Topology, connectedness, and modal logic

R. Kontchakov, I. Pratt-Hartmann, F. Wolter and M. Zakharyaschev

ABSTRACT. This paper presents a survey of topological spatial logics, taking as its point of departure the interpretation of the modal logic $\mathcal{S}4$ due to McKinsey and Tarski. We consider the effect of extending this logic with the means to represent topological *connectedness*, focusing principally on the issue of computational complexity. In particular, we draw attention to the special problems which arise when the logics are interpreted not over *arbitrary* topological spaces, but over (low-dimensional) *Euclidean* spaces.

Keywords: Spatial logic, modal logic, topology, connectedness.

1 Introduction: Spatial Logic and Modal Logic

In their seminal paper *The algebra of topology* [40], McKinsey and Tarski sought to provide 'an algebraic apparatus adequate for the treatment of portions of point-set topology.' In doing so, they created—*en passant*—a topological framework for the semantics of the modal logic $\mathcal{S}4$, exploiting the striking similarity between Gödel's [24] and Orlov's [43] axioms for 'provability' logic and Kuratowski's axioms for topological spaces. In this framework, proposition letters are interpreted as subsets of a topological space, Boolean connectives as set-theoretic operations on these sets, and the modal box as the topological interior operator. As McKinsey and Tarski showed, a modal formula φ is an $\mathcal{S}4$-validity if and only if, in any interpretation over a topological space T, φ denotes the whole of T. In fact, they showed more. Suppose we are interested not in topological spaces *in general*, but rather in some *specific* dense-in-itself, separable metric space—for example $\mathbb{R}^2$ or $\mathbb{R}^3$. For any such space T, a modal formula φ is an $\mathcal{S}4$-validity if and only if, in any interpretation over T, φ denotes the whole of T. In other words: $\mathcal{S}4$ is the logic of any dense-in-itself, separable metric space.

This situation invites generalization. By a *spatial logic*, let us understand any formal language interpreted over some class of geometrical structures, taken in the most general sense. That is: the variables of this language range over collections of figures in the relevant structures; and its non-logical primitives denote properties and operations defined over those figures. What makes a spatial logic a *logic* is that it has a regimented syntax and formal semantics interpreting it; what makes it *spatial* is that the operative notion of logical consequence is made to depend on the specifically geometrical features of the chosen interpretation. Thus, $\mathcal{S}4$ is a spatial logic whose (propositional) variables range over arbitrary subsets of a topological space in some given class.

Spatial logics, thus understood, have a long pedigree, tracing their origins back to the axiomatic tradition in geometry, which reached its zenith in Hilbert's *Grundlagen der Geometrie* [30]. Strikingly, Hilbert's axiomatization is couched in (lightly mathematicized) idiomatic German: notwithstanding its evident rigour, no attempt is made to articulate the implicit logical syntax or operative inference procedure. This feature prompted a further stage of formalization in another of Tarski's most significant papers: *What is elementary geometry?* [59]. Tarski's geometrical axioms are couched in a first-order language whose variables range over points in the standard model of Euclidean space, and whose non-logical predicates represent notions defined in terms of the metric structure of that space. Again, Tarski showed that the consequences of his axioms coincide with the true statements of that model. Of course, the real achievement here was not simply to shoe-horn Hilbert's perfectly good mathematics into the regimented syntax of a formal language, but rather, to ask what happens when that syntax is restricted. For Tarski showed that his elementary geometry is, on the one hand, *decidable*—there is an algorithmic procedure for determining the truth of any of its formulas—and, on the other, sufficiently expressive that it comes close (in a sense which Tarski was able to make precise) to fixing the familiar model of the plane as $\mathbb{R}^2$. On a practical level, Tarski's work has found application in spatial databases (see, e.g., [34]); from a theoretical point of view, we have the beginnings of one of the central themes in computational logic—the trade-off between expressive power and computational complexity. We remark that the precise complexity of Tarski's geometry (or $\mathrm{Th}(\mathbb{R}, +, \times, \leq)$) seems to be still unknown, with the current lower bound being NExpTime [22] and the upper bound ExpSpace [4].

A quite distinct intellectual tradition has also contributed to recent interest in spatial logics, however: Whitehead's theory of *extensive connection*, which appeared in its most complete form in his *Process and Reality* [65]. Whitehead's goal was to develop a purely *region-based* theory of space, whose sole geometrical primitive was the relation he called *connection*, but which is now (to avoid confusion with established terminology) generally referred to as *contact*. Roughly: two regions contact each other just in case they overlap or touch. Whitehead put forward a collection of postulates governing this relation, and gave reconstructions of various geometrical notions in terms of it. A similar—and in many ways more satisfactory—region-based theory of space was proposed at the same time by de Laguna [15]. In both cases the motivation was essentially metaphysical: to provide a spatial ontology whose basic entities are closer to the data of spatial experience than is the standard Cartesian model of space as $\mathbb{R}^3$. Paradoxically, perhaps, the methodology they employed was resolutely empiricist: the proposed system of postulates and definitions was to be evaluated by its conformity to (pre-theoretic spatial intuition and) spatial experience. The ensuing lack of any formal semantics for the languages in question impeded their mathematical development, despite sporadic revivals in the following decades [10, 11, 7].

Interest in region-based, qualitative spatial logics of this kind was rekindled, however, in the early 1990s, within Artificial Intelligence. (See the recent handbook chapters [49, 12] for comprehensive surveys.) The impetus for this development was the conviction that effective reasoning—geometrical or otherwise—depends on selecting a language with the appropriate representational resources: too little expressive power, and it cannot represent the information required; too much, and the reasoner is overwhelmed by the computational complexity of determining entailments within it. Hence the focus on languages whose variables range over spatial regions: while spatial regions *can* be modelled as sets of points, and so quantified over in second-order logic, a first-order logic whose object-level variables range over regions is, from the point of view of expressive economy, a preferable alternative. And once a region-based domain of quantification has been adopted, the focus on qualitative geometrical primitives follows naturally, since so many salient properties and relations involving regions are qualitative in character.

First-order qualitative theories of space, however, are generally undecidable or even non-recursively enumerable [28, 18, 13], a result which extends to some spatial logics based on the two-variable fragment of first-order logic [39]. Hence, attention has shifted to quantifier-free constraint systems such as 9-intersections or $\mathcal{RCC}$-8 [20, 46]. Intriguingly, research on such systems has led to a renewed and systematic investigation of spatial formalisms within the algebraic framework of Tarski [58, 40]. For spatial relations such as 'contact' or 'part of' form a natural subject for relation algebra; see the surveys [6, 61] and references therein. Furthermore, it turns out that many spatial constraint systems designed in AI can be regarded as natural fragments of $\mathcal{S}4$ augmented with the universal modality and known as $\mathcal{S}4_u$. Thus, the modal logic $\mathcal{S}4_u$ finds itself at a crossroads of different traditions and disciplines related to spatial logics. (See [62, 25] for a broader discussion of modal logics of space.)

Few practical problems in spatial reasoning are purely topological in character, of course; and this has recently prompted several extensions of $\mathcal{S}4_u$ with metric primitives (e.g., [35, 68, 54]). Yet, even from a topological point of view, $\mathcal{S}4$ and its near-relation $\mathcal{S}4_u$ can seem frustratingly inexpressive: for example, very few theorems from standard textbooks on topology can be formulated within them! Perhaps the most glaring expressive defect of these languages is their inability to express the property of *connectedness*—a concept of central theoretical and practical importance. To date, only sporadic attempts have been made to interpret $\mathcal{S}4_u$ over connected spaces, or to augment it with a primitive predicate expressing the property connectedness [9, 53, 66, 44].

The present paper has two main aims, therefore. The first is to present a survey of topological spatial logics, taking $\mathcal{S}4$ as its starting-point. The second is to investigate in detail the extension of these logics with the means to represent topological connectedness (and related notions), focusing principally on issues of computational complexity. A surprising discovery here was how the innocuous-looking *connectedness predicate* can increase complexity from NP to PSPACE, EXPTIME and, if component counting is allowed, to NEXPTIME. In particular, we draw the reader's attention to the special difficulties that arise when these logics are interpreted not over *arbitrary* topological spaces, but over (low-dimensional) *Euclidean* spaces. We also point out the sensitivity of such logics to the geometrical entities—polygons, disc-homeomorphs, *etc.*—over which their variables are taken to range.

2 $\mathcal{S}4_u$ over connected topological spaces

We begin by briefly reviewing the topological semantics for $\mathcal{S}4$, due to McKinsey and Tarski [40]. With a view to the ensuing generalizations, we present the language in unfamiliar guise. Specifically, we re-write the proposition letters as individual variables, the propositional connectives $\wedge$ and $\neg$ as the function-symbols $\cap$ and $^-$, respectively, and the modal box $\Box$ as the function-symbol $^\circ$. In this way, familiar modal formulas become *terms*. Formally, let $\mathcal{V} = \{v_i \mid i < \omega\}$ be a set of variables. Then the $\mathcal{S}4$-*terms* are given by

$$\tau \quad ::= \quad v_i \quad | \quad \overline{\tau} \quad | \quad \tau_1 \cap \tau_2 \quad | \quad \tau^\circ.$$

We abbreviate $\overline{(\overline{\tau}^\circ)}$ by τ^-, $\overline{(\overline{\tau_1} \cap \overline{\tau_2})}$ by $\tau_1 \cup \tau_2$, $v_0 \cap \overline{v_0}$ by $\mathbf{0}$, and $\overline{\mathbf{0}}$ by $\mathbf{1}$.

