Knowledge Compilation Properties of Trees-of-BDDs, Revisited*

Hélène Fargier
IRIT-CNRS, UMR 5505
Université de Toulouse, France
fargier@irit.fr

Pierre Marquis
CRIL-CNRS, UMR 8188
Université Lille-Nord de France, Artois, France
marquis@cril.univ-artois.fr

Abstract

Recent results have shown the interest of trees-of-BDDs [Subbarayan *et al.*, 2007] as a suitable target language for propositional knowledge compilation from the practical side. In the present paper, the concept of tree-of-BDDs is extended to additional classes of data structures C thus leading to trees-of-C representations (ToC). We provide a number of generic results enabling one to determine the queries/transformations satisfied by ToC depending on those satisfied by C. We also present some results about the spatial efficiency of the ToC languages. Focusing on the $\text{ToOBDD}_<$ language (and other related languages), we address a number of issues that remained open in [Subbarayan *et al.*, 2007]. We show that beyond **CO** and **VA**, the $\text{ToOBDD}_<$ fragment satisfies **IM** and **ME** but satisfies neither **CD** nor any query among **CE**, **SE** unless P = NP. Among other results, we prove that $\text{ToOBDD}_<$ is not comparable w.r.t. succinctness with any of CNF, DNF, DNNF unless the polynomial hierarchy collapses. This contributes to the explanation of some empirical results reported in [Subbarayan *et al.*, 2007].

1 Introduction

This paper is concerned with "knowledge compilation" (KC), a family of approaches proposed so far for addressing the intractability of a number of AI problems of various kinds (reasoning, decision making, etc.). The key idea underlying KC is to pre-process parts of the available information (i.e., turning them into a compiled form) for improving on-line computational efficiency (see among others [Darwiche, 2001; Cadoli and Donini, 1998; Selman and Kautz, 1996; del Val, 1994]).

A important research line in KC [Gogic *et al.*, 1995; Darwiche and Marquis, 2002] addresses the following issue: *How to choose a target language for knowledge compilation?* In [Darwiche and Marquis, 2002], the authors argue that the choice of a target language must be based both on the set of queries and transformations which can be achieved in polynomial time when the data are represented in the language, as well as the spatial efficiency of the language. They pointed out a KC map which can be viewed as a multi-criteria evaluation of a number of propositional fragments, including DNNF, d−DNNF, CNF, DNF, $\text{OBDD}_<$, OBDD (the union of all $\text{OBDD}_<$ when $<$ varies), etc. (see [Darwiche and Marquis, 2002] for details). From there, other propositional fragments have been considered so far and put in the KC map, see for instance [Wachter and Haenni, 2006; Fargier and Marquis, 2006; Subbarayan *et al.*, 2007; Pipatsrisawat and Darwiche, 2008; Fargier and Marquis, 2008a; 2008b].

Recent experimental results have shown the practical interest of trees-of-BDDs [Subbarayan *et al.*, 2007] as a target language for propositional knowledge compilation: it turns out that the tree-of-BDDs language renders feasible the compilation of a number of benchmarks which cannot be compiled into d−DNNF due to space limitations.

In the present paper, we elaborate on the tree-of-BDDs language. After some formal preliminaries (Section 2), we generalize the tree-of-BDDs language to the family of ToC representations where C is any complete propositional language (Section 3). We provide a number of generic results enabling one to determine the queries/transformations satisfied by ToC depending on the queries/transformations satisfied by C. We also present some results about the spatial efficiency of the ToC languages. Focusing on $\text{ToOBDD}_<$and some related languages, we then address a number of issues that remained open in [Subbarayan *et al.*, 2007] (Section 4): beyond **CO** and **VA**, the $\text{ToOBDD}_<$ language satisfies **IM** and **ME** but does not satisfy any query among **CE**, **SE** unless P = NP. Under similar assumptions from complexity theory, we demonstrate that $\text{ToOBDD}_<$ does not satisfy any transformation among **CD**, **FO**, $\wedge\mathbf{BC}$, $\vee\mathbf{C}$ or $\neg\mathbf{C}$. Among other succinctness results, we prove that the $\text{ToOBDD}_<$ language is not comparable w.r.t. succinctness with any of CNF, DNF or DNNF unless the polynomial hierarchy PH collapses. This contributes to the explanation of some empirical results reported in [Subbarayan *et al.*, 2007]. We conclude the paper by a discussion of the results and some perspectives (Section 5). Proofs are omitted for space reasons but are available at http://www.fr/~marquis/fargier-marquis-ijcai09.pdf.

*The authors have been partly supported by the ANR project PHAC (ANR-05-BLAN-0384).

2 Representations and the KC Map

Trees-of-BDDs and their forthcoming generalization are not *stricto sensu* formulae. Hence we need to extend the notions of queries, transformations and succinctness at work in the KC map to such representations. Roughly speaking, a propositional representation language is a way to represent Boolean functions. Such a representation language often takes the form of a standard propositional language but other data structures can be used as well (e.g. Karnaugh maps, truth tables, various graphs including those binary decision diagrams, ... and of course trees-of-BDDs) for the representation purpose.

