CDSAT for Nondisjoint Theories with Shared Predicates: Arrays With Abstract Length[⋆]

Maria Paola Bonacina[1,*], Stéphane Graham-Lengrand[2] and Natarajan Shankar[3]

[1] Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy
[2] SRI International, Menlo Park, California, USA
[3] SRI International, Menlo Park, California, USA

Abstract

CDSAT (*Conflict-Driven Satisfiability*) is a paradigm for *theory combination* that works by coordinating *theory modules* to reason in the *union of the theories* in a *conflict-driven* manner. We generalize CDSAT to the case of *nondisjoint theories* by presenting a new CDSAT theory module for a *theory of arrays with abstract length*, which is an abstraction of the *theory of arrays with length*. The length function is a *bridging function* as it forces theories to share symbols, but the proposed abstraction limits the sharing to one predicate symbol. The CDSAT framework handles shared predicates with minimal changes, and the new module satisfies the CDSAT requirements, so that completeness is preserved.

Keywords

Combination of theories, Nondisjoint theories, CDSAT, Theory of arrays

1. Introduction

CDSAT (*Conflict-Driven Satisfiability*) is a method to decide the satisfiability of a formula modulo a union of theories and an initial assignment of values to terms [1, 2]. CDSAT orchestrates *theory modules*, one for every theory in the union, to perform a *conflict-driven search* of a model of the input formula. A theory module is an abstraction of a theory satisfiability procedure. For proving properties such as soundness, completeness, and termination, a theory module is simply an inference system for the theory.

In this paper we generalize CDSAT to handle unions of theories that are *not necessarily disjoint*. Disjoint theories share only sorts and equality predicates on shared sorts. Nondisjoint theories share also symbols other than equality. For example, consider a theory of *arrays with length*. Length is usually thought of as a function from arrays to integers. Such a function is called a *bridging function* [3, 4, 5], because it constitutes a

SMT 2022: 20th Workshop on Satisfiability Modulo Theories, August 11–12, 2022, 9th FLoC, Haifa, Israel

*Corresponding author.
mariapaola.bonacina@univr.it (M. P. Bonacina); stephane.graham-lengrand@csl.sri.com (S. Graham-Lengrand); shankar@csl.sri.com (N. Shankar)
http://profs.sci.univr.it/~bonacina/ (M. P. Bonacina); http://www.csl.sri.com/users/sgl/ (S. Graham-Lengrand); http://www.csl.sri.com/users/shankar/ (N. Shankar)
0000-0001-9104-2692 (M. P. Bonacina); 0000-0002-2112-7284 (S. Graham-Lengrand)

bridge between arrays and linear integer arithmetic (LIA) that forces the two theories to share symbols (e.g., the *theory of arrays with MaxDiff* [6] and LIA share 0 and $\leq$).

We present a new abstract approach to nondisjoint theories with bridging functions, and we exemplify it with the theory ArrL of *arrays with abstract length.* In ArrL, the length of an array can be an integer, but does not have to be, and the concept of an index being within bounds is abstracted into that of an index being *admissible.* Admissibility is expressed by the *shared predicate* Adm, which remains uninterpreted for ArrL, while another theory $\mathcal{T}$, which is not necessarily LIA, provides its interpretation. In this manner, the two theories share a *minimum amount of information*, namely Adm and the sorts of its arguments, indices and lengths. In ArrL, an array is interpreted as a *partial updatable function*, whose domain of definition is given by the set of admissible indices. We define an axiomatization and a CDSAT theory module for the theory ArrL.

We show that CDSAT is *complete* for this kind of nondisjoint combination (*soundness* and *termination* are preserved). The completeness of CDSAT employs the concept of a *leading theory*, say $\mathcal{T}_1$, which may be one of the theories in the union or a theory that only needs to exist in principle. $\mathcal{T}_1$ acts as a hub: it has the information shared between any two theories, and it suffices that each theory agrees with $\mathcal{T}_1$ on the shared information to have an agreement among all the theories. If the theories are disjoint, they only need to agree on equalities and the cardinalities of shared sorts. Thus, $\mathcal{T}_1$ has all the sorts in the union and aggregates all the cardinality constraints on the shared sorts [1, 2]. If the theories are *not disjoint*, they also need to agree on shared symbols other than equality. Therefore, $\mathcal{T}_1$ has also all the symbols shared by any two theories. For example, ArrL and $\mathcal{T}$ share the predicate Adm with $\mathcal{T}_1$. The agreement between $\mathcal{T}$ and $\mathcal{T}_1$ and the agreement between ArrL and $\mathcal{T}_1$ imply the agreement between $\mathcal{T}$ and ArrL, on the interpretation of Adm, equalities, and cardinalities of shared sorts.

2. Preliminaries

A *signature* Σ is given by a set S of *sorts*, including the sort prop of Booleans, and a set F of *sorted symbols*, with equality symbols $\simeq_s$ for all sorts $s \in S$. A collection $\mathcal{V} = (\mathcal{V}^s)_{s\in S}$ of disjoint sets of *variables* is available. We use t and u for terms, and l for formulae that are the terms of sort prop. $\Sigma[\mathcal{V}]$-interpretations and Σ-structures are defined as usual.

A *theory* $\mathcal{T}$ is defined by a signature Σ and a set $\mathcal{A}$ of axioms that state properties of symbols in Σ, or as the class of Σ-structures that satisfy $\mathcal{A}$, called models of $\mathcal{T}$ or $\mathcal{T}$*-models*. Symbols that do not appear in the axioms are *free* or *uninterpreted.* Let $\mathcal{T}_1, \ldots, \mathcal{T}_n$ with signatures $\Sigma_k = (S_k, F_k)$, $\forall k$, $1 \leq k \leq n$, be the theories to be combined. Their union is denoted $\mathcal{T}_\infty$, with signature $\Sigma_\infty = (S_\infty, F_\infty)$, for $S_\infty = \bigcup_{k=1}^{n} S_k$ and $F_\infty = \bigcup_{k=1}^{n} F_k$. The symbols $\mathcal{T}$ and Σ stand for any $\mathcal{T}_k$ and Σ_k including $\mathcal{T}_\infty$ and Σ_∞. If the top symbol of a subterm u of a $\mathcal{T}_\infty$-term t is not in F_k, term u is a variable for theory $\mathcal{T}_k$: term u and its top symbol are dubbed Σ_k*-foreign.* Recall that $\trianglelefteq$ is the subterm ordering. The set $\mathsf{fv}_\Sigma(t)$ of the free Σ-variables of term t is the set of all $\triangleleft$-maximal subterms of t whose top symbol is Σ-foreign.

3. A Theory of Arrays with Abstract Length

The simplest theory of arrays is the theory Arr_0 of *arrays without extensionality*, which does not have axioms specifying when two arrays are equal. The theory Arr of *arrays with extensionality* adds to Arr_0 an *extensionality axiom* saying that two arrays are equal if and only if they have the same elements at *all indices*. In this section we define a theory ArrL of *arrays with extensionality and abstract length*, which extends Arr_0 in a different way, specifying different conditions for two arrays to be equal.

The signature of ArrL has sorts for *arrays*, *indices*, *elements*, and *lengths*. In order to allow arrays of different types, including arrays of arrays, one assumes a set of *basic sorts*, which includes prop, and an *array sort constructor*, denoted $\Rightarrow$, so that $I \overset{L}{\Rightarrow} V$ is the sort of arrays with indices of sort I, elements of sort V, and lengths of sort L. The set S_{ArrL} of the sorts of ArrL is the free closure of the set of basic sorts with respect to $\Rightarrow$.

The signature of ArrL includes the function symbols $\mathsf{select}\colon (I \overset{L}{\Rightarrow} V)\times I \to V$ for *select* or *read*, $\mathsf{store}\colon (I \overset{L}{\Rightarrow} V)\times I\times V \to (I \overset{L}{\Rightarrow} V)$ for *store* or *write*, and $\mathsf{len}\colon (I \overset{L}{\Rightarrow} V) \to L$ that maps an array a to its length $\mathsf{len}(a)$. Terms of the form $\mathsf{select}(a, i)$ may be abbreviated as $a[i]$. The signature also features a predicate symbol $\mathsf{Adm}\colon I\times L \to \mathsf{prop}$, such that if i is a term of sort I and a is a term of sort $I \overset{L}{\Rightarrow} V$, then $\mathsf{Adm}(i, \mathsf{len}(a))$ is true if index i is *admissible* with respect to $\mathsf{len}(a)$. Another theory shares with ArrL the symbol Adm and the sorts I and L (sharing the sort V is not necessary) and provides a concrete meaning of admissibility. The notion of admissibility is an abstraction that frees the theory of arrays from the commitment that lengths are positive integers and that the indices of array a are the consecutive nonnegative integers in the interval $[0, n)$ for $n = \mathsf{len}(a)$. Although this is a popular choice (e.g., [6]), it is not the only one.

Example 1. LIA *interprets both the sort L of lengths and the sort I of indices as the set $\mathbb{Z}$ of the integers, and defines $\mathsf{Adm}(i, n) \leftrightarrow 0 \le i < n$. A different theory may interpret I as a set S and L as the powerset of S, denoted $\mathcal{P}(S)$, and define $\mathsf{Adm}(i, n) \leftrightarrow i \in n$. In this case, $n \in \mathcal{P}(S)$ is the set of admissible indices, indices are not necessarily numbers, and n does not have to be an interval nor even an ordered set. Another theory may interpret I as $\mathbb{Z}$ and L as the set of pairs of the form $(addr, n)$, where addr is a binary number representing the start address of the array in memory, and n is an integer representing the number of admissible indices. Then, the axiom defining admissibility would be $\mathsf{Adm}(i, (addr, n)) \leftrightarrow 0 \le i < n$, where the start address plays no role in characterizing the set of admissible indices. With this axiom for admissibility, we can have two distinct arrays a and b with the same set of admissible indices $\{0, 1, 2, 3, 4\}$, but* $\mathsf{len}(a) = (000100, 5) \neq (010100, 5) = \mathsf{len}(b)$ *because a and b start at distinct addresses.*

Let a and b be variables of an $I \overset{L}{\Rightarrow} V$ sort, v and u be variables of sort V, i and j be variables of sort I, and n and m be variables of sort L. The axiomatization of ArrL includes *congruence* axioms (1)-(4), *select-over-store* axioms (5)-(6), an axiom saying that length is unaffected by a store (7), and an *extensionality* axiom (8):

$$\forall a, v, i, j.\ (a \simeq b \wedge i \simeq j) \to \mathsf{select}(a, i) \simeq \mathsf{select}(b, j), \qquad (1)$$

$$\forall a, v, i, j, u, v.\ (a \simeq b \wedge i \simeq j \wedge u \simeq v) \rightarrow \mathsf{store}(a,i,u) \simeq \mathsf{store}(b,j,v), \tag{2}$$

$$\forall a, b.\ a \simeq b \rightarrow \mathsf{len}(a) \simeq \mathsf{len}(b), \tag{3}$$

$$\forall n, m, i, j.\ (n \simeq m \wedge i \simeq j \wedge \mathsf{Adm}(i,n)) \rightarrow \mathsf{Adm}(j,m), \tag{4}$$

$$\forall a, v, i.\ \ \mathsf{Adm}(i, \mathsf{len}(a)) \rightarrow \mathsf{select}(\mathsf{store}(a,i,v),i) \simeq v, \tag{5}$$

$$\forall a, v, i, j.\ i \not\simeq j \rightarrow \mathsf{select}(\mathsf{store}(a,i,v),j) \simeq \mathsf{select}(a,j), \tag{6}$$

$$\forall a, i, v.\ \ \mathsf{len}(\mathsf{store}(a,i,v)) \simeq \mathsf{len}(a), \tag{7}$$

$$\forall a, b.\ [\mathsf{len}(a) \simeq \mathsf{len}(b) \wedge (\forall i.\ \ \mathsf{Adm}(i, \mathsf{len}(a)) \rightarrow \mathsf{select}(a,i) \simeq \mathsf{select}(b,i))] \rightarrow a \simeq b. \tag{8}$$

The other direction of axiom (8) is omitted as it follows from the congruence axioms. Axiom (6) is the same as in theories Arr_0 and Arr. Axiom (5), (7), and (8) are new. Axiom (8) says that a and b are equal if they have the same length and the same elements at *all admissible indices*. In other words, if a and b are different, either they differ in length or at an admissible index. On the other hand, in theory Arr, if a and b are different, they differ at an arbitrary index. As the following example shows, neither theory entails the extensionality axiom of the other.