In the sequel, we assume familiarity with basic general topology. If the topological space T is clear from context, and $X \subseteq T$, we denote the complement of X by $\overline{X}$, the topological closure of X by X^-, and the topological interior of X by X°. (The overloading of symbols here is deliberate.) We follow common practice in identifying topological spaces with their carrier sets, taking the topology to be implicit; in addition, we assume that topological spaces are non-empty. In this context, define a *topological frame* to be a pair $(T, \boldsymbol{S})$, where T is a topological space, and $\boldsymbol{S} \subseteq 2^T$ is a non-empty set of its subsets. A *topological model over* $(T, \boldsymbol{S})$ is a triple $\mathfrak{M} = (T, \boldsymbol{S}, \cdot^{\mathfrak{M}})$, where $\cdot^{\mathfrak{M}}$ is a map from $\mathcal{V}$ to $\boldsymbol{S}$. (The modifier 'topological' will generally be omitted in the sequel.) The *extension* $\tau^{\mathfrak{M}}$ of a term τ in a model $\mathfrak{M}$ is defined inductively by the equations:

$$(\overline{\tau})^{\mathfrak{M}} \ = \ \overline{(\tau^{\mathfrak{M}})}, \qquad (\tau_1 \cap \tau_2)^{\mathfrak{M}} \ = \ \tau_1^{\mathfrak{M}} \cap \tau_2^{\mathfrak{M}}, \qquad (\tau^\circ)^{\mathfrak{M}} \ = \ (\tau^{\mathfrak{M}})^\circ.$$

On the above semantics, variables are constrained to range over certain subsets of the underlying space, as specified by $\boldsymbol{S}$. We refer to the elements of $\boldsymbol{S}$ as *regions*. There is no formal requirement for $\boldsymbol{S}$ to be closed under the term-forming operations of our language (in the present case, $^-$, $\cap$ and $^\circ$). In the special case where every subset of T counts as a region—that is, where $\boldsymbol{S} = 2^T$—we identify the topological frame with the underlying topological space, and simply speak of a model $\mathfrak{M} = (T, \cdot^{\mathfrak{M}})$ over T.

Recall that a topological space is called an *Aleksandrov space* if arbitrary (not only finite) intersections of open sets are open. Aleksandrov spaces can be characterized in terms of pairs the form $F = (W, R)$, where $W \neq \emptyset$ and R is a transitive and reflexive relation (i.e., a *quasi-order*) on W. Every such pair—or *Kripke frame*—F induces the interior operator $\cdot^{\circ}_{F}$ on W:

$$X^{\circ}_{F} \;=\; \{x \in X \mid \forall y \in W\ (xRy \to y \in X)\}, \quad \text{for every } X \subseteq W.$$

In other words, the open sets of the topological space $T_F = W$ induced by F are the *upward closed* (or R-*closed*) subsets of W. It is well-known (see, e.g., [8]) that T_F is an Aleksandrov space and, conversely, every Aleksandrov space is induced by a quasi-order. Topological models over Aleksandrov spaces will be called *Aleksandrov models.*

We are now in a position to characterize the 'logic' of the term-language $\mathcal{S}4$ and its complexity.

THEOREM 1 ([40, 36]). (i) *Let τ be an $\mathcal{S}4$-term and T' a dense-in-itself separable space (e.g., $T' = \mathbb{R}^n$, for some $n \geq 1$). The following conditions are equivalent*:

- $\tau^{\mathfrak{M}} = T$ *for every model* $\mathfrak{M} = (T, \cdot^{\mathfrak{M}})$;
- $\tau^{\mathfrak{M}} = T'$ *for every model* $\mathfrak{M} = (T', \cdot^{\mathfrak{M}})$ *over* T';
- $\tau^{\mathfrak{M}} = T$ *for every (finite) Aleksandrov model* $\mathfrak{M} = (T, \cdot^{\mathfrak{M}})$.

(ii) *The problem of deciding, given an $\mathcal{S}4$-term τ, whether $\tau^{\mathfrak{M}} = T$ for all models $\mathfrak{M} = (T, \cdot^{\mathfrak{M}})$, is* PSPACE-*complete.*

With these resources at our disposal, we present our first topological logic, known under the name of $\mathcal{S}4_u$. (We remark that the original formulation of $\mathcal{S}4_u$ in [26], like those of $\mathcal{S}4$ in [43, 37, 24], made no reference to spatial logic or topology.) The language $\mathcal{S}4_u$ is the set of formulas given by:

$$\varphi \quad ::= \quad \tau_1 = \tau_2 \quad | \quad \neg\varphi \quad | \quad \varphi_1 \wedge \varphi_2,$$

where τ_1 and τ_2 range over $\mathcal{S}4$-terms. We employ the Boolean connectives $\vee$ and $\to$ as abbreviations in the standard way, additionally writing $\tau_1 \subseteq \tau_2$ for $\tau_1 \cap \overline{\tau_2} = \mathbf{0}$ and $\tau_1 \neq \tau_2$ for $\neg(\tau_1 = \tau_2)$. A formula will be called an *atom* if it involves no Boolean connectives, a *literal* if it is an atom or a negated atom, and *conjunctive* if it is a conjunction of literals. The *truth-relation* for $\mathcal{S}4_u$ is defined by setting:

$$\mathfrak{M} \models \tau_1 = \tau_2 \qquad \text{iff} \qquad \tau_1^{\mathfrak{M}} = \tau_2^{\mathfrak{M}},$$

and interpreting the Boolean connectives $\neg$ and $\wedge$ in the standard way. If $(T, \boldsymbol{S})$ is a topological frame, then φ is *satisfiable* over $(T, \boldsymbol{S})$ if $\mathfrak{M} \models \varphi$, for some model $\mathfrak{M} = (T, \boldsymbol{S}, \cdot^{\mathfrak{M}})$; if $\mathcal{F}$ is a class of topological frames, then φ is *satisfiable* over $\mathcal{F}$ if it is satisfiable over some $(T, \boldsymbol{S}) \in \mathcal{F}$. Similarly, *mutatis mutandis*, for the dual notion of *validity.* We denote by $Sat(\mathcal{S}4_u, \mathcal{F})$ the set of $\mathcal{S}4_u$-formulas that are satisfiable over $\mathcal{F}$.

In modal terms, $\mathcal{S}4_u$ in effect adds a 'universal modality' [26] to $\mathcal{S}4$, since an atom of the form $\tau = \mathbf{1}$ states that the modal formula corresponding to τ

is true everywhere in the relevant space. It is well known (see, e.g., [1]) that the language of $\mathcal{S}4_u$ defined as above is as expressive as the 'standard' one that allows nested applications of universal modalities. In many cases, addition of the universal modality to a modal logic increases its computational complexity (e.g., the modal logic $\mathcal{K}$ with universal modality is ExpTime-complete). For $\mathcal{S}4$ this turns out to be not the case. Denote by All the class of all topological frames and by Alek the class of all Aleksandrov frames (that is, topological frames based on Aleksandrov spaces).

THEOREM 2 ([53, 2]). *$Sat(\mathcal{S}4_u, \textsc{All}) = Sat(\mathcal{S}4_u, \textsc{Alek})$, and this set is* PSpace-*complete.*

In contrast to Theorem 1, the equality in Theorem 2 cannot be extended to the set $Sat(\mathcal{S}4_u, \mathcal{F})$, where $\mathcal{F}$ is any class of topological frames over $\mathbb{R}^n$. Recall that a topological space T is *connected* just in case it is not the union of two non-empty, disjoint, open sets; a subset $X \subseteq T$ is *connected in* T just in case either it is empty, or the topological space X (with the subspace topology) is connected. If $X \subseteq T$, a maximal connected subset of X is called a *component* of X. Every set X has at least one component, and a set is connected just in case it has at most one component. Denote by Con the class of all frames over connected spaces. The $\mathcal{S}4_u$-formula

$$(v_1 \neq \mathbf{0}) \ \wedge \ (v_2 \neq \mathbf{0}) \ \wedge \ (v_1 \cup v_2 = \mathbf{1}) \ \wedge \ (v_1^- \cap v_2 = \mathbf{0}) \ \wedge \ (v_1 \cap v_2^- = \mathbf{0})$$

is satisfiable in a topological space T iff T is not connected. It follows that $Sat(\mathcal{S}4_u, \textsc{All}) \neq Sat(\mathcal{S}4_u, \textsc{Con})$. The formula above was used in [53] to axiomatize the logic (in the standard language of $\mathcal{S}4_u$) of connected spaces.

Observe that an Aleksandrov space T_F induced by $F = (W, R)$ is connected iff F is *connected* in the sense that between any two points $x, y \in W$ there is a path along the relation $R \cup R^{-1}$, where R^{-1} is the inverse of R. Denote by ConAlek the class of all connected Aleksandrov frames.