Formally, given a finite set of propositional variables PS, we consider Boolean functions from $\{0,1\}^X$ to $\{0,1\}$, where $X \subseteq PS$. $Var(f) = X$ is called the scope of f. The support $\Omega(f)$ of f is the set of all assignments ω of $Var(f)$ to Boolean values such that $f(\omega) = 1$. For any $X \subseteq PS$, we note by $\overline{X}$ the set $PS \setminus X$. The set of Boolean functions is equipped with the three standard internal laws, $\wedge$, $\vee$ and $\neg$. Given $X \subseteq PS$ we note by $\exists X.f$ the Boolean function with scope $Var(f) \setminus X$ that maps 1 to an assignment $\omega_{Var(f)\setminus X}$ of $Var(f) \setminus X$ iff there exists an assignment ω of $Var(f)$ such that the restriction of ω over $Var(f) \setminus X$ and $\omega_{Var(f)\setminus X}$ coincide and $f(\omega) = 1$.

Definition 1 (representation language) *(inspired from [Gogic* et al.*, 1995]) A* (propositional) representation language *over a finite set of propositional variables PS is a set $\subset$ of data structures α (also referred to as $\subset$ representations) together with a scope function $Var : \subset \rightarrow 2^X$ with $X \subseteq PS$ and an interpretation function I which associates to each $\subset$ representation α a Boolean function $I(\alpha)$ the scope of which is $Var(\alpha)$. $\subset$ is also equipped with a size function from $\subset$ to IN that provides the size $|\alpha|$ of any $\subset$ representation α.*[1]

Definition 2 (complete language) *A propositional representation language $\subset$ is said to be* complete *iff for any Boolean function f with $Var(f) \subseteq PS$, there exists a $\subset$ representation α such that $I(\alpha) = f$.*

Clearly enough, formulae from a standard propositional language are representations of Boolean functions. The size of such a formula is the number of symbols in it. Slightly abusing words, when Σ is a propositional formula representing a Boolean function g one often says that a representation α of g is a representation of Σ instead of α is a representation of the semantics of Σ.

The DAG−NNF language [Darwiche and Marquis, 2002] is also a complete graph-based representation language of Boolean functions. Distinguished formulae from DAG−NNF are the literals over PS, the clauses (a clause is a finite disjunction of literals or the Boolean constant $\bot$) and the terms (a term is a finite conjunction of literals or the Boolean constant $\top$). We assume the reader to be familiar with the DAG−NNF fragments DNNF, d−DNNF, CNF, DNF, FBDD, $\text{OBDD}_<$, OBDD, MODS, etc.

Obviously, all the logical notions pertaining to formulae viewed up to logical equivalence can be easily extended to any representation language $\subset$ of Boolean functions. For instance, an asssignment ω of $Var(\alpha)$ to Boolean values is said to be a *model* of a $\subset$ representation α over $Var(\alpha)$ iff $I(\alpha)(\omega) = 1$. Similarly, two representations α and β (possibly from different representation formalisms) are said to be *equivalent*, noted $\alpha \equiv \beta$, when they represent the same Boolean function. A $\subset$ representation α is *consistent* (resp. *valid*) iff α does not represent the Boolean function 0 (resp. represents the Boolean function 1). α is a *logical consequence* of β, noted $\beta \models \alpha$, iff $\Omega(I(\beta)) \subseteq \Omega(I(\alpha))$.

We are now ready to extend the notions of queries, transformations and succinctness considered in the KC map to any propositional representation language. Their importance is discussed in depth in [Darwiche and Marquis, 2002], so we refrain from recalling it here.

Definition 3 (queries) *Let $\subset$ denote a propositional representation language.*

- *$\subset$ satisfies* **CO** *(resp.* **VA***) iff there exists a polytime algorithm that maps every $\subset$ representation α to 1 if α is consistent (resp. valid), and to 0 otherwise.*
- *$\subset$ satisfies* **CE** *iff there exists a polytime algorithm that maps every $\subset$ representation α and every clause δ to 1 if $\alpha \models \delta$ holds, and to 0 otherwise.*
- *$\subset$ satisfies* **EQ** *(resp.* **SE***) iff there exists a polytime algorithm that maps every pair of $\subset$ representations α, β to 1 if $\alpha \equiv \beta$ (resp. $\alpha \models \beta$) holds, and to 0 otherwise.*
- *$\subset$ satisfies* **IM** *iff there exists a polytime algorithm that maps every $\subset$ representation α and every term γ to 1 if $\gamma \models \alpha$ holds, and to 0 otherwise.*
- *$\subset$ satisfies* **CT** *iff there exists a polytime algorithm that maps every $\subset$ representation α to a nonnegative integer that represents the number of models of α over $Var(\alpha)$ (in binary notation).*
- *$\subset$ satisfies* **ME** *iff there exists a polynomial $p(.,.)$ and an algorithm that outputs all models of an arbitrary $\subset$ representation α in time $p(|\alpha|, m)$, where m is the number of its models (over $Var(\alpha)$).*

