on MAP. This can be explained by the different overall approach to knowledge compilation taken in the various solvers. *aspmc* always compiles a CNF encoding of the whole ground program, including all ground atoms, ProbLog compiles a similar encoding, but restricted to the part of the formula that is relevant to the task at hand, while PITA directly compiles a circuit for the truth of atoms of interest in terms of the relevant ground choice atoms. As MAP queries are limited to probabilistic facts, this allows PITA to compile a circuit for the truth of the evidence in terms of relevant probabilistic facts only, which especially for the graph setting can be significantly smaller than a complete encoding. For MEU, PITA needs one circuit per atom with a utility, which additionally need to be combined, putting PITA at a disadvantage.

For  $\text{SUCC}^{\text{sm}}$ , we plot running times on the modified smokers setting for the clingo baseline, *aspmc* and three variants of the dedicated ProbLog implementation, namely the original implementation of Totis *et al.* (2021) that compiles to d-DNNF as well as our modified implementation compiling to **X**-first and **X/D**-first SDDs. These problems appear to be hard in general, but we observe a clear benefit from the constrained compilation enabled in our approach.

**Q3:** *How does our second level approach compare to the first level approach when definedness reduces 2AMC to AMC?* On the regular smokers benchmark, SMProbLog and ProbLog semantics coincide, that is, all inner variables of the 2AMC task are defined,

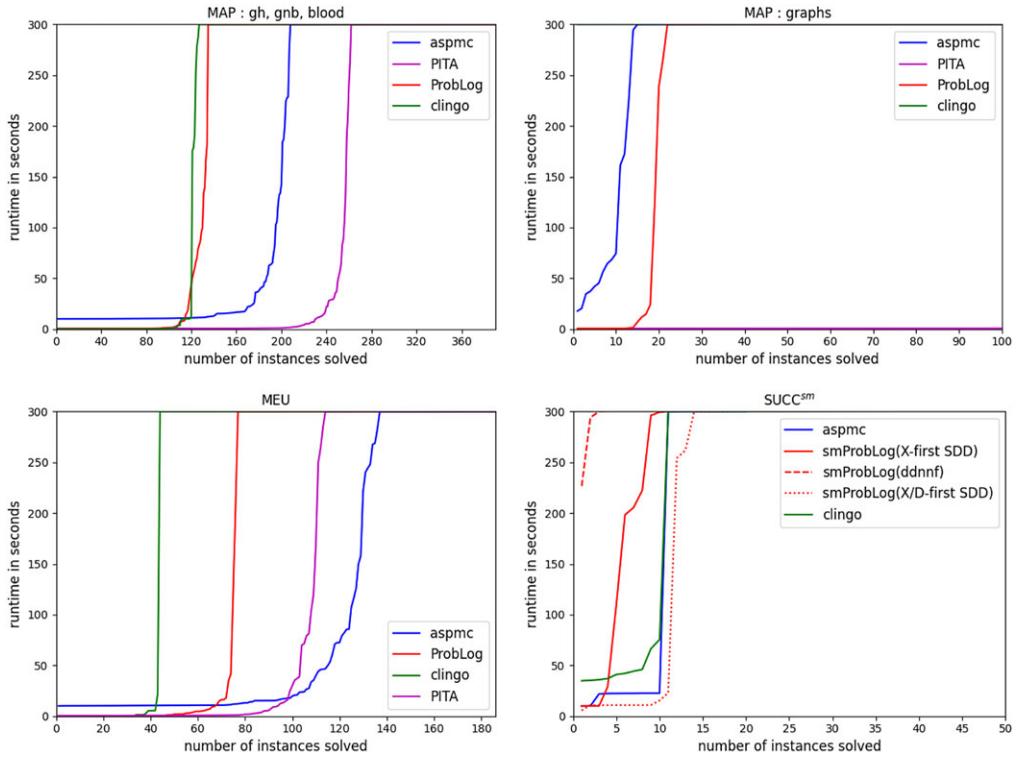


Fig. 4. Q2: Running times of different solvers on MAP problem sets (top), indicated above each plot, MEU problems (bottom left) and SUCC<sup>SM</sup> (bottom right).