Example 2. *Picture a model of* Arr*, extended with an interpretation of* len *and* Adm*, where arrays a and b have the same length, agree at all admissible indices, but disagree at an index that is not admissible: $a \simeq b$ is false in this model and hence the extensionality axiom of* ArrL *(axiom (8)) also is false. On the other hand, picture a model of* ArrL *where arrays a and b agree at all indices but have different lengths: $a \simeq b$ is false in this model and hence the extensionality axiom of* Arr *also is false. Note that this can happen even if a and b have the same set of admissible indices, as in the third case of Example 1, where arrays starting at distinct addresses have different lengths and hence are different. This interpretation of array equality is common in programming languages.*

Axioms (5) and (7) are designed having in mind the intuition that a store at an inadmissible index leaves the array unchanged. Therefore, first, the length is unchanged (axiom (7)), and, second, the value argument of the store is lost, so that axiom (5) requires index i to be admissible. The *theory of arrays with MaxDiff* [6] makes the same choices in the special case where the admissible indices of an array form an interval $[0, n)$.

Alternatively, one can have a theory where a store at an inadmissible index i in array a changes the length. This is captured by replacing axiom (7) with

$$\forall a, i, v.\ \ \mathsf{Adm}(i, \mathsf{len}(a)) \rightarrow \mathsf{len}(\mathsf{store}(a,i,v)) \simeq \mathsf{len}(a). \tag{9}$$

Then, one can drop $\mathsf{Adm}(i, \mathsf{len}(a))$ from the antecedent of axiom (5) restoring the *select-over-store* axioms of theory Arr. In the resulting theory (like in Arr), if $a \simeq \mathsf{store}(a,i,v)$, then by congruence $\mathsf{select}(a,i) \simeq \mathsf{select}(\mathsf{store}(a,i,v),i)$, and by the select-over-store axiom $\mathsf{select}(a,i) \simeq v$. In other words, $\mathsf{select}(a,i) \not\simeq v$ implies $a \not\simeq \mathsf{store}(a,i,v)$. However, by axiom (8) with b replaced by $\mathsf{store}(a,i,v)$, if $\mathsf{select}(a,i) \not\simeq v$ and $\neg\,\mathsf{Adm}(i, \mathsf{len}(a))$, then $\mathsf{len}(a) \not\simeq \mathsf{len}(\mathsf{store}(a,i,v))$. One way of further specifying the change of length is to impose that index i be admissible in $\mathsf{store}(a,i,v)$. This is obtained by adding the axiom

$$\forall a, j, i, v.\ (\mathsf{Adm}(j, \mathsf{len}(a)) \vee j \simeq i) \rightarrow \mathsf{Adm}(j, \mathsf{len}(\mathsf{store}(a,i,v))). \tag{10}$$

Models of this theory include data structures such as *finite maps* and *vectors* (aka *dynamic arrays*) which satisfy stronger versions of axiom (10). Maps satisfy the double implication

$$\forall a, j, i, v.\ (\mathsf{Adm}(j, \mathsf{len}(a)) \vee j \simeq i) \leftrightarrow \mathsf{Adm}(j, \mathsf{len}(\mathsf{store}(a, i, v))). \tag{11}$$

Vectors assume that indices are integers and satisfy the double implication

$$\forall a, j, i, v.\ (\mathsf{Adm}(j, \mathsf{len}(a)) \vee j \leq i) \leftrightarrow \mathsf{Adm}(j, \mathsf{len}(\mathsf{store}(a, i, v))), \tag{12}$$

which captures the growth of the vector as an effect of the store.

4. CDSAT for Nondisjoint Theories Sharing Predicates

In this section we summarize the CDSAT framework and we modify it sparingly to accommodate shared predicates. CDSAT works with *assignments* of values to terms, including formulae that get Boolean values. Thus, CDSAT treats Boolean and first-order assignments, initial assignments and generated assignments, as uniformly as possible. The values for a theory $\mathcal{T}$ with signature Σ are provided by a *conservative theory extension* $\mathcal{T}^+$ with signature Σ^+ that adds as many constant symbols as needed to name all the individuals in the sets used to interpret the sorts of $\mathcal{T}$. The added constants are called *$\mathcal{T}$-values*. In this way terms and values are kept separate. Conservativity means that if a Σ-formula is $\mathcal{T}$-satisfiable then it is also $\mathcal{T}^+$-satisfiable. All extensions add the values true and false, so that true and false are $\mathcal{T}$-values for all $\mathcal{T}$. The *trivial extension* adds only true and false. The signature of $\mathcal{T}_\infty^+$ is the union of the signatures of $\mathcal{T}_1^+, \ldots, \mathcal{T}_n^+$ so that all values are $\mathcal{T}_\infty$-values. We use $\mathfrak{b}$ for true or false and $\mathfrak{c}$ for generic values of arbitrary sort. A *$\mathcal{T}$-assignment* is one where all assigned values are $\mathcal{T}$-values.

Definition 1 (Assignment). *A set $J = \{u_1 \leftarrow \mathfrak{c}_1, \ldots, u_m \leftarrow \mathfrak{c}_m\}$ is a $\mathcal{T}$-*assignment *if for all i, $1 \leq i \leq m$, u_i is a $\mathcal{T}_\infty$-term and $\mathfrak{c}_i$ is a $\mathcal{T}$-value of the same sort.*

The set of terms that *occur* in J as above is $G(J) = \{t \mid t \trianglelefteq u_i, 1 \leq i \leq m\}$. If all values in J are Boolean, J is a *Boolean assignment.* If no value in J is Boolean, J is a *first-order assignment.* The *flip* of a Boolean singleton assignment L, written $\overline{L}$, assigns the opposite Boolean value to the same formula. Standard abbreviations are l for $l \leftarrow \mathsf{true}$, $\bar{l}$ for $l \leftarrow \mathsf{false}$, $t \not\simeq u$ for $(t \simeq u) \leftarrow \mathsf{false}$, $\top$ for $(x \simeq_{\mathsf{prop}} x) \leftarrow \mathsf{true}$, and $\bot$ for $x \not\simeq_{\mathsf{prop}} x$, where x is an arbitrary variable of sort prop. We use J for generic assignments, A for generic singletons, L for Boolean singletons, and H or E for $\mathcal{T}_\infty$-assignments. An unqualified assignment is a $\mathcal{T}_\infty$-assignment. A $\mathcal{T}$-assignment is *plausible* if it does not contain both $l \leftarrow \mathsf{true}$ and $l \leftarrow \mathsf{false}$. A plausible $\mathcal{T}$-assignment may contain first-order assignments $u \leftarrow \mathfrak{c}_1$ and $u \leftarrow \mathfrak{c}_2$ with $\mathfrak{c}_1 \neq \mathfrak{c}_2$, from which CDSAT deduces $\bot$. The reason for this difference is that CDSAT can generate terms $u_1 \simeq_s u_2$ and $u_1 \not\simeq_s u_2$ from terms u_1 and u_2 of sort s, except when s is prop. The exception prevents the construction of an infinite series such as $l_1 = (l \simeq_{\mathsf{prop}} l)$, $l_2 = (l_1 \simeq_{\mathsf{prop}} l_1)$, $l_3 = (l_2 \simeq_{\mathsf{prop}} l_2)$, etc. Input assignments are assumed to be plausible and CDSAT preserves plausibility. The *view* that a theory $\mathcal{T}_k$ has of a $\mathcal{T}_\infty$-assignment H is made of the $\mathcal{T}_k$-assignments in H, plus all equalities and inequalities between terms of a $\mathcal{T}_k$-sort that are entailed by first-order assignments in H.

$$
\begin{array}{rll}
t_1{\leftarrow}\mathfrak{c}, t_2{\leftarrow}\mathfrak{c} \vdash & t_1 \simeq_s t_2 & \text{if } \mathfrak{c} \text{ is a } \mathcal{T}\text{-value of sort } s \\
t_1{\leftarrow}\mathfrak{c}_1, t_2{\leftarrow}\mathfrak{c}_2 \vdash & t_1 \not\simeq_s t_2 & \text{if } \mathfrak{c}_1 \text{ and } \mathfrak{c}_2 \text{ are distinct } \mathcal{T}\text{-values of sort } s \\
\vdash & t_1 \simeq_s t_1 & \text{(reflexivity)} \\
t_1 \simeq_s t_2 \vdash & t_2 \simeq_s t_1 & \text{(symmetry)} \\
t_1 \simeq_s t_2, t_2 \simeq_s t_3 \vdash & t_1 \simeq_s t_3 & \text{(transitivity)}
\end{array}
$$

Figure 1: Equality inference rules, where t_1, t_2, and t_3 are terms of sort s

Definition 2 (Theory view). *Given a theory $\mathcal{T}$ with set of sorts S and a $\mathcal{T}_\infty$-assignment H, the* $\mathcal{T}$-view *$H_\mathcal{T}$ of H is the $\mathcal{T}$-assignment equal to the union of the following sets:*

- $\{\ u{\leftarrow}\mathfrak{c} \mid\ u{\leftarrow}\mathfrak{c}$ *is a $\mathcal{T}$-assignment in H*$\}$
- $\{\ u_1 \simeq_s u_2 \mid u_1{\leftarrow}\mathfrak{c}, u_2{\leftarrow}\mathfrak{c}$ *are in H and have sort* $s \in S\backslash\{\mathsf{prop}\}\}$
- $\{\ u_1 \not\simeq_s u_2 \mid u_1{\leftarrow}\mathfrak{c}_1, u_2{\leftarrow}\mathfrak{c}_2$ *are in H, have sort* $s \in S\backslash\{\mathsf{prop}\}$, *and* $\mathfrak{c}_1{\neq}\mathfrak{c}_2\}$.

Note that a Boolean assignment is in every theory view. A $\mathcal{T}^+$-model $\mathcal{M}$ *endorses* a $\mathcal{T}$-assignment J, written $\mathcal{M} \models J$, if $\mathcal{M}$ satisfies $u \simeq \mathfrak{c}$ for all pairs $(u{\leftarrow}\mathfrak{c}) \in J$. If $\{u{\leftarrow}\mathfrak{c}, t{\leftarrow}\mathfrak{c}\} \subseteq J$, then $\mathcal{M}$ also satisfies $u \simeq t$. Endorsing the $\mathcal{T}$-view $J_\mathcal{T}$ of J is generally stronger than endorsing J: if $\mathcal{M} \models J_\mathcal{T}$, then $\mathcal{M}$ also satisfies $u \not\simeq t$, for all pairs $u{\leftarrow}\mathfrak{c}_1$ and $t{\leftarrow}\mathfrak{c}_2$ in J such that $\mathfrak{c}_1 \neq \mathfrak{c}_2$ and the sort of u and t is a sort of $\mathcal{T}$. A $\mathcal{T}$-assignment J is *satisfiable* if there is a $\mathcal{T}^+$-model $\mathcal{M}$ such that $\mathcal{M} \models J_\mathcal{T}$ and it is *unsatisfiable* otherwise. The relation $J \models L$ holds if $\mathcal{M} \models L$ for all $\mathcal{T}^+$-models $\mathcal{M}$ such that $\mathcal{M} \models J_\mathcal{T}$. For a $\mathcal{T}_\infty$-assignment H, we say that $\mathcal{M}$ *globally endorses* H if $\mathcal{M} \models H_{\mathcal{T}_\infty}$ also written $\mathcal{M} \models^G H$ to emphasize "globally."