Definition 4 (transformations) *Let $\subset$ denote a propositional representation language.*

- *$\subset$ satisfies* **CD** *iff there exists a polytime algorithm that maps every $\subset$ representation α and every consistent term γ to a $\subset$ representation β of the restriction of $I(\alpha)$ to $I(\gamma)$, i.e., $Var(\beta) = Var(\alpha) \setminus Var(\gamma)$ and $I(\beta) = \exists Var(\gamma).(I(\alpha) \wedge I(\gamma))$.*
- *$\subset$ satisfies* **FO** *iff there exists a polytime algorithm that maps every $\subset$ representation α and every subset X of variables from PS to a $\subset$ representation of $\exists X.I(\alpha)$. If the property holds for each singleton X, we say that $\subset$ satisfies* **SFO**.
- *$\subset$ satisfies* $\wedge$**C** *(resp.* $\vee$**C***) iff there exists a polytime algorithm that maps every finite set of $\subset$ representations*

[1] I refers to the interpretation function associated to the $\subset$ language, so that $I_{\subset}$ would be a more correct notation for it; nevertheless, in order to keep the notations light and since no ambiguity is possible, we refrained from indexing the functions I (as well as Var and the size function) by the associated representation language.

$\alpha_1, \ldots, \alpha_n$ *to a* C *representation of* $I(\alpha_1) \wedge \cdots \wedge I(\alpha_n)$ *(resp.* $I(\alpha_1) \vee \cdots \vee I(\alpha_n)$*).*

- C *satisfies* $\wedge\mathbf{BC}$ *(resp.* $\vee\mathbf{BC}$*) iff there exists a polytime algorithm that maps every pair of* C *representations* α *and* β *to a* C *representation of* $I(\alpha) \wedge I(\beta)$ *(resp.* $I(\alpha) \vee I(\beta)$ *).*
- C *satisfies* $\neg\mathbf{C}$ *iff there exists a polytime algorithm that maps every* C *representation* α *to a* C *representation of* $\neg I(\alpha)$*.*

Definition 5 (succinctness) *Let* $\mathtt{C}_1$ *and* $\mathtt{C}_2$ *be two representation languages.* $\mathtt{C}_1$ *is at least as succinct as* $\mathtt{C}_2$*, noted* $\mathtt{C}_1 \leq_s \mathtt{C}_2$*, iff there exists a polynomial* p *such that for every* $\mathtt{C}_2$ *representation* α *there exists an equivalent* $\mathtt{C}_1$ *representation* β *where* $|\beta| \leq p(|\alpha|)$*.*

$\sim_s$ is the symmetric part of $\leq_s$ defined by $\mathtt{C}_1 \sim_s \mathtt{C}_2$ iff $\mathtt{C}_1 \leq_s \mathtt{C}_2$ and $\mathtt{C}_2 \leq_s \mathtt{C}_1$. $<_s$ is the asymmetric part of $\leq_s$ defined by $\mathtt{C}_1 <_s \mathtt{C}_2$ iff $\mathtt{C}_1 \leq_s \mathtt{C}_2$ and $\mathtt{C}_2 \not\leq_s \mathtt{C}_1$. Finally, $\mathtt{C}_1 \leq_s^* \mathtt{C}_2$ (resp. $\mathtt{C}_1 <_s^* \mathtt{C}_2$) means that $\mathtt{C}_1 \leq_s \mathtt{C}_2$ (resp. $\mathtt{C}_1 <_s \mathtt{C}_2$) unless the polynomial hierarchy PH collapses (which is considered very unlikely in complexity theory).

We also consider the following restriction of the succinctness relation:

Definition 6 (polynomial translation) *Let* $\mathtt{C}_1$ *and* $\mathtt{C}_2$ *be two representation languages.* $\mathtt{C}_1$ *is* polynomially translatable *into* $\mathtt{C}_2$*, noted* $\mathtt{C}_1 \geq_{\mathcal{P}} \mathtt{C}_2$*, iff there exists a polytime algorithm* A *such that for every* $\mathtt{C}_1$ *representation* α $A(\alpha)$ *is a* $\mathtt{C}_2$ *representation such that* $A(\alpha) \equiv \alpha$*.*

Like $\geq_s$, $\geq_{\mathcal{P}}$ is a preorder (i.e., a reflexive and transitive relation) over propositional representation languages. It refines the spatial efficiency preorder $\geq_s$ in the sense that for any $\mathtt{C}_1$ and $\mathtt{C}_2$, if $\mathtt{C}_1 \geq_{\mathcal{P}} \mathtt{C}_2$, then $\mathtt{C}_1 \geq_s \mathtt{C}_2$ (but the converse does not hold in general). We note by $\sim_{\mathcal{P}}$ the symmetric part of $\leq_{\mathcal{P}}$.