THEOREM 3. *$Sat(\mathcal{S}4_u, \textsc{Con}) = Sat(\mathcal{S}4_u, \textsc{ConAlek}) = Sat(\mathcal{S}4_u, T)$, for any connected dense-in-itself separable space T (in particular, for $T = \mathbb{R}^n$, $n \geq 1$). This set is* PSpace-*complete.*

The equations in Theorem 3 were proved in [53], and the complexity result follows from Theorem 5 below. Although of the same complexity as $Sat(\mathcal{S}4_u, \textsc{All})$, $Sat(\mathcal{S}4_u, \textsc{Con})$ requires a subtler treatment. We illustrate this by the following example.

EXAMPLE 4. Denote by $\textsc{Alek}^{\leq 1}$ the class of Aleksandrov frames induced by partial orders $F = (W, R)$ of depth 1, as in Fig. 1; i.e., R is the reflexive closure of a subset of $W_1 \times W_0$, where W_i is the set of points of depth i, $i = 0, 1$. Such partial orders will be called *quasi-saws.*

$Sat(\mathcal{S}4_u, \textsc{Alek}^{\leq 1})$ is NP-complete because formulas in this set enjoy the *polysize model property.* More precisely, it is easy to see that every formula $\varphi \in Sat(\mathcal{S}4_u, \textsc{Alek}^{\leq 1})$ is satisfied in a disjoint union of n many *m-brooms*, i.e., partial orders of the form $(\{x\} \cup W_0, R)$, where $|W_0| = m$ and R is the reflexive closure of $\{x\} \times W_0$, and both m and n are bounded by a

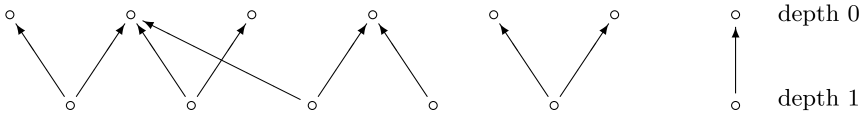

Figure 1. Quasi-saw.

linear function in $|\varphi|$. By contrast, formulas in $Sat(\mathcal{S}4_u, \textsc{ConAlek}^{\leq 1})$ may require *exponential* satisfying models, and this set is PSPACE-complete [66]. We show how one can construct such formulas. Using n variables $v_1, \dots, v_n$ one can represent (in binary) all natural numbers $< 2^n$. Now we can say that $\overline{v_n} \cap \dots \cap \overline{v_1}$ (i.e., 0) and $v_n \cap \dots \cap v_1$ (i.e., $2^n - 1$) are non-empty:

$$\overline{v_n} \cap \dots \cap \overline{v_1} \neq \mathbf{0}, \qquad v_n \cap \dots \cap v_1 \neq \mathbf{0},$$

and that the closure of the set representing a number m, $0 \leq m < 2^n - 1$, can only share points with the set representing $m+1$:

$$\begin{array}{lll} (v_j \cap \overline{v_k})^- \subseteq v_j, & (\overline{v_j} \cap \overline{v_k})^- \subseteq \overline{v_j}, & \text{for all } n \geq j > k \geq 1, \\ (\overline{v_k} \cap v_{k-1} \cap \dots \cap v_1)^- \subseteq (v_k \cap \overline{v_i}) \cup (\overline{v_k} \cap v_i), & & \text{for all } n \geq k > i \geq 1, \end{array}$$

and that $2^n - 1$ is a closed set:

$$(v_n \cap \dots \cap v_1)^- \subseteq v_n \cap \dots \cap v_1.$$

As the space is connected, there is a path between 0 and $2^n - 1$, and this path must contain all the numbers $< 2^n$ (see Fig. 2). Using this idea, we can simulate any polynomial-space-bounded deterministic Turing machine.

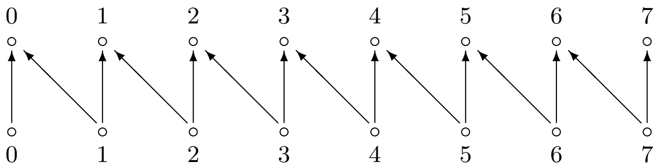

Figure 2. Satisfying the 'counter formulas' for $n = 3$.

Modal definability of separation properties and connectedness in $\mathcal{S}4_u$ and related hybrid logics, as well as their complexity, were studied in [57, 60].

3 Topological logics with connectedness

We now extend $\mathcal{S}4_u$ with an explicit *connectedness predicate* and denote the resulting language by $\mathcal{S}4_u c$. The *$\mathcal{S}4_u c$-formulas* are defined in the same way as the $\mathcal{S}4_u$-formulas, except that we have the additional clause

$$\varphi \quad ::= \quad \dots \quad | \quad c(\tau) \quad | \quad \dots,$$

where τ is an $\mathcal{S}4$-term. Given a topological model $\mathfrak{M} = (T, \boldsymbol{S}, \cdot^{\mathfrak{M}})$, the *truth-relation* for $\mathcal{S}4_u c$ is defined in the same way as for $\mathcal{S}4_u$, except that we have the additional clause

$$\mathfrak{M} \models c(\tau) \qquad \text{iff} \qquad \tau^{\mathfrak{M}} \text{ is connected in } T.$$

For example, most textbooks on general topology prove the following simple facts: (i) the union of two intersecting, connected sets is connected; (ii) any set sandwiched between a connected set and its closure is itself connected. These facts are expressible as $\mathcal{S}4_u c$-validities. That is, the formulas

$$c(v_1) \ \wedge \ c(v_2) \ \wedge \ (v_1 \cap v_2 \neq \mathbf{0}) \ \rightarrow \ c(v_1 \cup v_2), \tag{1}$$

$$c(v_1) \ \wedge \ (v_1 \subseteq v_2) \ \wedge \ (v_2 \subseteq v_1^-) \ \rightarrow \ c(v_2) \tag{2}$$

are valid in All.

Recalling that $\mathcal{S}4$ is a sub-language of $\mathcal{S}4_u c$, $Sat(\mathcal{S}4_u c, \textsc{All})$ is certainly PSpace-hard. But the matching upper bound holds only for the sublanguage $\mathcal{S}4_u c^1$ of $\mathcal{S}4_u c$ in which *at most one* subformula of the form $c(\tau)$ occurs with positive polarity.

THEOREM 5 ([31]). *$Sat(\mathcal{S}4_u c^1, \textsc{All})$ is PSpace-complete.*

Theorem 5 yields the promised result about $\mathcal{S}4_u$ interpreted over connected spaces: an $\mathcal{S}4_u$-formula φ is satisfiable in a connected space iff the $\mathcal{S}4_u c^1$-formula $\varphi \wedge c(\mathbf{1})$ is satisfiable in some topological space.

From a complexity-theoretic viewpoint, the main difference between the languages $\mathcal{S}4_u c$ and $\mathcal{S}4_u c^1$ is that when constructing a model for an $\mathcal{S}4_u c^1$-formula (using, say, a tableau-based technique) there is only one positive statement of the form $c(\tau)$ saying that points in τ have to be connected. We have seen above that connecting two points may require an exponentially long path. Nevertheless, 'connectivity' can be checked using a PSpace-algorithm because it is not necessary to keep in memory all the points on the path. However, if two statements $c(\tau_1)$ and $c(\tau_2)$ have to be satisfied, then, while connecting two τ_1-points using a path, one has to check whether the τ_2-points on that path can be connected by a path, which, in turn, can contain another τ_1-point, and so on. And this situation can indeed happen if we have two positive occurrences of sub-formulas like $c(\tau_1)$ and $c(\tau_2)$.

THEOREM 6 ([31]). *$Sat(\mathcal{S}4_u c, \textsc{All}) = Sat(\mathcal{S}4_u c, \textsc{Alek})$; and this set is ExpTime-complete.*

In fact, the lower bound holds already for $\textsc{Alek}^{\leq 1}$. It can be proved by reduction of polynomial-space-bounded *alternating* Turing machines or satisfiability in logics like modal $\mathcal{K}$ with the universal modality. In either case, the crucial point in the proof is simulating large binary (*non-transitive*) trees. We have already seen how connectedness can help us generate quasi-saws representing an exponential counter. But now we also need *branching.* One idea of simulating both is as follows. We start by representing the root of the tree as a point v_0 (see Fig. 3), which is forced to be connected to an auxiliary point z by means of some $c(\tau_0)$. On the connecting path from v_0 to z we represent the two successors of the root by v_1 and v_2, which are forced to be connected in their turn to z by some other $c(\tau_1)$. On each of the two connecting paths, we again take two points representing the successors of v_1 and v_2, respectively. We treat these four points in the same way as v_0, reusing $c(\tau_0)$, and proceed in this way *ad infinitum* alternating between τ_0 and τ_1 when forcing the paths which generate the required successors. Of course, in addition, certain information has to be passed from a node to its two successors (say, if $\Diamond\psi$ holds in the node, then ψ holds in one of its successors). Such information can be propagated along connected regions. Note now that all points are connected to z. Thus, to distinguish between the information we have to pass from distinct nodes of even (respectively, odd) level to their successors, we have to use *two* connectedness formulas of the form $c(f_i \cup a)$, $i = 0, 1$, in such a way that the f_i points form initial segments of the paths to z and a contains z. The f_i-segments are then used locally to pass information from a node to its successors without conflict. Note also that the points representing nodes of the tree belong to both f_0 and f_1 (except the root) and we have to separate them with auxiliary points in order to ensure proper tree structure (otherwise the 'tree' would collapse into a single node). For details the reader is referred to [31].