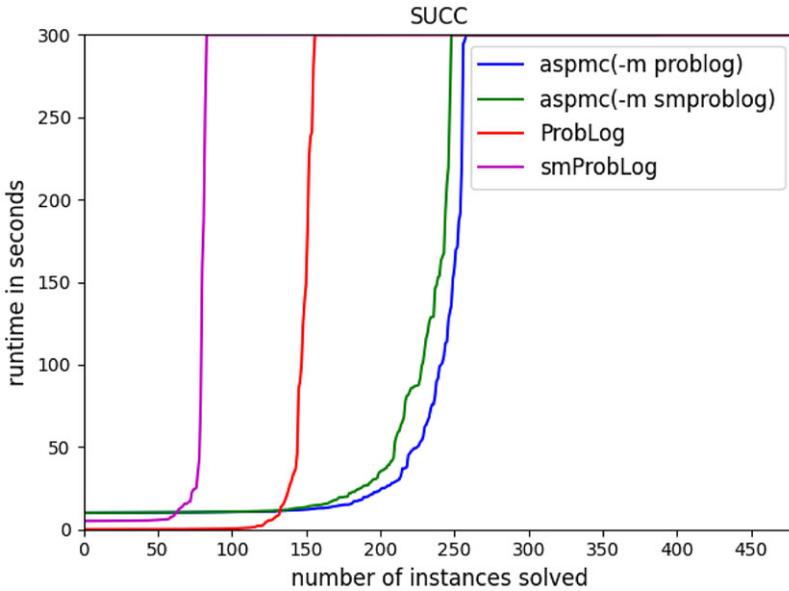


Fig. 5. Q3: Running times of different solvers on SUCC programs.

and it thus reduces to AMC. In Figure 5, we plot running times for the AMC and 2AMC variants of both *aspmc* and ProbLog. For ProbLog, there is a clear gap between the two approaches, which is at least in part due to the fact that ProbLog only compiles the relevant part of the program, whereas *SMProbLog* compiles the full theory. *aspmc* outperforms ProbLog on the harder instances, with limited overhead for the second level task.

## 7 Conclusion

2AMC is a hard problem, even harder than #SAT or AMC in general, as it imposes significant constraints on variable orders in KC. Our theoretical results show that these constraints can be weakened by exploiting definedness of variables. In practice, this allows us to (i) introduce a strategy to construct variable orders for compilation into **X/D**-first sd-DNNFs with a priori guarantees on complexity, and (ii) to use unit propagation to decide literals earlier than specified by the variable order during compilation. Our experimental evaluation shows that (ii) generally improves the performance and (i) can boost it when many variables are defined. Furthermore, we see that compilation usually performs much better than an enumeration based approach to solve 2AMC. Last but not least, our extensions of *aspmc* and *SMProbLog* are competitive with PITA and ProbLog, the state of the art solvers for MAP, MEU and  $\text{SUCC}^{\text{sm}}$  inference for logic programs, and even exhibit improved performance on MEU and  $\text{SUCC}^{\text{sm}}$ .

## Supplementary material

To view supplementary material for this article, please visit <http://doi.org/10.1017/S147106842200014X>.