3 The ToC Languages

We start with the definition of trees-of-BDDs as given in [Subbarayan *et al.*, 2007] (modulo the notations used):

Definition 7 (tree-of-BDDs)

- *A* decomposition tree *of a* CNF *formula* Σ *is a (finite) labelled tree* T *whose set of nodes is* N*. Each node* $n \in N$ *is labelled with* $Var(n)$*, a subset of* $Var(\Sigma)$*. For each* $n \in N$*, let* $clauses(n) = \{$*clause* δ *of* Σ *s.t* $Var(\delta) \subseteq Var(n)\}$*;* T *satisfies two conditions: for every clause* δ *of* Σ *there exists* $n \in N$ *such that* $\delta \in clauses(n)$*, and for every* $x \in Var(\Sigma)$*,* $\{n \in N \mid x \in Var(n)\}$ *forms a connected subtree of* T*.*
- *Let* $<$ *be a total strict ordering over* PS*. A* tree-of-BDDs *of a* CNF *formula* Σ *given* $<$ *consists of a decomposition tree* T *of* Σ *equipped with a further labelling function* B *such that for every* $n \in N$*,* $B(n)$ *is the* $\mathtt{OBDD}_<$ *representation of* $\exists \overline{Var(n)}.I(\Sigma)$*.*
  We have $Var(T) = \bigcup_{n \in N} Var(n)$ *and* $I(T) = \bigwedge_{n \in N} I(B(n))$*.* ToB *denotes the set of all trees-of-BDDs given* $<$*.*

Clearly, ToB is a complete representation language: for every Boolean function there is a CNF formula Σ representing it, and thus a tree-of-BDDs T of Σ such that $I(T) = I(\Sigma)$.

The above definition can be simplified and extended, allowing the representation of other formulae than CNF ones, and taking advantage of other target languages than $\mathtt{OBDD}_<$ for compiling the labels $B(n)$:

Definition 8 (ToC**)** *Let* C *be any complete propositional representation language. A* ToC *representation is a finite, labelled tree* T*, whose set of nodes is* N*. Each node* $n \in N$ *is labelled with* $Var(n)$*, a subset of* PS *and with a* C *representation* $B(n)$*.*

T *must satisfy:*

- the running intersection property: *for each* $x \in \bigcup_{n \in N} Var(n)$*,* $\{n \in N \mid x \in Var(n)\}$ *forms a connected subtree of* T*, and*
- the global consistency property: *for each* $n \in N$*,* $I(B(n)) \equiv \exists \overline{Var(n)}. \bigwedge_{n \in N} I(B(n))$*.*

We have $Var(T) = \bigcup_{n \in N} Var(n)$ *and* $I(T) = \bigwedge_{n \in N} I(B(n))$*. The size of a* ToC *representation* T *is the size of this tree, plus the sizes of the labels of the nodes of* T *(numbers of variables in* $Var(n)$ *and sizes of* $B(n)$*).*
ToC *denotes the set of all* ToC *representations.*

Taking C = $\mathtt{OBDD}_<$, we get the $\mathtt{ToOBDD}_<$ language. Clearly, this definition of $\mathtt{ToOBDD}_<$ is close to the previous one ToB from [Subbarayan *et al.*, 2007], except that a $\mathtt{ToOBDD}_<$ representation T is defined *per se*, i.e., independently from a given CNF formula Σ. Within this language, unlike with the $\mathtt{OBDD}_<$ one, a Boolean function may have several equivalent representations. For instance, let $\Sigma = (\neg a \wedge \neg b) \vee (\neg a \wedge c) \vee (b \wedge c)$. Whatever $<$, $I(\Sigma)$ can be represented by the $\mathtt{ToOBDD}_<$ representation T such that T has a single node n_0, such that $Var(n_0) = Var(\Sigma)$ and $B(n_0)$ is the $\mathtt{OBDD}_<$ representation equivalent to Σ; observing that $\Sigma \equiv (\neg a \vee b) \wedge (\neg b \vee c)$, $I(\Sigma)$ can also be represented by the $\mathtt{ToOBDD}_<$ representation T such that T has two nodes n_0 and n_1, the root of T is n_0, $Var(n_0) = \{a, b\}$, $Var(n_1) = \{b, c\}$, $B(n_0)$ is the $\mathtt{OBDD}_<$ formula equivalent to $(\neg a \vee b)$, and $B(n_1)$ is the $\mathtt{OBDD}_<$ formula equivalent to $(\neg b \vee c)$. In short, $\mathtt{ToOBDD}_<$ does not offer the property of canonical representation.