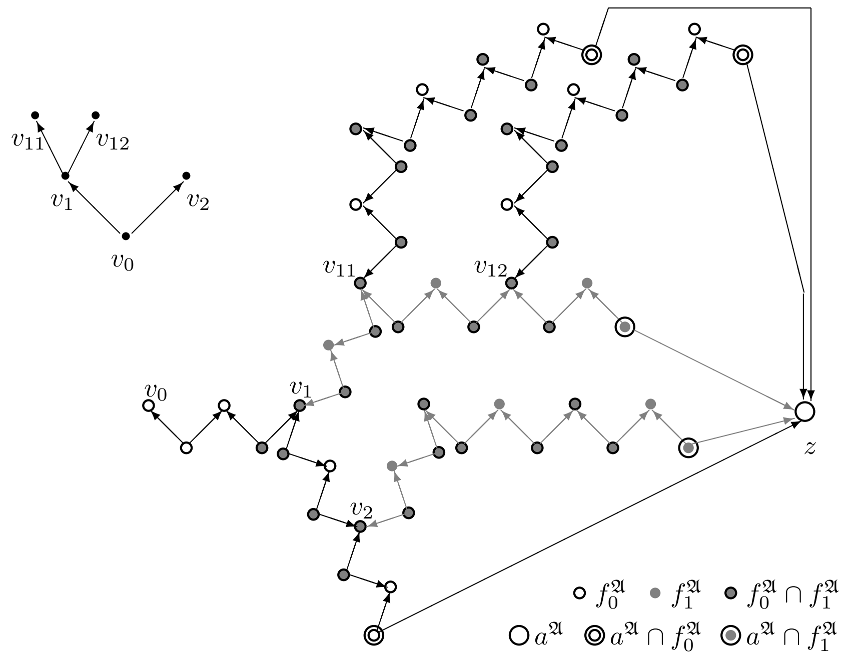

Figure 3. Encoding binary trees in $\mathcal{S}4_u c$.

To establish the upper bound for $Sat(\mathcal{S}4_u c, \textsc{Alek})$, we adapt the type-elimination technique first used to prove the ExpTime upper bound for $\mathcal{PDL}$; see, e.g., [29]. Let a formula φ of $\mathcal{S}4_u c$ be given; but suppose for the moment that φ contains no occurrences of c. One can test the satisfiability of φ over Alek by first computing the set of all φ-types (alias Hintikka sets), where a φ-type is a Boolean-saturated set of subterms of φ. Then one recursively eliminates all those φ-types t for which there is no witness type $t' \ni \tau$ for some $\tau^- \in t$. It can be shown that φ is satisfiable iff this elimination process terminates with a set of types corresponding to a model satisfying φ. Now suppose that φ involves some occurrences of c. Guess a set Ξ of subformulas of φ of the form $c(\tau)$ (those that one assumes to be true), and for each such $c(\tau)$, guess a φ-type t_τ containing τ. The elimination process described above can now be executed as before, except that one also eliminates those types t that contain a τ with $c(\tau) \in \Xi$ which 'cannot be connected' to t_τ (i.e., in the region corresponding to τ, one cannot find an $R \cup R^{-1}$-path from t to t_τ).

One can increase the expressive power of the connectedness predicate $c(\tau)$ by introducing the 'counting' predicates $c^{\leq k}(\tau)$ which state that τ has at most k connected components. We denote the language with such predicates by $\mathcal{S}4_u cc$. The *$\mathcal{S}4_u cc$-formulas* are defined in the same way as the $\mathcal{S}4_u$-formulas, except that we have the additional clause

$$\varphi \quad ::= \quad \dots \quad | \quad c^{\leq k}(\tau) \quad | \quad \dots,$$

where τ is an $\mathcal{S}4$-term and k a positive integer. Given a topological model $\mathfrak{M} = (T, \boldsymbol{S}, \cdot^{\mathfrak{M}})$, the *truth-relation* for $\mathcal{S}4_u cc$ is defined in the same way as for $\mathcal{S}4_u$, except that we have the additional clause:

$$\mathfrak{M} \models c^{\leq k}(\tau) \qquad \text{iff} \qquad \tau^{\mathfrak{M}} \text{ has at most } k \text{ components in } T.$$

We write $\neg c^{\leq k}(\tau)$ as $c^{\geq k+1}(\tau)$ and abbreviate $c^{\leq 1}(\tau)$ by $c(\tau)$. Thus, we may regard $\mathcal{S}4_u c$ as a sub-language of $\mathcal{S}4_u cc$. The numerical superscripts k in $c^{\leq k}$ are assumed to be coded in *binary* and so to have size $\lfloor \log k \rfloor + 1$.

The language $\mathcal{S}4_u cc$ is not essentially more expressive than $\mathcal{S}4_u c$. In particular, the $\mathcal{S}4_u cc$-literal $c^{\leq k}(\tau)$ is true in a model $\mathfrak{M}$ iff the $\mathcal{S}4_u c$-formula

$$\big(\tau = \bigcup_{1 \leq i \leq k} v_i\big) \quad \wedge \quad \bigwedge_{1 \leq i \leq k} c(v_i)$$

is true in some model $\mathfrak{M}'$ differing from $\mathfrak{M}$ at most in the assignments to the variables $v_1, \dots, v_k$. Thus, the $\mathcal{S}4_u cc$-literal $c^{\leq k}(\tau)$ may be 'encoded' by this $\mathcal{S}4_u c$-formula. Similarly, the $\mathcal{S}4_u cc$-literal $c^{\geq k}(\tau)$ may be likewise encoded by the $\mathcal{S}4_u c$-formula

$$\big(\tau = \bigcup_{1 \leq i \leq k} v_i\big) \quad \wedge \quad \bigwedge_{1 \leq i \leq k} \big(v_i \neq \mathbf{0}\big) \quad \wedge \quad \bigwedge_{1 \leq i < j \leq k} \big(\tau \cap v_i^- \cap v_j^- = \mathbf{0}\big).$$

Thus, any $\mathcal{S}4_u cc$-formula can be transformed into an equi-satisfiable $\mathcal{S}4_u c$-formula. However, this transformation involves a combinatorial explosion: the above formulas are exponentially larger than the literals they replace.

THEOREM 7 ([44]). *$Sat(\mathcal{S}4_u cc, \textsc{All}) = Sat(\mathcal{S}4_u cc, \textsc{Alek})$; and this set is* NExpTime*-complete.*

The upper complexity bound follows by establishing an exponential model property of $\mathcal{S}4_u cc$. We remark that this exponential model property holds even though constraints of the form $c^{\geq k}(\tau)$ can be used to succinctly enforce regions with many components. The matching lower bound is proved by reduction of the $2^n \times 2^n$ tiling problem [64]. As we have seen in the proof sketch of Theorem 3, using n variables and a polynomial (in n) number of formulas, one can create a sequence of points in a model representing all natural numbers $< 2^n$, where only points representing m and $m+1$ may be neighbours (see Fig. 2). By using $2n$ variables $v_n, \ldots, v_1$ and $u_n, \ldots, u_1$, one can create all points of the $2^n \times 2^n$ grid (for additional formulas required see [44]) such that a point representing (i,j) has neighbours representing $(i-1,j)$, $(i+1,j)$, $(i,j-1)$ and $(i,j+1)$ and only these pairs. However, the constructed grid may contain 'defects' because the 'counter formulas' are unable to prevent numbers repeating as in the sequence $\ldots, m, m+1, m+2, m+1, m, \ldots$. A key point in the proof is the following. Using the terms

$$\tau_{black} = (v_0 \cap u_0) \cup (\overline{v_0} \cap \overline{u_0}) \quad \text{and} \quad \tau_{white} = (\overline{v_0} \cap u_0) \cup (v_0 \cap \overline{u_0})$$

we can 'colour' the grid in a chessboard manner. But then the constraints $c^{\leq 2^{n-1}}(\tau_{black})$ and $c^{\leq 2^{n-1}}(\tau_{white})$ will ensure that all points representing a pair (i,j) are in the same connected component of either τ_{black} or τ_{white}, and so we can 'cover' all points in this component with the same tile.

Returning to the language $\mathcal{S}4_u c$, it is natural to consider what happens when this language is interpreted over restricted classes of topological spaces. Perhaps the most salient such classes in this context are the singleton classes $\{\mathbb{R}^n\}$, for various n, as well as their union.