Every theory $\mathcal{T}_k$ $(1 \leq k \leq n)$ is equipped with a *theory module* $\mathcal{I}_k$, whose inference rules produce inferences of the form $J \vdash_{\mathcal{I}_k} L$ (or $J \vdash_k L$ for short) where J is a $\mathcal{T}_k$-assignment and L is a Boolean assignment. All CDSAT theory modules include the *equality inference rules* in Fig. 1. CDSAT theory modules are required to be *sound*: if $J \vdash_k L$ then $J \models L$.

CDSAT works with a *trail* Γ which is a sequence of distinct singleton assignments that are either *decisions*, written ${}_?A$ to convey guessing, or *justified assignments*, written ${}_{H\vdash}A$. Decisions can be either Boolean or first-order assignments. The *justification* H in ${}_{H\vdash}A$ is a set of singleton assignments that appear *before* A in the trail. Input assignments are justified assignments with empty justification. All justified assignments are Boolean except for input first-order assignments. Given a trail $\Gamma = A_0, \ldots, A_m$, the *level* of an assignment is $\mathsf{level}_\Gamma(A_i) = 1 + \max\{\mathsf{level}_\Gamma(A_j) \mid j < i\}$, if A_i is a decision, and $\mathsf{level}_\Gamma(A_i) = \max\{\mathsf{level}_\Gamma(A) \mid A \in H\}$, if A_i is ${}_{H\vdash}A_i$ (where $\mathsf{level}_\Gamma(A_i) = 0$ if $H = \emptyset$).

The transition system of CDSAT, given in Fig. 2, comprises *trail rules* and *conflict state rules* (see [1, 2] for a detailed description). A *conflict state* is made of a trail and a conflict, where a *conflict* is an unsatisfiable assignment. Rule Decide expands the trail with a decision ${}_?A$, provided it is *acceptable* for a $\mathcal{T}$-module $\mathcal{I}$ in the $\mathcal{T}$-view of the trail.

Definition 3 (Acceptability). *A singleton $\mathcal{T}$-assignment $u{\leftarrow}\mathfrak{c}$ is* acceptable *for a $\mathcal{T}$-module $\mathcal{I}$ in a $\mathcal{T}$-assignment J, if (i) J does not assign a $\mathcal{T}$-value to u, (ii) if $u{\leftarrow}\mathfrak{c}$ is first-order, there are no $\mathcal{I}$-inferences $J' \cup \{u{\leftarrow}\mathfrak{c}\} \vdash_\mathcal{I} L$ for $J' \subseteq J$ and $\overline{L} \in J$, and (iii) u is relevant to $\mathcal{T}$ in J.*

TRAIL RULES (assume $1 \le k \le n$)			
Decide Γ	$\longrightarrow$	$\Gamma, {}_{?}A$	if A is an acceptable $\mathcal{T}_k$-assignment for $\mathcal{I}_k$ in $\Gamma_{\mathcal{T}_k}$
The next three rules share the conditions: $J \subseteq \Gamma$, $(J \vdash_k L)$, and $L \notin \Gamma$.			
Deduce Γ	$\longrightarrow$	$\Gamma, {}_{J\vdash}L$	if $\overline{L} \notin \Gamma$ and L is in $\mathcal{B}$
Fail Γ	$\longrightarrow$	unsat	if $\overline{L} \in \Gamma$ and $\mathsf{level}_\Gamma(J \cup \{\overline{L}\}) = 0$
ConflictSolve Γ	$\longrightarrow$	Γ'	if $\overline{L} \in \Gamma$, $\mathsf{level}_\Gamma(J \cup \{\overline{L}\}) > 0$, and $\langle \Gamma; J \cup \{\overline{L}\}\rangle \Longrightarrow^* \Gamma'$
CONFLICT STATE RULES (recall that $\uplus$ is disjoint union)			
UndoClear			
$\langle \Gamma; E \uplus \{A\}\rangle$	$\Longrightarrow$	$\Gamma^{\le m-1}$	if A is a first-order decision of level $m > \mathsf{level}_\Gamma(E)$
Resolve			
$\langle \Gamma; E \uplus \{A\}\rangle$	$\Longrightarrow$	$\langle \Gamma; E \cup H\rangle$	if $({}_{H\vdash}A) \in \Gamma$ and for no first-order decision $A' \in H$ $\mathsf{level}_\Gamma(A') = \mathsf{level}_\Gamma(E \uplus \{A\})$
UndoDecide			
$\langle \Gamma; E \uplus \{L\}\rangle$	$\Longrightarrow$	$\Gamma^{\le m-1}, {}_{?}\overline{L}$	if $({}_{H\vdash}L) \in \Gamma$ and for a first-order decision $A' \in H$ $m = \mathsf{level}_\Gamma(E) = \mathsf{level}_\Gamma(L) = \mathsf{level}_\Gamma(A')$
LearnBackjump			
$\langle \Gamma; E \uplus H\rangle$	$\Longrightarrow$	$\Gamma^{\le m}, {}_{E\vdash}L$	if L is a clausal form of H, L is in $\mathcal{B}$, $L \notin \Gamma$, $\overline{L} \notin \Gamma$, and $\mathsf{level}_\Gamma(E) \le m < \mathsf{level}_\Gamma(H)$

Figure 2: The CDSAT transition system

Condition (i) avoids multiple assignments to a term by the same theory and preserves plausibility. Condition (ii) blocks a first-order assignment that triggers an inference yielding a trivial conflict $\{L, \overline{L}\}$. Condition (iii) ensures that the assigned term u is *relevant*. The definition of *relevance* of a term to a theory in an assignment is *the first definition that has to be changed to accommodate shared predicates*. First, it makes sense that a $\mathcal{T}$-module $\mathcal{I}$ may decide a value for a term u if u occurs in the $\mathcal{T}$-view $\Gamma_\mathcal{T}$ of the trail and $\mathcal{T}$ has values for the sort of u. For equality, it also makes sense that $\mathcal{I}$ may decide $u \simeq t$ if u and t occur in $\Gamma_\mathcal{T}$, even if $u \simeq t$ does not, and $\mathcal{T}$ does not have values for the sort of u and t: indeed, if $\mathcal{T}$ has values for their sort, $\mathcal{I}$ can decide values for u and t, and glean the value of $u \simeq t$ by an equality inference. For shared predicates other than equality the latter subtlety is irrelevant.

Definition 4 (Nondisjoint Relevance). *Given a theory $\mathcal{T}$ with signature $\Sigma = (S, F)$ and a $\mathcal{T}$-assignment J, where $G(J)$ is the set of terms that occur in J, a term u is* relevant *to $\mathcal{T}$ in J, if either (i) $u \in G(J)$ and $\mathcal{T}$ has values for the sort of u; or (ii) u is an equality $u_1 \simeq_s u_2$ such that $u_1, u_2 \in G(J)$, $s \in S$, but $\mathcal{T}$ does not have values for sort s; or (iii) u is a Boolean term $p(u_1, \ldots, u_m)$ such that $p \in F$ is a shared predicate symbol $p\colon (s_1 \times \cdots \times s_m) \to \mathsf{prop}$, and for all i, $1 \le i \le m$, $u_i \in G(J)$ and $s_i \in S$.*

Example 3. *Consider* ArrL *with* Adm *interpreted by* LIA *(cf. Example 1). Given assignment $H = \{i \leftarrow 3,\ i \simeq j,\ \mathsf{len}(a) \simeq n,\ n \leftarrow 5,\ \mathsf{select}(\mathsf{store}(a, i, v), j) \not\simeq v\}$, the view of H for* LIA *is $H \cup \{i \not\simeq n\}$, whereas the view of H for* ArrL *contains the Boolean assignments in H and $\{i \not\simeq n\}$. The term $\mathsf{Adm}(i, n)$ does not occur in either view, but its arguments* *do. Thus,* $\mathsf{Adm}(i,n)$ *is relevant to both* LIA *and* ArrL *by Condition (iii) in Definition 4. Having the definition of* Adm*,* LIA *can decide wisely* $\mathsf{Adm}(i,n){\leftarrow}\mathsf{true}$*. If* ArrL *were to venture* $\mathsf{Adm}(i,n){\leftarrow}\mathsf{false}$*,* LIA *would detect a conflict.*

Rule Deduce expands the trail with a justified assignment ${}_{J\vdash}A$ supported by a theory inference $J \vdash_k A$ for some k, $1 \le k \le n$. These deductions cover *propagation* or *conflict detection/explanation*. Propagations put on the trail the consequences of decisions. Conflict detection detects a theory conflict. Conflict explanation transforms it into a Boolean conflict: L can be derived and $\overline{L}$ is on the trail. If such a conflict arises at level 0, rule Fail reports unsatisfiability. If such a conflict arises at a level greater than 0, the system enters conflict state with rule ConflictSolve. Rule Resolve transforms the conflict state until the conflict can be solved by either UndoClear or UndoDecide or LearnBackjump, producing a modified trail that ConflictSolve returns, allowing the search to resume. Trail $\Gamma^{\le m}$ is the *restriction* of trail Γ to its elements of level at most m (cf. UndoClear, UndoDecide, and LearnBackjump). The *clausal form* of a Boolean assignment $H = \{l_1, \ldots, l_n\}$ is $\neg l_1 \vee \ldots \vee \neg l_n$ (cf. LearnBackjump).

As theory module inferences can generate *new* (i.e., non-input) terms, every theory module $\mathcal{I}_k$ comes with a *local basis* denoted basis_k. Given a finite set X of terms (in practice, the set of input terms), $\mathsf{basis}_k(X)$ is a *finite* superset of X from which $\mathcal{I}_k$ can pick new terms. From the local bases it is possible to construct a *finite stable global basis* $\mathcal{B}$, where *stable* means $\mathsf{basis}_k(\mathcal{B}) \subseteq \mathcal{B}$ for all k, $1 \le k \le n$ (see [2] for the details of the construction). The sets produced by the local bases and hence $\mathcal{B}$ are required to be *closed*, meaning $\triangleleft$*-closed* (if u is a member so is every t such that $t \triangleleft u$) and *equality-closed* (if non-Boolean terms u and t are members so is $u \simeq t$). CDSAT checks that the terms generated during a derivation are in $\mathcal{B}$ (cf. Deduce and LearnBackjump).

CDSAT is *sound* if the theory modules are sound [1, Thm. 1]; and *terminating*, if $\mathcal{B}$ is *finite*, *closed*, and contains all input terms [1, Thm. 2]. Soundness and termination are not affected by the presence of nondisjoint theories, as long as their modules are sound, come with finite closed local bases, and there exists a $\mathcal{B}$ with the required properties.