Compiling a CNF formula Σ into a ToC representation T basically consists in computing first a decomposition tree of Σ, then taking advantage of any CNF-to-C compiler so as to turn the CNF $clauses(n)$ formulae (for each node n of the tree) into equivalent C representations, and finally to use the well-known message-passing propagation algorithm (see the **Propagate** function in [Subbarayan *et al.*, 2007], which applies also to ToC representations) from the leaves of the tree to its root then from the root to the leaves so as to ensure the global consistency property.This approach can be easily extended to deal with the compilation of any conjunctive representation into a ToC representation when compilers to C are available. The running intersection property enables one to replace a global computation on the resulting ToC represen-

tation T by a number of possibly easier, local computations on the corresponding $B(n)$.

Let us now present some generic properties about ToC fragments; such properties are about queries, transformations and succinctness, and are related to similar properties satisfied by the corresponding C languages. We first need the following definition:

Definition 9 (TE, CL) *Let* C *be any propositional representation language.*

- C *satisfies* **TE** *(the* term condition*) iff for every term γ over PS, a* C *representation equivalent to γ can be computed in time polynomial in $|\gamma|$.*
- C *satisfies* **CL** *(the* clause condition*) iff for every clause δ over PS, a* C *representation equivalent to δ can be computed in time polynomial in $|\delta|$.*

Clearly enough, those conditions are not very demanding and are satisfied by all complete languages considered in [Darwiche and Marquis, 2002], but MODS.

Proposition 1 *Let* C *be any complete propositional representation language.*

1. C *satisfies* **CO** *iff* ToC *satisfies* **CO**.
2. C *satisfies* **VA** *iff* ToC *satisfies* **VA**.
3. C *satisfies* **IM** *iff* ToC *satisfies* **IM**.
4. *If* C *satisfies* **CD**, *then* C *satisfies* **ME** *iff* ToC *satisfies* **ME**.
5. *If* C *satisfies* **CL**, *then* ToC *does not satisfy* **CE** *unless* P = NP.
6. *If* C *satisfies* **CL**, *then* ToC *does not satisfy* **SE** *unless* P = NP.

Points 1. to 4. show that the ToC languages typically satisfy all the queries **CO**, **VA**, **IM**and **ME** (just because the corresponding C languages typically satisfy them and **CD**). Similarly, points 5. and 6. show that the ToC languages typically do not satisfy any of **CE** or **SE** unless P = NP (because t because the corresponding C languages typically satisfy **CL**). Finally, since every propositional language satisfying **CO** and **CD** also satisfies **CE** (a straightforward extension of Lemma 1.4 from [Darwiche and Marquis, 2002] to any propositional representation language), we get as a corollary to points 1. and 5. that:

Corollary 1 *If* C *satisfies* **CO** *and* **CL**, *then* ToC *does not satisfy* **CD** *unless* P = NP.

Considering other transformations, we obtained the following results which hold for any propositional representation language (hence specifically for the ToC ones):

Proposition 2 *Let* C *be any propositional representation language.*

1. *If* C *satisfies* **CO** *and* **TE** *and* C *does not satisfy* **CE** *unless* P = NP, *then* C *does not satisfy* $\wedge$**BC** *unless* P = NP.
2. *If* C *satisfies* **VA** *and* **TE**, *then* C *does not satisfy* $\vee$**C** *unless* P = NP.
3. *If* C *satisfies* **IM** *and does not satisfy* **CE** *unless* P = NP, *then* C *does not satisfy* $\neg$**C** *unless* P = NP.

These results show that the ToC languages typically satisfy only few transformations among **CD**, $\wedge$**BC**, $\vee$**C** and $\neg$**C**. The conditions on C listed in Corollary 1 and Proposition 2 are indeed not very demanding.

It is interesting to note that the algorithms **Conditioning**, **Project**, **IsCE**, **IsEQ** reported in [Subbarayan *et al.*, 2007] (Figure 3), for respectively computing the conditioning of a ToOBDD$_<$ representation by a consistent term, computing the projection of a ToOBDD$_<$ representation T on a given set V of variables (or equivalently, forgetting all variables in T except those of V), deciding whether a clause is entailed by a ToOBDD$_<$ representation, deciding whether two ToOBDD$_<$ representations are equivalent, apply to ToC representations as well (the fact that each $B(n)$ of T is an OBDD$_<$ representation is not mandatory for ensuring the correctness of these algorithms). While these algorithms do not run in polynomial time in the general case, imposing further restrictions on C can be a way to achieve tractability. Thus, it is easy to show that if C has a linear-time algorithm for **FO** and a linear-time algorithm for $\wedge$**BC**, then **Project** is a polytime **FO** algorithm for the ToC languages. If C has a linear-time algorithm for **FO**, a linear-time algorithm for $\wedge$**BC**, and a polytime algorithm for **CD**, then **Conditioning** is a polytime **CD** algorithm for the ToC languages.