It is very easy to see that $Sat(\mathcal{S}4_u c, \mathbb{R})$ and $Sat(\mathcal{S}4_u c, \mathbb{R}^2)$ are both different from $Sat(\mathcal{S}4_u c, \textsc{Con})$ (and from each other). For instance, the formula

$$\bigwedge_{1 \leq i \leq 3} c(v_i) \quad \wedge \bigwedge_{1 \leq i < j \leq 3} \big(v_i \cap v_j \neq \mathbf{0}\big) \quad \wedge \quad \big(v_1 \cap v_2 \cap v_3 = \mathbf{0}\big) \tag{3}$$

is evidently satisfiable in $\mathbb{R}^n$ for all $n > 1$, but not satisfiable in $\mathbb{R}$, since connected, non-empty sets in $\mathbb{R}$ are simply intervals. Likewise, it is straightforward to write a formula satisfiable in $\mathbb{R}^n$ for all $n > 2$, but not satisfiable in $\mathbb{R}^2$. Let $v_{i,j}$, $1 \leq i < j \leq 5$, be distinct variables other than v_i, $1 \leq i \leq 5$; and let φ be the formula

$$\bigwedge_{i \in \{j,k\}} \big(v_i \subseteq (v_{j,k})^\circ\big) \quad \wedge \bigwedge_{1 \leq i \leq 5} \big(v_i \neq \mathbf{0}\big) \quad \wedge$$
$$\bigwedge_{\{i,j\} \cap \{k,l\} = \emptyset} \big(v_{i,j} \cap v_{k,l} = \mathbf{0}\big) \quad \wedge \bigwedge_{1 \leq i < j \leq 5} c((v_{i,j})^\circ). \tag{4}$$

Then φ is not satisfiable in $\mathbb{R}^2$, since otherwise, one could easily embed the non-planar graph K_5 in the plane. On the other hand, it is straightforward to satisfy φ in, say, $\mathbb{R}^3$. Slightly less obviously, it turns out that $Sat(\mathcal{S}4_u c, \{\mathbb{R}^n \mid n > 0\})$ is different from $Sat(\mathcal{S}4_u c, \textsc{Con})$.

FACT 8 ([41], p. 137). If D_1 and D_2 are non-intersecting closed sets in $\mathbb{R}^n$, and points x and y are connected in $\overline{D}_1$ and also in $\overline{D}_2$, then x and y are connected in $\overline{D}_1 \cap \overline{D}_2$.

Now consider the following formula:

$$(v_1 \cap v_2 = \mathbf{0}) \ \wedge \ \bigwedge_{i=1,2} \big((v_i^- \subseteq v_i) \ \wedge \ c(\overline{v_i})\big) \ \wedge \ \neg c(\overline{v_1} \cap \overline{v_2}). \tag{5}$$

This formula is is not satisfied over any space $\mathbb{R}^n$, by Fact 8. However, it is satisfiable in many natural, connected topological spaces: e.g., let T be a torus, and let v_1 and v_2 be interpreted as rings in T, arranged as in Fig. 4.

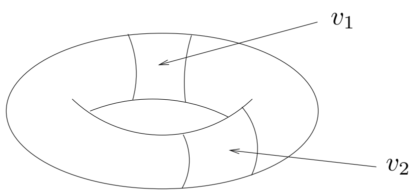

Figure 4. Two non-intersecting connected, closed sets v_1 and v_2 on a torus: note that $\overline{v_1}$ and $\overline{v_2}$ are connected, but $\overline{v_1} \cap \overline{v_2}$ is not.

Using an encoding of the topological interior and closure operators over $\mathbb{R}$ in standard temporal logic with 'since' and 'until' over $\mathbb{R}$ and Reynolds' [50] PSPACE-completeness result for this logic, one can prove the following:

THEOREM 9 ([31]). *$Sat(\mathcal{S}4_u c, \mathbb{R})$ and $Sat(\mathcal{S}4_u cc, \mathbb{R})$ are* PSPACE-*complete.*

Over higher-dimensional Euclidean spaces these languages turn out to be computationally more complex, because the proofs of the lower bounds in Theorems 6 and 7 can be restricted to such spaces:

THEOREM 10 ([31]). (i) *The sets $Sat(\mathcal{S}4_u c, \textsc{Con})$, $Sat(\mathcal{S}4_u c, \{\mathbb{R}^n \mid n > 2\})$ and $Sat(\mathcal{S}4_u c, \mathbb{R}^2)$ are all distinct; $Sat(\mathcal{S}4_u c, \textsc{Con})$ is* EXPTIME-*complete and the other two sets are* EXPTIME-*hard.*

(ii) *$Sat(\mathcal{S}4_u cc, \textsc{Con})$, $Sat(\mathcal{S}4_u cc, \{\mathbb{R}^n \mid n > 2\})$ and $Sat(\mathcal{S}4_u cc, \mathbb{R}^2)$ are all distinct; $Sat(\mathcal{S}4_u cc, \textsc{Con})$ is* NEXPTIME-*complete and the other two sets are* NEXPTIME-*hard.*

We conclude this section by mentioning two relevant research problems. First, it would be interesting to consider other modal logics with connectedness predicate (i.e. operator) $c(\tau)$, which is true in a Kripke model if any two distinct τ-points in the model are connected by a path of τ-points. For example, as in Theorem 6 one can show that basic modal logic $\mathcal{K}$ extended with the universal modality and connectedness predicate is EXPTIME-complete. The connectedness predicate can actually be expressed in the extension of $\mathcal{PDL}$ with converse programs and nominals, which is also EXPTIME-complete [14]. Another direction is to investigate the axiomatization problem for logics with connectedness predicate (see, e.g. [61]).

4 Regularized topological languages

So far, we have considered only frames $(T, \boldsymbol{S})$ in which $\boldsymbol{S}$ is the whole of 2^T—that is to say, frames in which every subset of the space counts as a region. When reasoning about spatial regions in practical situations, we may wish our variables to quantify only over 'sensible' subsets of space, corresponding to the regions potentially occupied by physical objects. In the same spirit, we may further wish to disregard differences between subsets of the space differing only with respect to boundary points. The following technical apparatus provides a convenient way to do this.

Let T be a topological space. A subset $X \subseteq T$ is called *regular closed* if $X = X^{\circ -}$. We denote the set of regular closed subsets of T by $\boldsymbol{RC}(T)$. It is easy to show that the regular closed subsets of T are in fact exactly those sets of the form $X^{\circ -}$, where X ranges over all subsets of T. The following fact is well-known (see, for example, [32], pp. 25–27).

FACT 11. Let T be a topological space. Then $\boldsymbol{RC}(T)$ is a Boolean algebra with top and bottom elements given by T and $\emptyset$, Boolean operations $\cdot, -$ given by $X \cdot Y = (X \cap Y)^{\circ -}$ and $-X = (\overline{X})^{-}$, and Boolean order $\leq$ given by the relation $\subseteq$.

In the context of $\mathbb{R}^2$, the regular closed sets are the closed sets with no 'filaments' or 'isolated points' (Fig. 5). Thus, we are led to consider logics interpreted over frames of the form $(T, \boldsymbol{RC}(T))$. We mention in passing

a) b)

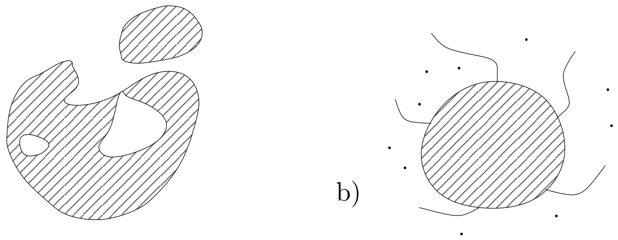

Figure 5. Shaded regions showing: a) a regular-closed subset of $\mathbb{R}^2$, and b) a (closed but) not regular-closed subset of $\mathbb{R}^2$.

that a set X is *regular open* if $X = X^{-\circ}$. The regular open subsets of T also form a Boolean algebra, $\boldsymbol{RO}(T)$, defined analogously to $\boldsymbol{RC}(T)$; in fact the map $X \mapsto X^-$ is a Boolean algebra isomorphism from $\boldsymbol{RO}(T)$ to $\boldsymbol{RC}(T)$. In this section, we speak only of regular closed sets; however, the same material can be presented (with minor changes) using regular open sets.

As $\boldsymbol{RC}(T)$ is not closed under complementation and intersection, it is not very natural to interpret $\mathcal{S}4$-terms over regular frames. This prompts us to define the term-language $\mathcal{B}$ as follows. Let $\mathcal{R} = \{r_i \mid i < \omega\}$ be a set of variables. The set of $\mathcal{B}$-*terms* is defined by:

$$\tau \quad ::= \quad r_i \quad | \quad -\tau \quad | \quad \tau_1 \cdot \tau_2.$$

We interpret $\mathcal{B}$-terms by taking variables to range over regular closed sets of topological spaces. More precisely, we confine attention to topological frames $(T, \boldsymbol{R})$, where $\boldsymbol{R}$ is a Boolean sub-algebra of $\boldsymbol{RC}(T)$. We call any such frame *regular*, and denote the class of regular frames by REG. The class of regular frames based on connected topological spaces will be denoted by CONREG. We may then inductively define the *extension* $\tau^{\mathfrak{M}}$ of a term τ in a model $\mathfrak{M}$ over a regular frame by the equations:

$$(-\tau)^{\mathfrak{M}} = -\tau^{\mathfrak{M}}, \qquad\qquad (\tau_1 \cdot \tau_2)^{\mathfrak{M}} = \tau_1^{\mathfrak{M}} \cdot \tau_2^{\mathfrak{M}}.$$