Completeness of CDSAT requires that there is a leading theory, its module is *complete*, the other modules are *leading-theory-complete*, $\mathcal{B}$ is *stable* and contains all input terms [1, Thms. 3, 4, 5]. The notions of a module being *complete* (for its own theory) or *leading-theory-complete* do not need to be reformulated for the nondisjoint case. First, we say that a $\mathcal{T}$-module $\mathcal{I}$ *expands* a $\mathcal{T}$-assignment J by adding either a $\mathcal{T}$-assignment A that is acceptable for $\mathcal{I}$ in J (cf. Decide), or a Boolean assignment $l{\leftarrow}\mathfrak{b}$ such that $J' \vdash_{\mathcal{I}} (l{\leftarrow}\mathfrak{b})$, for $J' \subseteq J$, $(l{\leftarrow}\mathfrak{b}) \notin J$, and $l \in \mathsf{basis}(J)$ (cf. Deduce, Fail, and ConflictSolve). Then, a $\mathcal{T}$-module $\mathcal{I}$ is *complete* if whenever it cannot expand a plausible $\mathcal{T}$-assignment J, there exists a $\mathcal{T}^+$-model $\mathcal{M}$ such that $\mathcal{M} \models J$ [1, Def. 12]. Also, a $\mathcal{T}$-module $\mathcal{I}$ is *leading-theory-complete* if whenever it cannot expand a plausible $\mathcal{T}$-assignment J, then J *is leading-theory-compatible* with $\mathcal{T}$ sharing the set of terms $G(J)$ [1, Def. 14]. In the disjoint case, leading-theory-compatibility says that if $\mathcal{M}_1 \models J_{\mathcal{T}_1}$ for a $\mathcal{T}_1^+$-model $\mathcal{M}_1$, then there exists a $\mathcal{T}^+$-model $\mathcal{M}$ such that $\mathcal{M} \models J$ ($J = J_{\mathcal{T}}$ as J cannot be expanded) and $\mathcal{M}$ agrees with $\mathcal{M}_1$ on the cardinality of shared sorts and on equalities between shared terms [1, Def. 13]. Leading-theory-compatibility is *the second definition that* *changes to allow shared predicates.* The change consists of extending the treatment of equality to all shared predicates.

Definition 5 (Nondisjoint leading-theory-compatibility). *Let $\mathcal{T}_1$ be the leading theory, $\mathcal{T}$ and $\Sigma = (S, F)$ stand for $\mathcal{T}_k$ and $\Sigma_k = (S_k, F_k)$, $2 \le k \le n$, and N be a set of terms. A $\mathcal{T}$-assignment J is* leading-theory-compatible with $\mathcal{T}$ sharing N, *if for all $\mathcal{T}_1^+[\mathcal{V}_1]$-models $\mathcal{M}_1$ such that $\mathcal{M}_1 \models J_{\mathcal{T}_1}$ with $\mathsf{fv}_{\Sigma_1}(J \cup N) \subseteq \mathcal{V}_1$, there exists a $\mathcal{T}^+[\mathcal{V}]$-model $\mathcal{M}$ with $\mathsf{fv}_{\Sigma}(J \cup N) \subseteq \mathcal{V}$, such that (i) $\mathcal{M} \models J$; (ii) for all shared predicates $p \in F \cap F_1$ with $p\colon (s_1 \times \cdots \times s_m) \to \mathsf{prop}$ and for all terms $u_1, \ldots, u_m \in N$ of sorts $s_1, \ldots, s_m$, $\mathcal{M}_1(p(u_1, \ldots, u_m)) = \mathcal{M}(p(u_1, \ldots, u_m))$; and (iii) for all sorts $s \in S$, there exists a bijection f_s from domain $s^{\mathcal{M}}$ to domain $s^{\mathcal{M}_1}$ (so that $|s^{\mathcal{M}}| = |s^{\mathcal{M}_1}|$), such that for all shared predicates $p \in F \cap F_1$ with $p\colon (s_1 \times \cdots \times s_m) \to \mathsf{prop}$ and for all inhabitants $v_1, \ldots, v_m$ of $s_1^{\mathcal{M}}, \ldots, s_m^{\mathcal{M}}$, $p^{\mathcal{M}}(v_1, \ldots, v_m) = p^{\mathcal{M}_1}(f_{s_1}(v_1), \ldots, f_{s_m}(v_m))$.*

When equality is the only shared predicate, property *(ii)* in the above definition reduces to $\mathcal{M}(u_1) = \mathcal{M}(u_2)$ if and only if $\mathcal{M}_1(u_1) = \mathcal{M}_1(u_2)$ for all sorts $s \in S$ and terms $u_1, u_2 \in N$ of sort s. Property *(iii)* reduces to $|s^{\mathcal{M}}| = |s^{\mathcal{M}_1}|$ for all $s \in S$, because all models interpret equality as identity. With other shared predicates, the property is stated explicitly, relying on named bijections between the interpretations of a shared sort.

5. A CDSAT Module for Arrays with Abstract Length

In previous work we gave a CDSAT module for theory Arr [1] and proved its leading-theory-completeness [2, Thm. 4]. In this section we give a CDSAT theory module $\mathcal{I}_{\mathsf{ArrL}}$ for theory ArrL (cf. Sect. 3 for Arr, ArrL and the axioms of ArrL). The reduction to clausal form of the extensionality axiom (8) of ArrL introduces the Skolem function symbols $\mathsf{diff}\colon (I \overset{I}{\Rightarrow} V) \times (I \overset{I}{\Rightarrow} V) \to I$ that map two arrays to an index, called a *witness*, where they differ. Module $\mathcal{I}_{\mathsf{ArrL}}$ augments the equality rules of Fig. 1 with the following rules:

$$a \simeq b,\ i \simeq j,\ \mathsf{select}(a,i) \not\simeq \mathsf{select}(b,j)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{13}$$
$$a \simeq b,\ i \simeq j,\ u \simeq v,\ \mathsf{store}(a,i,u) \not\simeq \mathsf{store}(b,j,v)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{14}$$
$$a \simeq b \vdash_{\mathsf{ArrL}}\ \mathsf{len}(a) \simeq \mathsf{len}(b) \tag{15}$$
$$n \simeq m,\ i \simeq j,\ \mathsf{Adm}(i,n),\ \neg\,\mathsf{Adm}(j,m)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{16}$$
$$a \simeq c,\ b \simeq d,\ \mathsf{diff}(a,b) \not\simeq \mathsf{diff}(c,d)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{17}$$
$$i \simeq j,\ \mathsf{len}(a) \simeq n,\ \mathsf{Adm}(i,n),\ b \simeq \mathsf{store}(a,i,v),\ \mathsf{select}(b,j) \not\simeq v\ \vdash_{\mathsf{ArrL}}\ \bot \tag{18}$$
$$i \not\simeq j,\ k \simeq j,\ b \simeq \mathsf{store}(a,i,v),\ a \simeq c,\ \mathsf{select}(b,k) \not\simeq \mathsf{select}(c,j)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{19}$$
$$\mathsf{len}(\mathsf{store}(a,i,v)) \not\simeq \mathsf{len}(a)\ \vdash_{\mathsf{ArrL}}\ \bot \tag{20}$$
$$a \not\simeq b,\ \mathsf{len}(a) \simeq \mathsf{len}(b)\ \vdash_{\mathsf{ArrL}}\ \mathsf{select}(a, \mathsf{diff}(a,b)) \not\simeq \mathsf{select}(b, \mathsf{diff}(a,b)) \tag{21}$$
$$a \not\simeq b,\ \mathsf{len}(a) \simeq \mathsf{len}(b)\ \vdash_{\mathsf{ArrL}}\ \mathsf{Adm}(\mathsf{diff}(a,b), \mathsf{len}(a)) \tag{22}$$

where rules (13)-(16) correspond to axioms (1)-(4), rule (17) adds congruence for diff, rules (18)-(19) correspond to axioms (5)-(6) with premises flattened by introducing new variables, rule (20) corresponds to axiom (7), and rules (21) and (22) correspond to the clauses for axiom (8). The flattening conveys that in order to fire, for example, rule (18), it suffices to have on the trail terms of the form $b \simeq \mathsf{store}(a, i, v)$ and $\mathsf{select}(b, j) \not\simeq v$, and not necessarily the term $\mathsf{select}(\mathsf{store}(a, i, v), j) \not\simeq v$. This is relevant for completeness, because the equality rules of Fig. 1 do not include a rule for replacement of equals by equals and hence cannot deduce $\mathsf{select}(\mathsf{store}(a, i, v), j) \not\simeq v$ from $b \simeq \mathsf{store}(a, i, v)$ and $\mathsf{select}(b, j) \not\simeq v$.

The first requirement when designing a CDSAT module is that its rules are *sound*, which is satisfied by $\mathcal{I}_{\mathsf{ArrL}}$. The second requirement is that it is possible to define a *local basis*. Rules that generate $\bot$ are convenient, because they are useful for conflict detection and they are trivial for the construction of the local basis, since it suffices that it contains $\top$ (the flip of $\top$ is $\bot$). The third requirement is that the module is *leading-theory-complete*. In order to prove this property, the rules of the module must put on the trail the terms needed for defining a model. This is why rules (15), (21) and (22) produce terms other than $\bot$. Thus, the design of a CDSAT module demands a balancing act between the local basis requirement, which suggests to minimize the generation of new terms, and the completeness requirement.

The *local basis* for ArrL maps any given finite set X of terms to a set $\mathsf{basis}_{\mathsf{ArrL}}(X)$ defined as the smallest closed set Y such that $X \subseteq Y$, $\top \in Y$, and:

1. For all terms l_1 and l_2 of sort prop that occur as subterms of terms in Y with select, store, len, or diff as top symbol, $(l_1 \simeq_{\mathsf{prop}} l_2) \in Y$;
2. For all terms $t \in Y$ and $u \in Y$ of the same array sort, $\mathrm{Ded}(t, u) \subseteq Y$, where $\mathrm{Ded}(t, u)$ contains precisely the terms $\mathsf{len}(t)$, $\mathsf{select}(t, \mathsf{diff}(t, u))$, $\mathsf{select}(u, \mathsf{diff}(t, u))$, $\mathsf{Adm}(\mathsf{diff}(t, u), \mathsf{len}(t))$, and $\mathsf{Adm}(\mathsf{diff}(t, u), \mathsf{len}(u))$.

Clause (1) adds equalities between formulae that may be needed (e.g., picture arrays where indices or elements are Boolean) and whose presence is not guaranteed by equality-closedness that applies to non-Boolean terms. Clause (2) adds the terms that may be generated by rules (15), (21), and (22).

The following reasoning shows that Y is finite. For Clause (1), for terms l_1 and l_2 of sort prop, let $P^1(l_1, l_2)$ stand for the conjunction of the conditions $l_1 \lhd t$, $t \in X$, $l_2 \lhd u$, $u \in X$, $top(t) \in \{\mathsf{select},\ \mathsf{store},\ \mathsf{len},\ \mathsf{diff}\}$, and $top(u) \in \{\mathsf{select},\ \mathsf{store},\ \mathsf{len},\ \mathsf{diff}\}$. Let $\mathrm{Sat}^1(X)$ be the union of X and $\bigcup_{P^1(l_1,l_2)}\{l_1 \simeq_{\mathsf{prop}} l_2\}$. For Clause (2), for all terms t of an array sort, let $\mathrm{depth}(t)$ be the number of occurrences of the array sort constructor $\Rightarrow$ in the sort of t. Let $k = max\{\mathrm{depth}(t) \mid t \in X,\ t \text{ of an array sort}\}$. For terms t and u of the same array sort, let $P^2_q(t, u)$ stand for $t \in X \wedge u \in X \wedge \mathrm{depth}(t) = q \wedge \mathrm{depth}(u) = q$. Let $\mathrm{Sat}^2_q(X)$ be the union of X and $\bigcup_{P^2_q(t,u)} \mathrm{Ded}(t, u)$. All terms in $\bigcup_{P^2_q(t,u)} \mathrm{Ded}(t, u)$ have depth smaller than q, because even in the case of an array-indexed array the depth of an array term used as index is smaller. Thus, the closure $Y = \mathrm{Sat}^1(\mathrm{Sat}^2_1(\mathrm{Sat}^2_2(\ldots \mathrm{Sat}^2_k(X) \ldots)))$ is finite. A theory ArrL with array sorts $s_1, \ldots, s_n$ ($n > 1$) can be viewed as the union of n theories ArrL with one array sort each. Then, the above finiteness argument is an instance of the proof showing how to construct a finite global basis for a union of theories from the local bases of the component theories [2].