The fact that many queries/transformations are NP-hard in the general case does not discard ToOBDD$_<$ (and beyond it the ToC languages) as interesting target languages for KC *from the practical side*.[2] Indeed, if the width of a ToC representation T, i.e., $max_{n \in N}(\{|Var(n)| - 1\})$, is (upper) bounded by a constant,[3] then the time complexity of the **Propagate** function becomes linear in the tree size; as a consequence, many other queries and transformations may become tractable as well; for instance if C satisfies **CD**, we get that both conditioning and clausal entailment can be achieved in polynomial time in the tree size.

As to succinctness, we got the following results:

Proposition 3 *Let* C *be any complete propositional representation language.*

1. ToC $\leq_{\mathcal{P}}$ C.
2. *Let* C′ *be any complete propositional fragment. If* C $\leq_s$ C′, *then* ToC $\leq_s$ ToC′.
3. *If* C *satisfies* **CL** *and* C' *satisfies* **CE** *then* C' $\not\leq_s^*$ ToC.
4. *If* C *satisfies* **IM**, *then* ToC $\not\leq_s^*$ DNF.

Proposition 3 has many interesting consequences:

- From point 1., we directly get that ToC $\leq_s$ C, and that ToC is complete (since C is). This result *cannot* be strengthened to ToC $<_s$ C in the general case (for every C satisfying $\wedge$**C**, e.g. C = CNF, we can prove that C $\sim_{\mathcal{P}}$ ToC).

[2] See [Marquis, 2008] for more details on this issue.

[3] The price to be paid by such a restriction is a lack of expressiveness: none of the languages of ToC representations of width bounded by c (where c is a parameter) is complete.

- Point 2. allows one to take advantage of previous results describing how propositional languages C are organized w.r.t. spatial efficiency in order to achieve similar results for the corresponding ToC languages.
- Point 3. implies that the DNNF language, which satisfies **CE**, is typically (i.e., whenever C satisfies **CL**) not more succinct than the corresponding ToC language; hence none of the languages C which are less succinct than DNNF (e.g. C = DNF) can be more succinct than such ToC languages; thus, we get for instance that DNF $\not\leq_s^*$ ToDNF (which together with point 1. shows that ToDNF $<_s^*$ DNF).
- Another consequence of point 3. is that if C satisfies **CL** then DNNF $\not\leq_s^*$ ToC (hence d−DNNF $\not\leq_s^*$ ToC). With point 1. this shows ToDNNF to be spatially (strictly) more efficient than DNNF, while keeping **CO** and **ME**.

Finally, an interesting issue is to determine whether, at the "instance level", i.e., considering a given Boolean function to be compiled, targeting ToC in a compilation process leads always to save space w.r.t. targeting C. The answer is "not always" (even in the cases when we have ToC $<_s^*$ C). We showed it by considering the notion of decomposition set:

Definition 10 (decomposition) *Let f be a Boolean function. Let V_1, ..., V_k be k subsets of PS. $D = \{V_1, \ldots, V_k\}$ is a* decomposition set *for f iff we have $f = \bigwedge_{i=1}^{k} \exists \overline{V_i}.f$.*

Clearly enough, for each ToC representation T whose set of nodes is N, $\{Var(n) \mid n \in N\}$ is a decomposition set for $I(T)$. We proved that:

Lemma 1 *Let f be a Boolean function. Let δ be an essential prime implicate of f, i.e., a prime implicate of f which is not implied by the conjunction of the other prime implicates of f. Then for every decomposition set D for f, there exists $V \in D$ such that $Var(\delta) \subseteq V$.*

This lemma shows that when f has an essential prime implicate containing all its variables, no ToC representation of f can be more compact than each of its C representations. This lemma also shows that when f has an essential prime implicate δ such that $\exists \overline{Var(\delta)}.f$ has no C representation of reasonable size, chosing ToC as the target language is not a way to save space.

Finally, Lemma 1 also explains why imposing a fixed decomposition tree T for defining a ToC language is not so a good idea (despite the fact it may offer a property of canonicity in some cases): either T has a node n such that $Var(n) = \{x_1, \ldots, x_p\}$ (all the variables of interest), and in this case the corresponding ToC language mainly amounts to C, or T does not contain such a node, and in this case the ToC language is incomplete: the Boolean function which is the semantics of the clause $\bigvee_{i=1}^{p} x_i$ cannot be represented in ToC.

4 Back to ToOBDD$_<$ Representations

Let us now fix C to OBDD$_<$ in order to get some further results. Beyond ToOBDD$_<$ we have investigated the properties of U(ToOBDD$_<$) (the union of all ToOBDD$_<$ for each total order $<$ over PS) and of ToOBDD, as target languages for propositional knowledge compilation, along the lines of the KC map. To make the differences between these languages clearer, observe that OBDD representations $B(n), n \in N$ where N is the set of nodes of a given ToOBDD T may rely on different variable orders $<$, while all the OBDD$_<$ representations in a given U(ToOBDD$_<$) are based on the same order. Hence, U(ToOBDD$_<$) is a proper subset of ToOBDD.