Again, we have overloaded the symbols $\cdot$ and $-$: on the right-hand sides of these equations, they denote the Boolean algebra operations defined in Fact 11. We abbreviate $-((-\tau_1) \cdot (-\tau_2))$ by $\tau_1 + \tau_2$, $r_0 \cdot -r_0$ by $\mathbf{0}$, and $-\mathbf{0}$ by $\mathbf{1}$. The language of $\mathcal{B}$-terms can form the basis of topological logics just as well as $\mathcal{S}4$. In particular, we can introduce the languages $\mathcal{B}$, $\mathcal{B}c$ and $\mathcal{B}cc$ by defining their formulas as, respectively,

$$\begin{array}{lllllllll}
\varphi & ::= & \tau_1 = \tau_2 & \mid & \neg\varphi & \mid & \varphi_1 \wedge \varphi_2; & & \\
\varphi & ::= & \tau_1 = \tau_2 & \mid & \neg\varphi & \mid & \varphi_1 \wedge \varphi_2 & \mid & c(\tau_1); \\
\varphi & ::= & \tau_1 = \tau_2 & \mid & \neg\varphi & \mid & \varphi_1 \wedge \varphi_2 & \mid & c^{\leq k}(\tau_1),
\end{array}$$

where τ_1 and τ_2 range over $\mathcal{B}$-terms and k is a positive integer. The semantics of these predicates is exactly as for the languages $\mathcal{S}4_u$, $\mathcal{S}4_u c$ and $\mathcal{S}4_u cc$. We abbreviate $\tau_1 \cdot (-\tau_2) = \mathbf{0}$ by $\tau_1 \leq \tau_2$ (preferring this to $\tau_1 \subseteq \tau_2$).

Observe that the languages from the $\mathcal{B}$-family can be viewed as a *syntactic* restriction of the respective languages from the $\mathcal{S}4$-family, as follows. Let τ be a $\mathcal{B}$-term. Define the $\mathcal{S}4$-term $h(\tau)$ recursively by:

$$h(r_i) = v_i^{\circ\,-}, \quad h(\tau_1 \cdot \tau_2) = (h(\tau_1) \cap h(\tau_2))^{\circ\,-}, \quad h(-\tau_1) = \left(\overline{h(\tau_1)}\right)^{-};$$

and if φ is a $\mathcal{B}$-formula ($\mathcal{B}c$- or $\mathcal{B}cc$-formula), define $h(\varphi)$ to be the result of replacing each (maximal) $\mathcal{B}$-term τ occurring in φ by the corresponding $\mathcal{B}$-term $h(\tau)$. Thus, $h(\varphi)$ is an $\mathcal{S}4_u$-formula ($\mathcal{S}4_u c$- or $\mathcal{S}4_u cc$-formula, respectively). It is easy to check that a $\mathcal{B}cc$-formula φ is satisfiable over a frame $(T, \boldsymbol{RC}(T))$ iff $h(\varphi)$ is satisfiable over T.

The minimal logic $\mathcal{B}$ is as expressive as the modal logic $\mathcal{S}5$, with $\tau = \mathbf{1}$ playing the role of the $\mathcal{S}5$-box. Topologically, every satisfiable $\mathcal{B}$-formula φ is satisfied in a discrete topological space (= Aleksandrov frame of depth 0) with $\leq |\varphi|$ points. Hence, $Sat(\mathcal{B}, \text{REG})$ is NP-complete. It also follows that $\mathcal{B}$ does not distinguish between REG, CONREG, $\boldsymbol{RC}(\mathbb{R}^n)$, $n \geq 1$.

The language $\mathcal{B}c$ is less trivial. For example, the smallest Aleksandrov model satisfying the formula $\neg c(r_1) \wedge c(\mathbf{1})$ is shown in Fig. 6. Another important example is the formula $c(\tau_1) \wedge c(\tau_2) \wedge \neg c(\tau_1 + \tau_2)$, which says that both τ_1 and τ_2 are connected and do not intersect, i.e., $\tau_1^{\mathfrak{M}} \cap \tau_2^{\mathfrak{M}} = \emptyset$ in every model satisfying it. In fact, using formulas of this kind one can simulate binary trees (as in the proof of Theorem 6) and obtain the following rather surprising result:

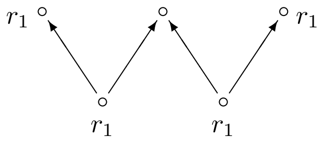

Figure 6. An Aleksandrov model for $\neg c(r_1) \wedge c(\mathbf{1})$.

THEOREM 12 ([31]). (i) *Sat*$(\mathcal{B}c, \textsc{Reg}) \neq$ *Sat*$(\mathcal{B}c, \textsc{ConReg})$*, with both sets being* ExpTime*-complete.*

(ii) *Sat*$(\mathcal{B}cc, \textsc{Reg}) \neq$ *Sat*$(\mathcal{B}cc, \textsc{ConReg})$; *both are* NExpTime*-complete.*

It is also of interest to note that $\mathcal{B}c$ (and $\mathcal{B}cc$) can distinguish between $\boldsymbol{RC}(\mathbb{R})$, $\boldsymbol{RC}(\mathbb{R}^2)$ and $\boldsymbol{RC}(\mathbb{R}^n)$, for $n > 2$.

Consider now the language $\mathcal{B}c^\circ$ defined in the same way as $\mathcal{B}c$, except that the predicate c is replaced by the predicate c° with the interpretation:

$$\mathfrak{M} \models c^\circ(\tau) \quad \text{iff} \quad \text{the interior of } \tau^{\mathfrak{M}} \text{ is connected.}$$

It is a simple exercise in general topology to show that the analogues, in $\mathcal{B}c$ and $\mathcal{B}c^\circ$, of the formula (1), namely,

$$\begin{array}{c} c(r_1) \;\wedge\; c(r_2) \;\wedge\; (r_1 \cdot r_2 \neq \mathbf{0}) \;\;\rightarrow\;\; c(r_1 + r_2) \\ c^\circ(r_1) \;\wedge\; c^\circ(r_2) \;\wedge\; (r_1 \cdot r_2 \neq \mathbf{0}) \;\;\rightarrow\;\; c^\circ(r_1 + r_2) \end{array}$$

are both valid over Reg. The language $\mathcal{B}c^\circ$ is a natural choice for describing arrangements in the Euclidean plane, particularly when variables are taken to range only over well-behaved regions. To understand the issues that arise in this context, consider the $\mathcal{B}c^\circ$-formula

$$\bigwedge_{1 \leq i \leq 3} c^\circ(r_i) \;\;\wedge \bigwedge_{1 \leq i < j \leq 3} (r_i \cdot r_j = \mathbf{0}) \;\;\wedge\;\; c^\circ(\sum_{1 \leq i \leq 3} r_i) \;\;\wedge \bigwedge_{i=2,3} \neg c^\circ(r_1 + r_i). \quad (6)$$

Formula (6) 'says' that r_1, r_2 and r_3 are interior-connected, pairwise disjoint, regular closed sets having an interior-connected sum, such that the first forms an interior-connected sum with neither of the other two. This formula is satisfiable in $\boldsymbol{RC}(\mathbb{R}^2)$. For let $\mathfrak{M}$ be a model over $\boldsymbol{RC}(\mathbb{R}^2)$ in which

$$\begin{aligned} r_1^{\mathfrak{M}} &= \{(x,y) \in \mathbb{R}^2 \mid -1 \leq x \leq 0,\; -1-x \leq y \leq 1+x\}, \\ r_2^{\mathfrak{M}} &= \{(x,y) \in \mathbb{R}^2 \mid 0 < x \leq 1,\; -1-x \leq y \leq \sin(1/x)\} \;\cup \\ &\qquad\qquad\qquad\qquad \{(0,y) \in \mathbb{R}^2 \mid -1 \leq y \leq 1\}, \\ r_3^{\mathfrak{M}} &= \{(x,y) \in \mathbb{R}^2 \mid 0 < x \leq 1,\; \sin(1/x) \leq y \leq 1+x\} \;\cup \\ &\qquad\qquad\qquad\qquad \{(0,y) \in \mathbb{R}^2 \mid -1 \leq y \leq 1\}, \end{aligned}$$

as depicted in Fig. 7. It is easy to check that $(r_1+r_2+r_3)^\circ$ is the interior of the large triangle, and so is certainly connected, but that neither $(r_1+r_2)^\circ$ nor $(r_1+r_3)^\circ$ is connected. However, the regions r_2 and r_3 in Fig. 7 are rather 'wild': they cannot sensibly be used to represent regions of the plane

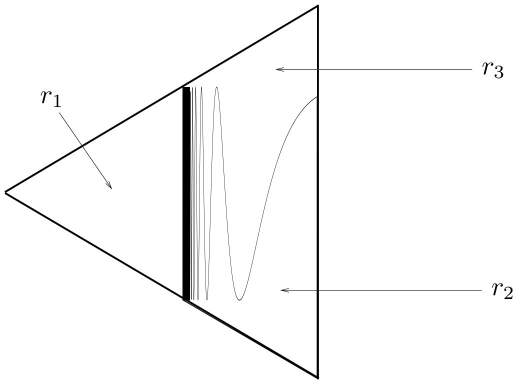

Figure 7. Three elements in $\boldsymbol{RC}(\mathbb{R}^2)$.

occupied (or left unoccupied) by physical objects. The question therefore arises as to whether (6) is satisfiable if only 'tame' regions are allowed.