## References

- AZIZ, R. A., CHU, G., MUISE, C. J. AND STUCKEY, P. J. 2015. Stable model counting and its application in probabilistic logic programming. In *AAAI*, 3468–3474.
- BARABÁSI, A.-L. AND BONABEAU, E. 2003. Scale-free networks. *SCI AM* 288, 5, 60–69.
- BARAL, C., GELFOND, M. AND RUSHTON, N. 2009. Probabilistic reasoning with answer sets. *TPLP* 9, 1, 57–144.
- BELLODI, E., ALBERTI, M., RIGUZZI, F. AND ZESE, R. 2020. MAP inference for probabilistic logic programming. *TPLP* 20, 5, 641–655.
- DARWICHE, A. 2004. New advances in compiling CNF into decomposable negation normal form. In *ECAI*, 328–332.
- DARWICHE, A. 2011. SDD: A new canonical representation of propositional knowledge bases. In *IJCAI*.
- DELL, H., KOMUSIEWICZ, C., TALMON, N. AND WELLER, M. 2017. PACE: The second iteration. In *International Symposium on Parameterized and Exact Computation (IPEC)*, 30:1–30:13.
- DE RAEDT, L., KIMMIG, A. AND TOIVONEN, H. 2007. ProbLog: A probabilistic Prolog and its application in link discovery. In *IJCAI*, vol. 7, 2462–2467.
- DERKINDEREN, V. AND DE RAEDT, L. 2020. Algebraic circuits for decision theoretic inference and learning. In *ECAI*, vol. 325, 2569–2576.

- EITER, T., HECHER, M. AND KIESEL, R. 2021. Treewidth-aware cycle breaking for algebraic answer set counting. In *KR*, 269–279.
- FIERENS, D., DEN BROECK, G. V., RENKENS, J., SHTERIONOV, D. S., GUTMANN, B., THON, I., JANSSENS, G. AND DE RAEDT, L. 2015. Inference and learning in probabilistic logic programs using weighted boolean formulas. *TPLP* 15, 3, 358–401.
- GEBBER, M., KAMINSKI, R., KAUFMANN, B. AND SCHAUB, T. 2014. Clingo= ASP+ control: Preliminary report. arXiv preprint arXiv:1405.3694.
- KIMMIG, A., VAN DEN BROECK, G. AND DE RAEDT, L. 2011. An algebraic prolog for reasoning about possible worlds. In *AAAI*.
- KIMMIG, A., VAN DEN BROECK, G. AND DE RAEDT, L. 2017. Algebraic model counting. *JAL* 22, 46–62.
- KORHONEN, T. AND JÄRVISALO, M. 2021. Integrating tree decompositions into decision heuristics of propositional model counters (short paper). In *CP*. LIPIcs, vol. 210, 8:1–8:11.
- LAGNIEZ, J.-M., LONCA, E. AND MARQUIS, P. 2016. Improving model counting by leveraging definability. In *IJCAI*, 751–757.
- LATOUR, A. L. D., BABAKI, B., DRIES, A., KIMMIG, A., VAN DEN BROECK, G. AND NIJSSEN, S. 2017. Combining stochastic constraint optimization and probabilistic programming - from knowledge compilation to constraint solving. In *CP*, 495–511.
- LEE, J. AND YANG, Z. 2017. LPMLN, weak constraints, and P-log. In *AAAI*.
- MANHAEVE, R., DUMANCIC, S., KIMMIG, A., DEMEESTER, T. AND DE RAEDT, L. 2018. Deep-ProbLog: Neural probabilistic logic programming. In *NeurIPS*, 3749–3759.
- OZTOK, U. AND DARWICHE, A. 2015. A top-down compiler for sentential decision diagrams. In *IJCAI*, 3141–3148.
- RIGUZZI, F. AND SWIFT, T. 2011. The pita system: Tabling and answer subsumption for reasoning under uncertainty. *TPLP* 11, 4-5, 433–449.
- SKRYAGIN, A., STAMMER, W., OCHS, D., DHAMI, D. S. AND KERSTING, K. 2021. SLASH: Embracing probabilistic circuits into neural answer set programming. *CoRR*, abs/2110.03395.
- TOTIS, P., KIMMIG, A. AND DE RAEDT, L. 2021. SMProbLog: Stable model semantics in ProbLog and its applications in argumentation. *StarAI*, abs/2110.01990.
- VAN DEN BROECK, G., THON, I., VAN OTTERLO, M. AND DE RAEDT, L. 2010. DTPProbLog: A decision-theoretic probabilistic Prolog. In *AAAI*.
- YANG, Z., ISHAY, A. AND LEE, J. 2020. Neurasp: Embracing neural networks into answer set programming. In *IJCAI*, 1755–1762.