As arrays represent functions that can be updated, a model of Arr interprets an array as an *updatable function* from indices (meaning a set interpreting the sort of indices) to elements (meaning a set interpreting the sort of elements). Given generic sets $\mathcal{U}$ and $\mathcal{V}$, and the set $\mathcal{V}^{\mathcal{U}}$ of the functions from $\mathcal{U}$ to $\mathcal{V}$, we say that $\mathcal{W} \subseteq \mathcal{V}^{\mathcal{U}}$ is an *updatable function set from $\mathcal{U}$ to $\mathcal{V}$*, if every function obtained by a finite number of updates to a function in $\mathcal{W}$ is in $\mathcal{W}$. A model of Arr interprets an array sort as an updatable function set. A model of ArrL interprets an array as a *partial updatable function*, whose domain of definition is the set of admissible indices. Therefore, the cardinality of an array sort depends on the interpretation of Adm.

Definition 6 (ArrL-suitability). *A leading theory $\mathcal{T}_1$ is* suitable for ArrL, *or* ArrL-suitable, *if it has all the sorts in S_{ArrL}, it shares with* ArrL *only the equality symbols $\simeq_s$ for all sorts $s \in S_{\mathsf{ArrL}}$ and the symbol* Adm, *and for all $\mathcal{T}_1$-models $\mathcal{M}_1$ and array sorts $I \overset{L}{\Rightarrow} V$ there exists a length-indexed collection $(X_n)_{n \in L^{\mathcal{M}_1}}$ of nonempty sets such that*

$$|(I \overset{L}{\Rightarrow} V)^{\mathcal{M}_1}| = |\biguplus_{n \in L^{\mathcal{M}_1}} X_n|$$

where X_n is an updatable function set from $I_n = \{i \mid i \in I^{\mathcal{M}_1} \wedge \mathsf{Adm}^{\mathcal{M}_1}(i, n)\}$ to $V^{\mathcal{M}_1}$ for all $n \in L^{\mathcal{M}_1}$.

The set X_n is the set of updatable functions that interprets the arrays of length n. The functions in X_n are partial as they are defined only on the set I_n of admissible indices for length n and not on the set $I^{\mathcal{M}_1}$ of all indices. Note that the interpretation of select remains nonetheless a total function, because every term $\mathsf{select}(a, i)$ is interpreted. ArrL-suitability does not restrict the realm of theories with which ArrL can be combined, because ArrL-suitability merely formalizes sensible requirements on the cardinalities of array sorts. As usual in CDSAT, the leading theory simply aggregates appropriately the requirements on cardinalities coming from the theories in the union.

Example 4. *Consider the first case of Example 1. Suppose that* ArrL *interprets also V as $\mathbb{Z}$. A leading theory that interprets L, I, and* Adm *as stipulated by* LIA, *and V as stipulated by* ArrL *is* ArrL*-suitable: for all $n \in \mathbb{Z}$, the set I_n of admissible indices is $\{i \mid i \in \mathbb{Z} \wedge 0 \leq i < n\}$. Since X_n is countably infinite for all n, $n > 0$, the cardinality of the interpretation of $I \overset{L}{\Rightarrow} V$ is countably infinite. Suppose that* ArrL *interprets V as a finite set of cardinality m ($m > 0$). A leading theory that interprets L, I, and* Adm *as stipulated by* LIA, *and V as stipulated by* ArrL *is* ArrL*-suitable: since X_n has cardinality m^n for all n, $n > 0$, the cardinality of the interpretation of $I \overset{L}{\Rightarrow} V$ is countably infinite. In both cases, a leading theory that interprets $I \overset{L}{\Rightarrow} V$ as being finite is not* ArrL*-suitable.*

Example 5. *Consider the union of* ArrL *and the theory* BV *of bitvectors, where* $\mathsf{BV}[n]$ *is the set of bitvectors of length n. Assume that* BV *interprets I as* $\mathsf{BV}[1]$, *L as* $\mathsf{BV}[2]$, *and* Adm *as* true *everywhere except for the pairs $(0, 00)$, $(1, 00)$, and $(1, 01)$. Suppose that the two theories share also V and that* BV *interprets it as* $\mathsf{BV}[1]$. *A leading theory that interprets L, I,* Adm, *and V as stipulated by* BV *is* ArrL*-suitable: the sets of admissible indices are $I_{00} = \emptyset$, $I_{01} = \{0\}$, and $I_{10} = I_{11} = \{0, 1\}$, so that the cardinalities of the updatable function sets are $|X_{00}| = 2^0 = 1$, $|X_{01}| = 2^1 = 2$, and $|X_{10}| = |X_{11}| = 2^2 = 4$, and the cardinality of the interpretation of $I \overset{L}{\Rightarrow} V$ is 11. On the other hand, a leading theory that interprets $I \overset{L}{\Rightarrow} V$ as being countably infinite is not* ArrL*-suitable.*

The extension ArrL^+ for ArrL may either be trivial, or add a countably infinite set of ArrL-values for each sort in $S\backslash\{\mathsf{prop}\}$. We prove that $\mathcal{I}_{\mathsf{ArrL}}$ is leading-theory-complete assuming that ArrL^+ is nontrivial.

Lemma 1. *If J is a plausible ArrL-assignment that $\mathcal{I}_{\mathsf{ArrL}}$ cannot expand, for all terms t of an array sort, if t is in $G(J)$, then the term $\mathsf{len}(t)$ is also in $G(J)$.*

Proof: By reflexivity $(t \simeq t) \in J$, by rule (15) $(\mathsf{len}(t) \simeq \mathsf{len}(t)) \in J$, so that $\mathsf{len}(t) \in G(J)$.

This lemma (and the form of rule (15)) may be surprising, as one may expect that $\mathcal{I}_{\mathsf{ArrL}}$ needs to be concerned only with the lengths of arrays that differ. The point is that in ArrL the length is an essential part of an array (since the definition of arrays sorts as $I \overset{L}{\Rightarrow} V$), and the model construction in the proof of leading-theory-completeness of $\mathcal{I}_{\mathsf{ArrL}}$ needs to define a length function as a step towards the functional interpretation of arrays.

Theorem 1. *Module $\mathcal{I}_{\mathsf{ArrL}}$ is leading-theory-complete for all ArrL-suitable leading theories.*

Proof: Let J be a plausible ArrL-assignment that $\mathcal{I}_{\mathsf{ArrL}}$ cannot expand. We show that J is leading-theory-compatible with ArrL sharing $G(J)$. We begin by observing that every formula $l \in G_{\mathsf{prop}}(J)$ is relevant to ArrL by Condition (i) of Definition 4, and therefore J assigns a value to l (see [2, Lemma 1, Claim 2]). For a sort s other than prop, every term $u \in G_s(J)$ is relevant to ArrL by Condition (i) of Definition 4, as ArrL^+ has (infinitely many) values for all sorts I, V, L, and $I \overset{L}{\Rightarrow} V$. Moreover, the only ArrL-inferences using first-order assignments are equality inferences, and therefore J assigns a value to every such term u (see [2, Lemma 1, Claim 3]). It follows that J assigns values to all terms in $G(J)$ (†) and $\mathsf{fv}_{\Sigma_{\mathsf{ArrL}}}(G(J)) = \mathsf{fv}_{\Sigma_{\mathsf{ArrL}}}(J)$ (see [2, Corollary 1]). Let $\mathcal{T}_1$ be an ArrL-suitable leading theory, Σ_1 its signature, $\mathcal{T}_1^+$ its extension, $\mathcal{M}_1$ a $\mathcal{T}_1^+[\mathcal{V}_1]$-model such that $\mathsf{fv}_{\Sigma_1}(J) \subseteq \mathcal{V}_1$ and $\mathcal{M}_1 \models J_{\mathcal{T}_1}$, and $(X_n)_{n \in L^{\mathcal{M}_1}}$ the length-indexed family of updatable function sets for $\mathcal{M}_1$ of Definition 6. We start the construction of the $\mathsf{ArrL}^+[\mathcal{V}]$-model $\mathcal{M}$ with $\mathsf{fv}_{\Sigma_{\mathsf{ArrL}}}(J) \subseteq \mathcal{V}$ by interpreting

- All sorts in S and all variables $t \in \mathsf{fv}_{\Sigma_{\mathsf{ArrL}}}(J)$ as $\mathcal{M}_1$ does;
- The shared predicate Adm as $\mathcal{M}_1$ does, to get Parts (ii) and (iii) of Definition 5;
- All ArrL-values $\mathfrak{c}$ such that $(t \leftarrow \mathfrak{c}) \in J$ as $\mathcal{M}_1$ interprets t; and
- All other ArrL-values arbitrarily.

We need to define how $\mathcal{M}$ interprets symbols len, store, select, and diff. To this end, for every inhabitant a of $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$, we construct a *functional interpretation* mapping indices in $I^{\mathcal{M}}$ to elements in $V^{\mathcal{M}}$. More precisely, we will define

- A function *len* from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to $L^{\mathcal{M}}$ mapping arrays to lengths;
- A function ψ from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to $\biguplus_{n \in L^{\mathcal{M}_1}} X_n$ such that $\psi(a)$ is in X_n for $n = \mathit{len}(a)$, so that $\psi(a)$ is an updatable function from the set of admissible indices I_n to $V^{\mathcal{M}}$;
- A function ϕ from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to an updatable function set from $I^{\mathcal{M}}$ to $V^{\mathcal{M}}$ so that $\phi(a)$ is a total updatable function from $I^{\mathcal{M}}$ to $V^{\mathcal{M}}$ that agrees with $\psi(a)$ on I_n;
- A function *diff* from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}} \times (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to $I^{\mathcal{M}}$ mapping pairs of arrays to indices.

The functions *len*, ψ, ϕ, and *diff* will be used to construct the $\mathcal{M}$-interpretation of symbols len, store, select, and diff, respectively. We build this interpretation so as to satisfy:

1. The axioms of ArrL so that $\mathcal{M}$ is an $\mathsf{ArrL}^+[\mathsf{fv}_{\Sigma_{\mathsf{Arr}}}(J)]$-model;
2. The assignment J so that $\mathcal{M} \models J$ and Part (i) of Definition 5 is fulfilled;
3. The cardinality constraints conveyed by $\mathcal{M}_1$ to get Part (iii) of Definition 5.