Any $(n-1)$-dimensional hyper-plane of $\mathbb{R}^n$ cuts $\mathbb{R}^n$ into two closed sets, in the obvious way, which we shall call *half-spaces*. It is easy to see that these half-spaces are regular closed, with each being the complement of the other in the Boolean algebra $\boldsymbol{RC}(\mathbb{R}^n)$. Hence, we can speak about the sums, products and complements of half-spaces in $\boldsymbol{RC}(\mathbb{R}^n)$.

A *basic polytope* in $\mathbb{R}^n$ is the product, in $\boldsymbol{RC}(\mathbb{R}^n)$, of finitely many half-spaces. A *polytope* in $\mathbb{R}^n$ is the sum, in $\boldsymbol{RC}(\mathbb{R}^n)$, of any finite set of basic polytopes. We denote the set of polytopes in $\mathbb{R}^n$ by $\boldsymbol{RCP}(\mathbb{R}^n)$; we call the polytopes in $\boldsymbol{RCP}(\mathbb{R}^2)$ *polygons*. Thus, polytopes (in our sense) may be unbounded, disconnected, and may have disconnected complements. (In alternative parlance, the elements of $\boldsymbol{RCP}(\mathbb{R}^n)$ are the regular closed *semi-linear* sets.) It is obvious that $\boldsymbol{RCP}(\mathbb{R}^n)$ is a Boolean sub-algebra of $\boldsymbol{RC}(\mathbb{R}^n)$. Polytopes are well-behaved in two crucial respects.

Let T be a topological space, and $\boldsymbol{M} \subseteq \boldsymbol{RC}(X)$. We call $\boldsymbol{M}$ *finitely decomposable* if, for all $R \in \boldsymbol{M}$, there exist $R_1, \ldots, R_n \in \boldsymbol{M}$ such that $R = R_1 + \cdots + R_n$. Let T be a topological space, $X \subseteq T$, and p a point on the frontier of X. An *end-cut to p in X* is a Jordan arc g in T such that $g(1) = p$ and $g([0, 1[) \subseteq X$. We say that S *has curve-selection* if, for any point p in the frontier of X, there exists an end-cut in X to p. A set of subsets of T has *has curve-selection* if each of its members does.

LEMMA 13. *$\boldsymbol{RCP}(\mathbb{R}^n)$ is finitely decomposable, and has curve-selection.*

Indeed, basic polytopes are convex, and so trivially have curve selection. But if $R = R_1 + \cdots + R_n$, then $\partial R \subseteq \partial R_1 \cup \cdots \cup \partial R_n$ and $R^\circ \supseteq R_1^\circ \cup \cdots \cup R_n^\circ$, where ∂X denotes the frontier of X. The significance of these properties is that they affect the satisfiability of formula (6).

LEMMA 14 ([45], p. 40). *Let $\boldsymbol{M}$ be a finitely decomposable Boolean subal-*

gebra of $\boldsymbol{RC}(\mathbb{R}^2)$ *forming a closed basis for the usual topology on* $\mathbb{R}^2$*, and having curve-selection. Then* (6) *is not satisfiable over the frame* $(\mathbb{R}^2, \boldsymbol{M})$*.*

In particular, formula (6) is unsatisfiable over the frame $(\mathbb{R}^2, \boldsymbol{RCP}(\mathbb{R}^2))$; hence $Sat(\mathcal{B}c^\circ, \boldsymbol{RC}(\mathbb{R}^2)) \neq Sat(\mathcal{B}c^\circ, \boldsymbol{RCP}(\mathbb{R}^2))$. We remark that many other natural collections of 'tame' regions exhibit the property of finite decomposability and curve-selection—most notably, the regular closed *semi-algebraic* sets (see, e.g. [63]).

Unfortunately, little is known about these logics. In particular, it would be interesting to investigate the complexity of sets like $Sat(\mathcal{B}c, \boldsymbol{RC}(\mathbb{R}^2))$, $Sat(\mathcal{B}c, \boldsymbol{RCP}(\mathbb{R}^2))$, $Sat(\mathcal{B}c^\circ, \boldsymbol{RC}(\mathbb{R}^2))$, $Sat(\mathcal{B}c^\circ, \boldsymbol{RCP}(\mathbb{R}^2))$. The computational behaviour of $\mathcal{B}c^\circ$ over simpler classes of frames such as Reg and ConReg remains also open for investigation.

5 Boolean contact algebras

By restricting interpretations of variables to regular closed (or open) sets and, correspondingly, the language of $\mathcal{S}4_u$ to its fragment $\mathcal{B}$, we considerably restrict the expressive capabilities of our spatial logics. In particular, Whitehead's 'extensive connection' [65] $C(\tau_1, \tau_2)$, which has historically played a prominent role in region-based theories of space, cannot be expressed by means of $\mathcal{B}$-formulas despite its very simple intended meaning:

$$\mathfrak{M} \models C(\tau_1, \tau_2) \quad \text{iff} \quad (\tau_2^{\mathfrak{M}})^- \cap (\tau_2^{\mathfrak{M}})^- \neq \emptyset.$$

In $\mathcal{S}4_u$, we clearly have $C(\tau_1, \tau_2) \ \equiv \ (\tau_1^- \cap \tau_2^- \neq \mathbf{0})$.

So we define the language $\mathcal{C}$ by extending $\mathcal{B}$ with the binary predicate C, interpreted as above. Thus the $\mathcal{C}$*-terms* are precisely the $\mathcal{B}$-terms and the $\mathcal{C}$*-formulas* are defined in the same way as the $\mathcal{B}$-formulas, except that we have the additional clause

$$\varphi \quad ::= \quad \ldots \quad | \quad C(\tau_1, \tau_2) \quad | \quad \ldots,$$

where τ_1 and τ_2 are $\mathcal{C}$-terms. Since Whitehead's term 'extensive connection' risks confusion with the standard topological notion of *connectedness*, we follow more recent usage and read $C(\tau_1, \tau_2)$ as 'τ_1 contacts τ_2.'

It turns out that $\mathcal{C}$ is adequate for the reconstruction of topology within a 'region-based' ontology, in the following sense. A *closed mereotopology* is a topological frame $(T, \boldsymbol{M})$, such that $\boldsymbol{M}$ is (i) a Boolean sub-algebra of $\boldsymbol{RC}(T)$; and (ii) $\boldsymbol{M}$ is a closed basis for the topology on T. It is easy to verify that, over any mereotopology, the following $\mathcal{C}$-formulas are valid:

$$\neg C(r, \mathbf{0}), \tag{7}$$

$$(r \neq \mathbf{0}) \rightarrow C(r, r), \tag{8}$$

$$C(r, s) \rightarrow C(s, r), \tag{9}$$

$$C(r, s) \wedge (s \leq t) \rightarrow C(r, t), \tag{10}$$

$$C(r, s + t) \rightarrow C(r, s) \vee C(r, t). \tag{11}$$

$\mathsf{DC}(r,s)$	$\neg C(r,s)$	r and s are disconnected
$\mathsf{EC}(r,s)$	$(r \cdot s = \mathbf{0}) \wedge C(r,s)$	r and s are externally connected
$\mathsf{EQ}(r,s)$	$r = s$	r and s are equal
$\mathsf{PO}(r,s)$	$(r \cdot s \neq \mathbf{0}) \wedge ((-r) \cdot s \neq \mathbf{0}) \wedge (r \cdot (-s) \neq \mathbf{0})$	r and s partially overlap
$\mathsf{TPP}(r,s)$	$(r \cdot (-s) = \mathbf{0}) \wedge C(r,-s)$	r is a tangential proper part of s
$\mathsf{NTPP}(r,s)$	$\neg C(r,-s)$	r is a non-tangential proper part of s
$\mathsf{TPP}^{-1}(r,s)$	$(s \cdot (-r) = \mathbf{0}) \wedge C(s,-r)$	s is a tangential proper part of r
$\mathsf{NTPP}^{-1}(r,s)$	$\neg C(s,-r)$	s is a non-tangential proper part of r

Table 1. The $\mathcal{RCC}$-8 relations in the language $\mathcal{C}$.

Structures (in the first-order sense) satisfying the usual axioms of Boolean algebras together with the universal closures of (7)–(11) are known as *Boolean contact algebras*. Thus, any mereotopology is a Boolean contact algebra. In fact we have a converse: every Boolean contact algebra is isomorphic to some mereotopology $(T, \boldsymbol{M})$, where T is a semiregular and compact T_0-space [16, 17]. Axiom sets corresponding to closed mereotopologies over spaces satisfying certain separation properties have also been obtained [16, 17, 19, 51]; see also [56].

The complexity of reasoning in $\mathcal{C}$ was studied in [66], where this logic was introduced under the name $\mathcal{BRCC}$-8 in recognition of the fact that it is able to express the eight relationships in Table 1, which have played an important role in the recent development of spatial logics.