	CE	VA	CO	IM	EQ	SE	CT	ME
ToOBDD	○	√	√	√	?	○	?	√
U(ToOBDD$_<$)	○	√	√	√	?	○	?	√
ToOBDD$_<$	○	√	√	√	?	○	?	√
OBDD	√	√	√	√	√	√	√	√
OBDD$_<$	√	√	√	√	√	√	√	√

	CD	FO	SFO	∧BC	∧C	∨BC	∨C	¬C
ToOBDD	○	○*	?	○	•	○	○	○
U(ToOBDD$_<$)	○	○*	?	○	•	○	○	○
ToOBDD$_<$	○	○*	?	○	•	?	○	○
OBDD	√	•	√	○	•	○	•	√
OBDD$_<$	√	•	√	√	•	√	•	√

Table 1: √ means "satisfies", • means "does not satisfy", ○ means "does not satisfy unless P = NP", and ○* means "does not satisfy unless PH collapses." Results for OBDD$_<$ and OBDD are from [Darwiche and Marquis, 2002] and are given here as a baseline.

Proposition 4 *The results in Table 1 hold.*

The fact that ToOBDD, U(ToOBDD$_<$), and ToOBDD$_<$ satisfy **CO**, **VA**, **IM**, and **ME** and that none of these languages satisfies any of **CE**, **SE**, **CD**, ∧**BC**, ∧**C**, ∨**C** or ¬**C**, unless P = NP is a direct corollary of Propositions 1 and 2. Except **CO** and **VA**, all those results concern some issues left open in [Subbarayan *et al.*, 2007]. Especially, there exist polytime algorithms for **IM** and **ME** which are not based on the message-passing propagation algorithm (those given in [Subbarayan *et al.*, 2007] do not run in polynomial time in the general case). Furthermore, contrary to what was expected in [Subbarayan *et al.*, 2007], ¬**C** is *not* trivial: the negation of a conjunction of OBDD$_<$ representations is equivalent to the *disjunction* of their negations. We actually showed that the ¬**C** transformation on ToOBDD$_<$ cannot be achieved in polynomial time unless P = NP.

As to succinctness, we proved the following results:

Proposition 5

1. *For each $<$,* ToOBDD$_<$ $<_s^*$ OBDD$_<$.
2. *For each $<$,* DNNF $\not\leq_s^*$ ToOBDD$_<$.
3. ToOBDD $\not\leq_s^*$ DNF.
4. ToOBDD $\not\leq_s$ CNF.

Points 1. to 3. are direct consequences of Proposition 3 and results from [Darwiche and Marquis, 2002]). A direct consequence of Proposition 5 is that d−DNNF $\not\leq_s^*$ ToOBDD$_<$. This explains in some sense the space savings which can be offered by ToOBDD$_<$ over d−DNNF and observed empirically as reported in [Subbarayan *et al.*, 2007]. More generally, from Proposition 3 and some results given in [Darwiche and Marquis, 2002] we get that:

Corollary 2 *Unless* PH *collapses,* ToOBDD, U(ToOBDD$_<$) *and* ToOBDD$_<$ *are* incomparable w.r.t. succinctness *with the languages* CNF, DNF, *and* DNNF.

5 Conclusion

In this paper, the concept of tree-of-BDDs has been extended to any complete propositional representation language C thus leading to the family of ToC languages. A number of generic results are provided, which allow to determine the queries/transformations satisfied by ToC depending on the ones satisfied by C, as well as results about the spatial efficiency of the ToC languages. Focusing on the $\texttt{ToOBDD}_<$ language, we have addressed a number of issues that remained open in [Subbarayan *et al.*, 2007]; especially, we have shown that beyond **CO** and **VA**, $\texttt{ToOBDD}_<$ satisfies **IM** and **ME** but does not satisfy any query among **CE**, **SE**. We have also proved that $\texttt{ToOBDD}_<$ does not satisfy any transformation among **CD**, **FO**, $\wedge\mathbf{BC}$, $\vee\mathbf{C}$ or $\neg\mathbf{C}$ and that this fragment is not comparable for succinctness w.r.t. any of CNF, DNF and DNNF unless PH collapses.

From this investigation, it turns out that the $\texttt{ToOBDD}_<$ language (and in general the ToC languages) satisfies only few queries and transformations. Subsequently, in applications where some queries/transformations not satisfied by $\texttt{ToOBDD}_<$ must be achieved under some guaranteed response time, considering $\texttt{ToOBDD}_<$ as a target language for KC is not always the best choice. From the practical side, as reported in [Subbarayan *et al.*, 2007] (and despite the fact that $\texttt{ToOBDD}_< \not\leq_s \texttt{CNF}$), there are CNF formulae which can be compiled into $\texttt{ToOBDD}_<$ using a reasonable amount of computational resources, while it turned out impossible to generate d−DNNF representations for them. Such empirical results cohere with our succinctness result $\texttt{d−DNNF} \not\leq_s^* \texttt{ToOBDD}_<$. Nevertheless, our result $\texttt{ToOBDD} \not\leq_s^* \texttt{DNNF}$ shows that this empirical evidence can be argued (this result implies that some DNNF representations do not have "small" $\texttt{ToOBDD}_<$ equivalent representations under the standard assumptions of complexity theory), so DNNF remains a very attractive language for the KC purpose.