For (3), we make sure that ψ is a bijection from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to $\biguplus_{n \in L^{\mathcal{M}}} X_n$. In order to define the functional interpretations of inhabitants of $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ we pick functions $f_n \in X_n$ (i.e., from I_n to $V^{\mathcal{M}}$) for all n in $L^{\mathcal{M}}$, and we complete every f_n into a total function g_n from $I^{\mathcal{M}}$ to $V^{\mathcal{M}}$. These functions will be used as defaults in the construction. The rest of the construction is subdivided in four parts. We start by considering those inhabitants of $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ that are used by $\mathcal{M}_1$ to interpret terms in $G(J)$. Let Y be the finite subset of $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ consisting of those elements a such that $\mathcal{M}_1(t) = a$ for some term $t \in G(J)$. The first step is to define len_Y, ϕ_Y, and ψ_Y, the respective cores of *len*, ψ, and ϕ that are only defined on Y.

1. *Definition of len_Y, ϕ_Y, and ψ_Y*:
   Let a be an element of Y with $a = \mathcal{M}_1(t)$ for term $t \in G(J)$. By Lemma 1, $\mathsf{len}(t) \in G(J)$. Model $\mathcal{M}_1$ sees $\mathsf{len}(t)$ as a variable in $\mathsf{fv}_{\Sigma_1}(J)$, since len is a Σ_1-foreign symbol. We define $len_Y(a) = \mathcal{M}_1(\mathsf{len}(t))$. Let $\mathcal{R}_a \subseteq I^{\mathcal{M}} \times V^{\mathcal{M}}$ be the set of index-element pairs dictated by J. Formally, R_a is the relation defined by the union of three sets:
   $\{(\mathcal{M}_1(i), \mathcal{M}_1(t[i])) \mid \mathsf{select}(t,i) \in G(J),\ \mathcal{M}_1(t) = a\}$
   $\{(\mathcal{M}_1(i), \mathcal{M}_1(u)) \mid \mathsf{store}(t,i,u) \in G(J),\ \mathcal{M}_1(\mathsf{store}(t,i,u)) = a,\ \mathcal{M}_1(i) \in I_{len_Y(a)}\}$
   $\{(\mathcal{M}_1(i), \mathcal{M}_1(t[i])) \mid \mathsf{store}(t,j,u) \in G(J),\ \mathsf{select}(t,i) \in G(J),$
   $\mathcal{M}_1(\mathsf{store}(t,j,u)) = a,\ \mathcal{M}_1(i) \neq \mathcal{M}_1(j)\}$.
   In other words, $\mathcal{R}_a$ is dictated by the terms in $G(J)$ where either select is applied to an array term that $\mathcal{M}_1$ interprets as a or the application of store forms an array term that $\mathcal{M}_1$ interprets as a. Since $G(J)$ is finite, $\mathcal{R}_a$ is finite. Also, $\mathcal{R}_a$ is a functional relation from $I^{\mathcal{M}}$ to $V^{\mathcal{M}}$, because otherwise $\mathcal{I}_{\mathsf{ArrL}}$ could expand J by rules (18)-(19). Let $\phi_Y(a)$ be the total function that is identical to $\mathcal{R}_a$ where $\mathcal{R}_a$ is defined, and maps every $e \in I^{\mathcal{M}_1}$ where $\mathcal{R}_a$ is undefined to $g_n(e) \in V^{\mathcal{M}_1}$ for $n = len_Y(a)$. Let $\psi_Y(a)$ be the restriction of $\phi_Y(a)$ to I_n. Since $\mathcal{R}_a$ is finite, $\phi_Y(a)$ differs from g_n by finitely many updates. Hence $\psi_Y(a)$ differs from f_n by finitely many updates, so that it is in X_n. The second step is to show that ψ_Y is injective and in the same context define $diff_Y$. The injectivity of ψ_Y will allow us to define ψ, *len*, ϕ, and *diff* as extensions of ψ_Y, len_Y, ϕ_Y, and $diff_Y$.
2. *Injectivity of ψ_Y and definition of $diff_Y$*:
   By way of contradiction, suppose that there are two elements $a, b \in Y$ such that $a \neq b$ and $\psi_Y(a) = \psi_Y(b)$. Since $\psi_Y(a)$ is a function in X_n for $n = len_Y(a)$ and $\psi_Y(b)$ is a function in X_m for $m = len_Y(b)$, the equality $\psi_Y(a) = \psi_Y(b)$ means that X_n and X_m have non-empty intersection. Since the collection $(X_n)_{n \in L^{\mathcal{M}}}$ is pairwise disjoint, it must be $n = m$. Since $a, b \in Y$, we have $a = \mathcal{M}_1(t)$ and $b = \mathcal{M}_1(u)$ for some terms $t, u \in G(J)$. This means that $\mathcal{M}_1 \models t \not\simeq u$. By (†) J assigns values to t and u,

and therefore it also assigns a truth value $\mathfrak{b}$ to $t \simeq u$, because otherwise an equality inference could expand J. Also, $((t \simeq u) \leftarrow \mathfrak{b}) \in J_{\mathcal{T}_1}$ by definition of theory view. Since $\mathcal{M}_1 \models t \not\simeq u$ and $\mathcal{M}_1 \models J_{\mathcal{T}_1}$, the truth value $\mathfrak{b}$ must be false, or, equivalently, $(t \not\simeq u) \in J$. Moreover by Lemma 1, $\mathsf{len}(t)$ and $\mathsf{len}(u)$ are also in $G(J)$, and J assigns them values by (†). Thus, J assigns a truth value $\mathfrak{b}'$ to $\mathsf{len}(t) \simeq \mathsf{len}(u)$ and so does $J_{\mathcal{T}_1}$. Since $\mathit{len}_Y(a) = \mathit{len}_Y(b)$, by definition of len_Y we have $\mathsf{len}^{\mathcal{M}_1}(a) = \mathsf{len}^{\mathcal{M}_1}(b)$. Since $\mathcal{M}_1 \models J_{\mathcal{T}_1}$, the truth value $\mathfrak{b}'$ must be true (i.e., $(\mathsf{len}(t) \simeq \mathsf{len}(u)) \in J$). By rule (21), also $t[\mathsf{diff}(t,u)] \not\simeq u[\mathsf{diff}(t,u)]$ is in J (*) and hence in $J_{\mathcal{T}_1}$. Since $\mathcal{M}_1 \models J_{\mathcal{T}_1}$, it follows that $\mathcal{M}_1(t[\mathsf{diff}(t,u)]) \neq \mathcal{M}_1(u[\mathsf{diff}(t,u)])$. Now we define diff_Y. By (*) $\mathsf{diff}(t,u) \in G(J)$. Model $\mathcal{M}_1$ sees $\mathsf{diff}(t,u)$ as a variable in $\mathsf{fv}_{\Sigma_1}(J)$, since diff is a Σ_1-foreign symbol. For all $a, b \in Y$, if $a \neq b$ and $\mathit{len}_Y(a) = \mathit{len}_Y(b)$, let $\mathit{diff}_Y(a,b) = \mathcal{M}_1(\mathsf{diff}(t,u))$, and let $\mathit{diff}_Y(a,b)$ be arbitrary otherwise. We resume the proof of the injectivity of ψ_Y. By rule (22), also $\mathsf{Adm}(\mathsf{diff}(t,u), \mathsf{len}(t))$ is in J and hence in $J_{\mathcal{T}_1}$. Since $\mathcal{M}_1 \models J_{\mathcal{T}_1}$, it follows that $\mathcal{M}_1(\mathsf{diff}(t,u))$ is an admissible index (i.e., it is in I_n for $n = \mathsf{len}^{\mathcal{M}_1}(a)$). By definition of $\psi_Y(a)$ (based on $\mathcal{R}_a$) for a generic a, we have:

$$\psi_Y(a)(\mathcal{M}_1(\mathsf{diff}(t,u))) = \mathcal{M}_1(t[\mathsf{diff}(t,u)])$$
$$\psi_Y(b)(\mathcal{M}_1(\mathsf{diff}(t,u))) = \mathcal{M}_1(u[\mathsf{diff}(t,u)]).$$

Since the two right hand sides are different, the two left hand sides are also different, so that $\psi_Y(a) \neq \psi_Y(b)$, a contradiction.

3. *Definition of ψ, len, ϕ, and diff*:
   - Since ψ_Y is an injective function from Y to $\biguplus_{n \in L^{\mathcal{M}}} X_n$, we can extend it to a bijection ψ from $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ to $\biguplus_{n \in L^{\mathcal{M}}} X_n$ which have the same cardinality.
   - For all $a \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ let $\mathit{len}(a)$ be the unique n in $L^{\mathcal{M}}$ such that $\psi(a)$ is in X_n. Note that for $a \in Y$ we have $\mathit{len}(a) = \mathit{len}_Y(a)$.
   - For all $a \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$, if $a \in Y$ let $\phi(a) = \phi_Y(a)$; otherwise, let $\phi(a)$ be the function that agrees with $\psi(a)$ on I_n, where $n = \mathit{len}(a)$, and with g_n everywhere else.
   - For all $a, b \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$, if $a, b \in Y$ let $\mathit{diff}(a,b) = \mathit{diff}_Y(a,b)$; otherwise, if $a = b$ or ($a \neq b$ and $\mathit{len}(a) \neq \mathit{len}(b)$), let $\mathit{diff}(a,b)$ be arbitrary. If $a \neq b$ and $\mathit{len}(a) = \mathit{len}(b) = n$, let $\mathit{diff}(a,b) = j$ for any index $j \in I_n$ such that $\psi(a)(j) \neq \psi(b)(j)$, where at least one such j exists, because $a \neq b$ implies $\psi(a) \neq \psi(b)$ by injectivity of ψ.
4. *How $\mathcal{M}$ interprets* len, diff, select, *and* store *for all array sorts $I \overset{L}{\Rightarrow} V$*:
   - For all $a \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ let $\mathsf{len}^{\mathcal{M}}(a) = \mathit{len}(a) \in L^{\mathcal{M}}$;
   - For all $a, b \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ let $\mathsf{diff}^{\mathcal{M}}(a,b) = \mathit{diff}(a,b) \in I^{\mathcal{M}}$;
   - For all pairs $(a,e) \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}} \times I^{\mathcal{M}}$ let $\mathsf{select}^{\mathcal{M}}(a,e) = \phi(a)(e) \in V^{\mathcal{M}}$;
   - For all triples $(a,e,v) \in (I \overset{L}{\Rightarrow} V)^{\mathcal{M}} \times I^{\mathcal{M}} \times V^{\mathcal{M}}$ we define $\mathsf{store}^{\mathcal{M}}(a,e,v)$ by considering two cases with $\mathit{len}(a) = n$:
     - If $e \notin I_n$: let $\mathsf{store}^{\mathcal{M}}(a,e,v) = a$;
     - If $e \in I_n$: let f be the function from I_n to $V^{\mathcal{M}}$ that maps e to v and

every other $j \in I_n$ to $\psi(a)(j) \in V^{\mathcal{M}}$; function f is in X_n as it differs from $\psi(a) \in X_n$ by one update; since ψ is a bijection, take $\psi^{-1}(f)$ which is in $(I \overset{L}{\Rightarrow} V)^{\mathcal{M}}$ and set $\mathsf{store}^{\mathcal{M}}(a, e, v) = \psi^{-1}(f)$;

Part (ii) of Definition 5 for equality follows by induction on the term structure.

The claim holds also if ArrL^+ is the trivial extension. The proof is similar, except that non-Boolean terms are not assigned ArrL-values. Leading-theory-completeness is preserved if $\mathcal{I}_{\mathsf{ArrL}}$ is enriched with rules obtained from those deriving $\bot$ by removing the last premise and adding its flip as conclusion (see [2, Lemma 2]).

6. Completeness of CDSAT in the Nondisjoint Case

In this section we show that CDSAT is complete also in the case of nondisjoint theories sharing predicates. Let a *predicate-sharing union* of theories be a union $\mathcal{T}_\infty$ of theories $\mathcal{T}_1, \ldots, \mathcal{T}_n$, such that the signatures are disjoint or share predicate symbols, and there exists a leading theory, say $\mathcal{T}_1$, which has all sorts and all shared symbols in its signature.

As a consequence of generalizing leading-theory-compatibility to a predicate-sharing union, the concept of *model-describing assignment* from [1, Def. 19] is generalized accordingly. Preliminarily, given an assignment H, the set $\mathcal{V}_{\mathsf{sh}}(H)$ of the shared terms in H contains the left-hand sides of pairs in H and all their subterms that are shared variables or foreign terms for any theory [1, Def. 18]. An assignment H is *model-describing* if (1) there exists a $\mathcal{T}_1^+[\mathcal{V}]$-model $\mathcal{M}_1$ such that $\mathcal{M}_1 \models H_{\mathcal{T}_1}$ (assuming $\mathsf{fv}_{\Sigma_1}(H_{\mathcal{T}_1}) \subseteq \mathcal{V}$), and (2) for all k, $2 \leq k \leq n$, the theory view $H_{\mathcal{T}_k}$ is leading-theory-compatible with $\mathcal{T}_k$ sharing $\mathcal{V}_{\mathsf{sh}}(H)$. Note how the generic assignment J and the generic set N of shared terms of Definition 5 are replaced by $H_{\mathcal{T}_k}$ and $\mathcal{V}_{\mathsf{sh}}(H)$.

The core of the proof of completeness is to show that a model-describing assignment is globally endorsed [1, Thm. 4]. The generalization of that statement to a predicate-sharing union, below, requires generalizing its proof.

Theorem 2. *In a predicate-sharing union of theories, if an assignment H is model-describing, there exists a $\mathcal{T}_\infty^+[\mathsf{fv}(H)]$-model $\mathcal{M}$ such that $\mathcal{M} \models^G H$.*

Proof: The proof is structured in eight short parts like that of [1, Thm. 4]. In order to accommodate shared predicates it suffices to modify Parts (2) and (3). Therefore, we reproduce Parts (1), (2), and (3), referring the reader to the proof of [1, Thm. 4] for the remaining ones.

1. *Existence of a leading-theory model $\mathcal{M}_1$*: by the hypothesis that H is model-describing, there exists a $\mathcal{T}_1^+[\mathcal{V}_1]$-model $\mathcal{M}'_1$, with $\mathsf{fv}_{\Sigma_1}(H_{\mathcal{T}_1}) \subseteq \mathcal{V}_1$, such that $\mathcal{M}'_1 \models H_{\mathcal{T}_1}$. Note that for all k, $1 \leq k \leq n$, $\mathsf{fv}_{\Sigma_k}(H_{\mathcal{T}_k}) = \mathsf{fv}_{\Sigma_k}(H) \subseteq \mathsf{fv}_{\Sigma_k}(\mathcal{V}_{\mathsf{sh}}(H))$ $(*)$. Thus, we have $\mathsf{fv}_{\Sigma_1}(H) \subseteq \mathcal{V}_1$, but there may be terms in $\mathsf{fv}_{\Sigma_1}(\mathcal{V}_{\mathsf{sh}}(H))$ that are not in $\mathcal{V}_1$. Therefore, we pick arbitrary elements in the domains of $\mathcal{M}'_1$ to interpret terms in $\mathsf{fv}_{\Sigma_1}(\mathcal{V}_{\mathsf{sh}}(H)) \setminus \mathcal{V}_1$, if any, and we extend $\mathcal{M}'_1$ into a $\mathcal{T}_1^+[\mathsf{fv}_{\Sigma_1}(\mathcal{V}_{\mathsf{sh}}(H))]$-model $\mathcal{M}_1$ such that $\mathcal{M}_1 \models H_{\mathcal{T}_1}$.

2. *Existence of the other $\mathcal{T}_k$-models $\mathcal{M}_k$*: by the hypothesis that H is model-describing, for all k, $2 \le k \le n$, there exists a $\mathcal{T}_k^+[\mathcal{V}_k]$-model $\mathcal{M}_k$ where $\mathsf{fv}_{\Sigma_k}(H_{\mathcal{T}_k} \cup \mathcal{V}_{\mathsf{sh}}(H)) \subseteq \mathcal{V}_k$ and hence $\mathsf{fv}_{\Sigma_k}(\mathcal{V}_{\mathsf{sh}}(H)) \subseteq \mathcal{V}_k$ by $(*)$, with the following properties: (i) $\mathcal{M}_k \models H_{\mathcal{T}_k}$, (ii) for all sorts $s \in S_k$, there exists a bijection f^k_s from domain $s^{\mathcal{M}_k}$ to domain $s^{\mathcal{M}_1}$, such that for all shared predicates $p \in F_k \cap F_1$ with $p\colon (s_1 \times \cdots \times s_m) \to \mathsf{prop}$: (iii) for all terms $u_1, \ldots, u_m \in \mathcal{V}_{\mathsf{sh}}(H)$ of sorts $s_1, \ldots, s_m$, $\mathcal{M}_1(p(u_1, \ldots, u_m)) = \mathcal{M}_k(p(u_1, \ldots, u_m))$, and (iv) for all inhabitants $v_1, \ldots, v_m$ of $s_1^{\mathcal{M}_k}, \ldots, s_m^{\mathcal{M}_k}$, $p^{\mathcal{M}_k}(v_1, \ldots, v_m) = p^{\mathcal{M}_1}(f^k_{s_1}(v_1), \ldots, f^k_{s_m}(v_m))$.
3. *Bijection between any $\mathcal{M}_k$ and $\mathcal{M}_1$*:
   For all k, $1 \le k \le n$, we construct a collection of bijections $\phi^s_k\colon s^{\mathcal{M}_k} \to s^{\mathcal{M}_1}$ indexed by $s \in S_k$, that satisfies the same properties as the $(f^k_s)_{s \in S_k}$ collection, but also satisfies the additional property $\phi^s_k(\mathcal{M}_k(t)) = \mathcal{M}_1(t)$ for all shared terms $t \in \mathcal{V}^s_{\mathsf{sh}}(H)$ of sort s.
   Let Y^s_1 (resp. Y^s_k) be the (finite) subset of $s^{\mathcal{M}_1}$ (resp. $s^{\mathcal{M}_k}$) consisting of those inhabitants of the form $\mathcal{M}_1(t)$ (resp. $\mathcal{M}_k(t)$) for some term t in $\mathcal{V}_{\mathsf{sh}}(H)$.
   For a family $(f^k_s)_{s \in S_k}$ of bijections satisfying the conditions of leading-theory compatibility, let $\Psi(f^k_s)_{s \in S_k}$ be the (finite) number of terms t in $\mathcal{V}_{\mathsf{sh}}(H)$ such that $f^k_s(\mathcal{M}_k(t)) \neq \mathcal{M}_1(t)$. We aim at producing a family $(\phi^s_k)_{s \in S_k}$ with $\Psi(\phi^s_k)_{s \in S_k} = 0$. We define a transformation Φ such that, if $\Psi(f^k_s)_{s \in S_k} > 0$, then $\Psi(\Phi(f^k_s)_{s \in S_k}) < \Psi(f^k_s)_{s \in S_k}$.
   Assume $f^k_s(\mathcal{M}_k(t)) \neq \mathcal{M}_1(t)$. Let $v_1 = f^k_s(\mathcal{M}_k(t))$ and let $v_k = (f^k_s)^{-1}(\mathcal{M}_1(t))$. The family $\Phi(f^k_s)_{s \in S_k}$ is the family that updates $(f^k_s)_{s \in S_k}$ by replacing f^k_s by g^k_s, where $g^k_s(v_k) = v_1$, $g^k_s(\mathcal{M}_k(t)) = \mathcal{M}_1(t)$, and for every other v, $g^k_s(v) = f^k_s(v)$. Hence, $\Psi(\Phi(f^k_s)_{s \in S_k}) < \Psi(f^k_s)_{s \in S_k}$. Also note that $\Phi(f^k_s)_{s \in S_k}$ satisfies the same properties (from leading-theory compatibility) as $(f^k_s)_{s \in S_k}$ does.
   We keep applying Φ to the family $(f^k_s)_{s \in S_k}$ from the leading-theory compatibility, until we obtain a family $(\phi^s_k)_{s \in S_k}$ that also satisfies the additional property $\phi^s_k(\mathcal{M}_k(t)) = \mathcal{M}_1(t)$ for all shared terms $t \in \mathcal{V}^s_{\mathsf{sh}}(H)$ of sort s.

The rest of the proof is as in [1, Thm. 4].

Given a predicate-sharing union of theories $\mathcal{T}_1, \ldots, \mathcal{T}_n$, a collection of theory modules $\mathcal{I}_1, \ldots, \mathcal{I}_n$ for $\mathcal{T}_1, \ldots, \mathcal{T}_n$ is *complete*, if module $\mathcal{I}_1$ is complete for the leading theory, and modules $\mathcal{I}_k$'s, $2 \le k \le n$, are leading-theory-complete. With this assumption, one shows the generalized version of [1, Thm. 3], where an assignment H *is in* $\mathcal{B}$ if $(t \leftarrow \mathfrak{c}) \in H$ implies $t \in \mathcal{B}$.

Theorem 3. *In a predicate-sharing union of theories equipped with a complete collection of theory modules and a stable global basis $\mathcal{B}$, for all input assignments H in $\mathcal{B}$, whenever a CDSAT derivation from H halts in a state Γ other than* unsat*, Γ is model-describing.*

Proof: The proof is the same as that of [1, Thm. 3] because the CDSAT transition system is unchanged.

Theorems 2 and 3 directly entail the completeness of CDSAT for predicate-sharing unions, which subsumes the completeness property for disjoint unions [1, Thm. 5].

Theorem 4 (Completeness). *In a predicate-sharing union of theories equipped with a complete collection of theory modules and a stable global basis $\mathcal{B}$, for all input assignments H in $\mathcal{B}$, whenever a CDSAT derivation from H halts in a state Γ other than* unsat*, there exists a $\mathcal{T}^+_\infty[\mathsf{fv}(\Gamma)]$-model $\mathcal{M}$ such that $\mathcal{M} \models^G \Gamma$ and hence $\mathcal{M} \models^G H$ (as $H \subseteq \Gamma$).*

7. Discussion

The *equality-sharing method* (aka *Nelson-Oppen scheme*) yields a decision procedure for the satisfiability of a conjunction of literals in a union of theories, by combining the respective decision procedures for the component theories [7]. The integration of the equality-sharing method in the CDCL($\mathcal{T}$) transition system[1] yields a decision procedure for the satisfiability of a quantifier-free formula in a union of theories [11, 12]. The theories are required to be *disjoint* and *stably-infinite*, where a theory $\mathcal{T}$ is *stably-infinite* if every $\mathcal{T}$-satisfiable formula has a $\mathcal{T}$-model with countably infinite domain.

Polite theory combination (e.g., [13, 14, 15, 16]) extends the equality-sharing method so as to combine a non-stably-infinite theory with a *polite* theory. Politeness is a stronger cardinality requirement than stable infiniteness, and in general the theories are still required to be disjoint. However, polite theory combination was generalized [5] to the nondisjoint case represented by the theories of *absolutely free data structures with bridging functions*, which are polite [5]. The theory of *arrays with extensionality* is polite [13], but arrays are not an absolutely free data data structure with constructors and selectors.