Our results also suggest a number of ToC languages as quite promising. Consider for instance the ToFBDD language. From our results, it comes easily that ToFBDD satisfies **CO**, **VA**, **IM**, **ME** (hence the same queries as $\texttt{ToOBDD}_<$); since ToFBDD is at least as succinct as ToOBDD, it appears as a challenging fragment. Furthermore, a compiler to FBDD is already available (see e.g. www.eecg.utoronto.ca/∼jzhu/fbdduser11.ps). When none of **VA** or **IM** is expected, the ToDNNF language looks also valuable; indeed, from our results we know that ToDNNF satisfies **CO** and **ME**, while being quite compact: $\texttt{ToDNNF} <_s^* \texttt{ToOBDD}_<$ and $\texttt{ToDNNF} <_s^* \texttt{DNNF}$ hold; beyond the spatial dimension, targeting the ToDNNF language may also reduce the on-line computation time needed for achieving queries/transformations based on **Propagate** function (as well as the off-line CNF-to-ToC compilation time) since DNNF satisfies **FO**, which is one of the two key operations of the propagation algorithm. The $\texttt{ToDNNF}_T$ language, based on $\texttt{DNNF}_T$ [Pipatsrisawat and Darwiche, 2008], also looks interesting in this respect since it satisfies both **FO** and $\wedge\mathbf{BC}$, the other key operation of the propagation algorithm.

This is what the "theory" says in some sense about such languages. Going further requires to implement compilers and perform experiments in order to determine whether, from the practical side, representations from those languages can be computed using a reasonable amount of resources. This is an issue for further research. Another perspective for further work is to complete the missing results about queries, transformations and succinctness for the ToC languages and to extend the KC map accordingly. Especially, it would be interesting to characterize some families of propositional formulae each of DNNF and ToOBDD are "effective" on.

References

[Cadoli and Donini, 1998] M. Cadoli and F.M. Donini. A survey on knowledge compilation. *AI Communications*, 10(3–4):137–150, 1998.

[Darwiche and Marquis, 2002] A. Darwiche and P. Marquis. A knowledge compilation map. *Journal of Artificial Intelligence Research*, 17:229–264, 2002.

[Darwiche, 2001] A. Darwiche. Decomposable negation normal form. *Journal of the ACM*, 48(4):608–647, 2001.

[del Val, 1994] A. del Val. Tractable databases: How to make propositional unit resolution complete through compilation. In *Proc. of KR'94*, pages 551–561, 1994.

[Fargier and Marquis, 2006] H. Fargier and P. Marquis. On the use of partially ordered decision graphs in knowledge compilation and quantified Boolean formulae. In *Proc. of AAAI'06*, pages 42–47, 2006.

[Fargier and Marquis, 2008a] H. Fargier and P. Marquis. Extending the knowledge compilation map: Closure principles. In *Proc. of ECAI'08*, pages 50–54, 2008.

[Fargier and Marquis, 2008b] H. Fargier and P. Marquis. Extending the knowledge compilation map: Krom, Horn, affine and beyond. In *Proc. of AAAI'08*, pages 442–447, 2008.

[Gogic *et al.*, 1995] G. Gogic, H.A. Kautz, Ch.H. Papadimitriou, and B. Selman. The comparative linguistics of knowledge representation. In *Proc. of IJCAI'95*, pages 862–869, 1995.

[Marquis, 2008] P. Marquis. Knowledge compilation: a sightseeing tour. In *Tutorial notes, ECAI'08*, 2008. available on-line.

[Pipatsrisawat and Darwiche, 2008] K. Pipatsrisawat and A. Darwiche. New compilation languages based on structured decomposability. In *Proc. of AAAI'08*, pages 517–522, 2008.

[Selman and Kautz, 1996] B. Selman and H.A. Kautz. Knowledge compilation and theory approximation. *Journal of the ACM*, 43:193–224, 1996.

[Subbarayan *et al.*, 2007] S. Subbarayan, L. Bordeaux, and Y. Hamadi. Knowledge compilation properties of tree-of-BDDs. In *Proc. of AAAI'07*, pages 502–507, 2007.

[Wachter and Haenni, 2006] M. Wachter and R. Haenni. Propositional DAGs: A new graph-based language for representing Boolean functions. In *Proc. of KR'06*, pages 277–285, 2006.