Another approach to the problem of reasoning in a union of theories consists of applying a superposition-based inference system to the axioms and the target formula. If superposition is a decision procedure for each of the component theories, it is a decision procedure for their union, provided the theories are *disjoint* and *variable-inactive* [17]. The latter property implies stable-infiniteness. The theory of *arrays with extensionality* is decidable by superposition [18] and is variable-inactive [17]. This approach was extended to unions of theories that share a theory of counter arithmetic [19, 20]. However, there are theories, such as arithmetic or bitvectors, that do not lend themselves to reasoning by generic theorem proving.

Therefore, the CDCL($\Gamma + \mathcal{T}$) transition system[2] integrates a superposition-based inference system (the Γ parameter) in the CDCL($\mathcal{T}$) transition system, with the *model-based theory combination* version [22] of equality sharing. The resulting method can reason in a union of theories comprising both built-in and axiomatized theories, provided the theories are *disjoint* and either *stably-infinite* (for the built-in theories) or *variable-inactive* (for the axiomatized theories). CDCL($\Gamma+\mathcal{T}$) is a semidecision procedure in general, but it may yield decision procedures by employing *speculative axioms* [21]. A survey of the methods mentioned up to here appeared [23].

MCSAT [24] offered for the first time a transition system that composes the transitions

[1]The original name is DPLL($\mathcal{T}$) [8], but the recent literature uses CDCL($\mathcal{T}$), since the DPLL (Davis-Putnam-Logemann-Loveland) [9] and CDCL (Conflict-Driven Clause Learning) [10] procedures have been recognized as distinct.

[2]Here too the original name is DPLL($\Gamma+\mathcal{T}$) [21] and the renaming follows that of DPLL($\mathcal{T}$).

for CDCL with those for another conflict-driven decision procedure. CDSAT [1, 2] generalized MCSAT to generic unions of *disjoint* theories, accommodating both conflict-driven and non-conflict-driven decision procedures. Stable infiniteness is not required, because an agreement on the cardinalities of shared sorts is reached via a *leading theory*. Equality sharing is covered as a special case with a leading theory that assigns countably infinite cardinality to the interpretation of all sorts other than prop.

Here we presented an extension of CDSAT to the *nondisjoint* case, motivated by the problem of enriching the *theory of arrays with extensionality* with a notion of *length* of an array. Previous approaches considered this problem in the case where the indices of an array form an interval in a linearly ordered set.

The theory of arrays in [25] assumes that the indices are integers, and defines the *bounded equality* of two arrays as having equal elements at all indices between a lower bound and an upper bound. The resulting axiomatization belongs to the *array property fragment*, whose decision procedure reduces the problem to reasoning about uninterpreted functions, LIA, and the theory of the array elements [25].

The *theory of arrays with MaxDiff* [6] is parametrized with respect to a theory of indices that is required to extend the theory of *linear orderings* with an element 0. LIA, LRA, and the theory IDL of *integer difference logic* (i.e., the theory with 0, successor, predecessor, and the ordering), satisfy this requirement. The theory of arrays with MaxDiff features a symbol $\perp$ for the undefined element and a symbol ϵ for the array that has element $\perp$ at all indices. The axioms impose that an array has element $\perp$ at all indices smaller than 0, and that $\mathsf{diff}(a, b)$ is the largest index where a and b differ and 0 otherwise. Thus, the length of an array a is given by $\mathsf{diff}(a, \epsilon)$. The theory of arrays with MaxDiff and the theory of indices need to share the symbols for the element 0 and for the ordering.

Our approach is more general. The theory ArrL of *arrays with extensionality and abstract length* features an abstract *admissibility predicate* Adm for array indices. This predicate can be interpreted in such a way that the indices of an array form an interval in a linearly ordered set, but it does not have to. Thanks to this abstraction, ArrL only needs to share the symbol Adm with another theory and with the leading theory (these two theories may coincide, but they do not have to). Thus, it suffices to extend CDSAT to allow the theories to share *predicate symbols* other than equality. This requires only minimal changes to the CDSAT framework of definitions and none to the CDSAT transition system itself. We proved that CDSAT is *complete for predicate-sharing unions*.

Directions for future work include developing the abstract approach of this paper to handle in CDSAT a version of theory ArrL enriched with a *concatenation* operator, the theories of *finite maps* and *dynamic arrays* or *vectors* (cf. Sect. 3), and other theories made nondisjoint by bridging functions. An implementation of CDSAT is under way.

Acknowledgements This work was done while the first author was visiting the Computer Science Laboratory of SRI International, whose support is greatly appreciated. This material is based upon work supported in part by NSF grants 1816936, 1817204, and 2016597. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

References

[1] M. P. Bonacina, S. Graham-Lengrand, N. Shankar, Conflict-driven satisfiability for theory combination: transition system and completeness, J. Autom. Reason. 64 (2020) 579–609. doi:10.1007/s10817-018-09510-y.

[2] M. P. Bonacina, S. Graham-Lengrand, N. Shankar, Conflict-driven satisfiability for theory combination: lemmas, modules, and proofs, J. Autom. Reason. 66 (2022) 43–91. doi:10.1007/s10817-021-09606-y.

[3] H. Ganzinger, H. Rueß, N. Shankar, Modularity and refinement in inference systems, Technical Report CSL-SRI-04-02, CSL, SRI International, Menlo Park, CA, USA, 2004.

[4] V. Sofronie-Stokkermans, Locality results for certain extensions of theories with bridging functions, in: R. A. Schmidt (Ed.), Proc. of CADE-22, volume 5663 of *LNAI*, Springer, 2009, pp. 67–83. doi:10.1007/978-3-642-02959-2_5.

[5] P. Chocron, P. Fontaine, C. Ringeissen, Politeness and combination methods for theories with bridging functions, J. Autom. Reason. 64 (2020) 97–134. doi:10.1007/s10817-019-09512-4.

[6] S. Ghilardi, A. Gianola, D. Kapur, Interpolation and amalgamation for arrays with MaxDiff, in: S. Kiefer, C. Tasson (Eds.), Proc. of FoSSaCS-24, volume 12650 of *LNCS*, Springer, 2021, pp. 268–288. doi:10.1007/978-3-030-71995-1_14.

[7] G. Nelson, D. C. Oppen, Simplification by cooperating decision procedures, ACM Trans. Prog. Lang. Syst. 1 (1979) 245–257. doi:10.1145/357073.357079.

[8] R. Nieuwenhuis, A. Oliveras, C. Tinelli, Solving SAT and SAT modulo theories: from an abstract Davis-Putnam-Logemann-Loveland procedure to DPLL(T), J. ACM 53 (2006) 937–977.

[9] M. Davis, G. Logemann, D. Loveland, A machine program for theorem-proving, C. ACM 5 (1962) 394–397. doi:10.1145/368273.368557.

[10] J. Marques Silva, I. Lynce, S. Malik, Conflict-driven clause learning SAT solvers, in: A. Biere, M. Heule, H. Van Maaren, T. Walsh (Eds.), Handbook of Satisfiability, volume 185 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2009, pp. 131–153. doi:10.3233/978-1-58603-929-5-131.

[11] C. W. Barrett, R. Nieuwenhuis, A. Oliveras, C. Tinelli, Splitting on demand in SAT modulo theories, in: M. Hermann, A. Voronkov (Eds.), Proc. of LPAR-13, volume 4246 of *LNAI*, Springer, 2006, pp. 512–526. doi:10.1007/11916277_35.

[12] S. Krstić, A. Goel, Architecting solvers for SAT modulo theories: Nelson-Oppen with DPLL, in: F. Wolter (Ed.), Proc. of FroCoS-6, volume 4720 of *LNAI*, Springer, 2007, pp. 1–27. doi:10.1007/978-3-540-74621-8_1.

[13] S. Ranise, C. Ringeissen, C. G. Zarba, Combining data structures with nonstably infinite theories using many-sorted logic, in: B. Gramlich (Ed.), Proc. of FroCoS-5, volume 3717 of *LNAI*, Springer, 2005, pp. 48–64. doi:10.1007/11559306_3.

[14] D. Jovanović, C. W. Barrett, Polite theories revisited, in: C. G. Fermüller, A. Voronkov (Eds.), Proc. of LPAR-17, volume 6397 of *LNAI*, Springer, 2010, pp. 402–41. doi:10.1007/978-3-642-16242-8_29.

[15] Y. Sheng, Y. Zohar, C. Ringeissen, J. Lange, P. Fontaine, C. W. Barrett, Politeness

for the theory of abstract data types, in: N. Peltier, V. Sofronie-Stokkermans (Eds.), Proc. of IJCAR-10, volume 12166 of *LNAI*, Springer, 2020, pp. 238–255. doi:10.1007/978-3-030-51074-9_14.

[16] Y. Sheng, Y. Zohar, C. Ringeissen, A. Reynolds, C. W. Barrett, C. Tinelli, Politeness and stable infiniteness: stronger together, in: A. Platzer, G. Sutcliffe (Eds.), Proc. of CADE-28, volume 12699 of *LNAI*, Springer, 2021, pp. 148–165. doi:10.1007/978-3-030-79876-5_9.

[17] A. Armando, M. P. Bonacina, S. Ranise, S. Schulz, New results on rewrite-based satisfiability procedures, ACM Trans. Comput. Log. 10 (2009) 129–179. doi:10.1145/1459010.1459014.

[18] A. Armando, S. Ranise, M. Rusinowitch, A rewriting approach to satisfiability procedures, Inf. Comput. 183 (2003) 140–164. doi:10.1016/S0890-5401(03)00020-8.

[19] E. Nicolini, C. Ringeissen, M. Rusinowitch, Combining satisfiability procedures for unions of theories with a shared counting operator, Fundam. Inform. 105 (2010) 163–187. doi:10.3233/FI-2010-362.

[20] C. Ringeissen, V. Senni, Modular termination and combinability for superposition modulo counter arithmetic, in: C. Tinelli, V. Sofronie-Stokkermans (Eds.), Proc. of FroCoS-8, volume 6989 of *LNAI*, Springer, 2011, pp. 211–226. doi:10.1007/978-3-642-24364-6_15.

[21] M. P. Bonacina, C. A. Lynch, L. de Moura, On deciding satisfiability by theorem proving with speculative inferences, J. Autom. Reason. 47 (2011) 161–189. doi:10.1007/s10817-010-9213-y.

[22] L. de Moura, N. Bjørner, Model-based theory combination, in: S. Krstić, A. Oliveras (Eds.), Proc. of SMT-5, volume 198(2) of *ENTCS*, Elsevier, 2008, pp. 37–49. doi:10.1016/j.entcs.2008.04.079.

[23] M. P. Bonacina, P. Fontaine, C. Ringeissen, C. Tinelli, Theory combination: beyond equality sharing, in: C. Lutz, U. Sattler, C. Tinelli, A.-Y. Turhan (Eds.), Description Logic, Theory Combination, and All That: Essays Dedicated to Franz Baader, volume 11560 of *LNCS*, Springer, 2019, pp. 57–89. doi:10.1007/978-3-030-22102-7_3.

[24] L. de Moura, D. Jovanović, A model-constructing satisfiability calculus, in: R. Giacobazzi, J. Berdine, I. Mastroeni (Eds.), Proc. of VMCAI-14, volume 7737 of *LNCS*, Springer, 2013, pp. 1–12. doi:10.1007/978-3-642-35873-9_1.

[25] A. R. Bradley, Z. Manna, H. B. Sipma, What's decidable about arrays?, in: E. A. Emerson, K. S. Namjoshi (Eds.), Proc. of VMCAI-7, volume 3855 of *LNCS*, Springer, 2006, pp. 427–442. doi:10.1007/11609773_28.