THEOREM 15 ([66]). *Sat*$(\mathcal{C}, \textsc{Reg})$ *is* NP-*complete.*

This result follows from the fact that every satisfiable $\mathcal{C}$-formula φ can be satisfied in a frame belonging to $\textsc{Alek}^{\leq 1}$ which is a disjoint union of linearly many (in the length of φ) forks; see Fig. 8.

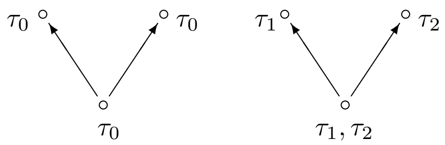

Figure 8. Satisfying $(\tau_0 \neq \mathbf{0}) \;\wedge\; C(\tau_1, \tau_2)$ in a disjoint union of forks.

THEOREM 16 ([66]). *Sat*$(\mathcal{C}, \textsc{ConReg}) =$ *Sat*$(\mathcal{C}, \boldsymbol{RC}(\mathbb{R}^n))$; *and this set is* PSPACE-*complete.*

To show that the two sets coincide, one can use two observations: (i) every $\mathcal{C}$-formula satisfiable in a frame from $\textsc{ConReg}$ is satisfiable in a finite *saw* Aleksandrov model, i.e., a model induced by a partial order R on $\{x_0, z_1, x_1, z_2, x_2, \dots, z_n, x_n\}$ such that $z_i R x_{i-1}$ and $z_i R x_i$, for $1 \leq i \leq n$;

and (ii) every saw Aleksandrov model can be embedded into $\mathbb{R}^n$. The first observation follows from the fact that every formula in $Sat(\mathcal{C}, \textsc{ConReg})$ is satisfiable in a connected quasi-saw Aleksandrov model, which can be transformed (by duplicating points) into a saw model. The proof of (ii) for $n = 1$ is illustrated in Fig. 9 (points of depth 0 correspond to closed intervals and points of depth 1 to the end-points of those intervals).

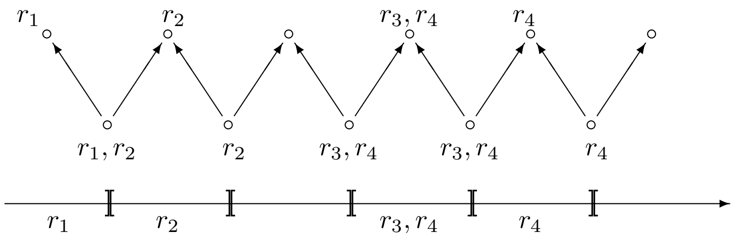

Figure 9. Embedding a saw model in $\mathbb{R}$.

The complexity result is proved similarly to Theorem 5; a full proof can be found in [31]. In fact, it turns out that $\mathcal{C}$ is powerful enough to model binary counters (as in Section 3).

We also mention here two results from [3, Section 9]: (i) axioms (7)–(11) together with the axioms for Boolean algebras axiomatize the validities of $\mathcal{C}$ over $\textsc{Reg}$; and (ii) the extra axiom

$$(r \neq \mathbf{0}) \wedge (r \neq \mathbf{1}) \quad \rightarrow \quad C(r, -r)$$

is required to axiomatize $\mathcal{C}$-validities over $\textsc{ConReg}$.

As with the languages $\mathcal{S}4_u$ and $\mathcal{B}$, it is of interest to consider the extensions $\mathcal{C}c$ and $\mathcal{C}cc$ of $\mathcal{C}$ with the predicates c and $c^{\leq k}$. Surprisingly enough, these languages are of the same complexity as their $\mathcal{S}4_u$ counterparts:

THEOREM 17 ([31]). (i) *$Sat(\mathcal{C}c, \textsc{Reg})$ and $Sat(\mathcal{C}c, \textsc{ConReg})$ are distinct and* ExpTime*-complete*; *$Sat(\mathcal{C}c, \boldsymbol{RC}(\mathbb{R}^n))$, for $n \geq 2$, is* ExpTime*-hard.*

(ii) *$Sat(\mathcal{C}cc, \textsc{Reg})$ and $Sat(\mathcal{C}cc, \textsc{ConReg})$ are distinct and* NExpTime*-complete*; *$Sat(\mathcal{C}cc, \boldsymbol{RC}(\mathbb{R}^n))$, for $n \geq 2$, is* NExpTime*-hard.*

Another surprising result is that the satisfiability problem for $\mathcal{C}c$ (and $\mathcal{C}cc$) is reducible to the satisfiability problem for $\mathcal{B}c$ ($\mathcal{B}cc$, respectively). Clearly, two connected closed sets are in contact iff their union is connected; that is to say, the formula $c(\tau_1) \wedge c(\tau_2) \rightarrow \big(C(\tau_1, \tau_2) \;\leftrightarrow\; c(\tau_1 + \tau_2)\big)$ is a $\mathcal{C}cc$-validity. However, this 'reduction' of C to c assumes the arguments of $C(\tau_1, \tau_2)$ to be *connected*, which is not in general the case. Roughly, the idea behind the reduction is as follows. If $\mathfrak{M} \models C(\tau_1, \tau_2)$ then there are connected components X_i of $\tau_i^{\mathfrak{M}}$ such that $X_1 \cap X_2 \neq \emptyset$. So we can introduce fresh variables t_i for X_i, for which $\mathfrak{M} \models c(t_1 + t_2) \wedge c(t_1) \wedge c(t_2)$. On the other hand, if $\mathfrak{M} \models \neg C(\tau_1, \tau_2)$ then we can extend the Aleksandrov space T underlying $\mathfrak{M}$ with two extra points u_1 and u_2 that 'connect' all the points of τ_1 and τ_2, respectively, and consider the new connected sets $X_i = \tau_i^{\mathfrak{M}} \cup \{u_i\}$. By introducing fresh variables t_i for X_i, we then have $\mathfrak{M} \models \neg c(t_1 + t_2) \wedge c(t_1) \wedge c(t_2)$. For more details consult [31].

6 $\mathcal{RCC}$-8

In Sections 2 and 3, we considered topological logics based on the term-language $\mathcal{S}4$. In Section 4, we investigated the result of (in effect) restricting this term-language to those terms that denote regular closed sets. In this section, we consider topological languages in which terms have no structure at all: they are simply variables.

What topological primitives might we employ over such an impoverished term-language? The possibilities are almost endless; historically, however, one particular collection of primitives has held centre-stage. Define the language $\mathcal{RCC}$-8 as follows. The $\mathcal{RCC}$-8-*terms* are simply the variables. The $\mathcal{RCC}$-8-*formulas*, φ, are given by

$$\begin{aligned}\varphi \quad ::= \quad & r \,\mathsf{DC}\, s \quad | \quad r \,\mathsf{EC}\, s \quad | \quad r \,\mathsf{EQ}\, s \quad | \quad r \,\mathsf{PO}\, s \quad | \quad r \,\mathsf{TPP}\, s \quad | \\ & r \,\mathsf{NTPP}\, s \quad | \quad r \,\mathsf{TPP}^{-1}\, s \quad | \quad r \,\mathsf{NTPP}^{-1}\, s \quad | \quad \neg\varphi \quad | \quad \varphi_1 \wedge \varphi_2,\end{aligned}$$

where r and s are variables (i.e. $\mathcal{RCC}$-8-terms). The semantics of $\mathcal{RCC}$-8 is defined according to Table 1, under the restriction that $\mathcal{RCC}$-8-terms are interpreted by non-empty, regular closed sets of topological spaces. $\mathcal{RCC}$-8 and similar formalisms were originally introduced in the area of knowledge representation and reasoning in AI, in particular, geographical information systems; see [20, 21, 55, 46]. The fact that it is actually a simple fragment of $\mathcal{S}4_u$ was first observed by Bennett [5]; see also [48, 42] (in fact, $\mathcal{RCC}$-8 can be embedded in $\mathcal{S}5$ [67]).

$\mathcal{RCC}$-8 is rather inexpressive. As was observed by Renz [47], we have:

THEOREM 18 ([47]). *Sat*$(\mathcal{RCC}\text{-}8, \textsc{Reg})$ = *Sat*$(\mathcal{RCC}\text{-}8, \textsc{ConReg})$ = *Sat*$(\mathcal{RCC}\text{-}8, \boldsymbol{RC}(\mathbb{R}^n))$*, for any* $n \geq 1$*; this set is* NP*-complete.*

NP-completeness follows from the fact that—similarly to $\mathcal{C}$—every satisfiable $\mathcal{RCC}$-8-formula φ is satisfied in a disjoint union of linearly many (in the length of φ) forks. Tractable fragments of $\mathcal{RCC}$-8 were analyzed in [48].

Actually, the NP-hardness result here arises entirely from the Boolean combinations available in formulas. Indeed, the following holds:

THEOREM 19 ([27]). *The problem of determining whether a conjunctive* $\mathcal{RCC}$*-8-formula is satisfiable in* $\textsc{Reg}$ *is* $\textsc{NLogSpace}$*-complete.*

$\mathcal{RCC}$-8's lack of expressiveness at the level of the term-language opens up additional possibilities for restrictions on the topological frames considered, for it is perfectly natural to interpret $\mathcal{RCC}$-8 over topological frames $(T, \boldsymbol{M})$ in which $\boldsymbol{M}$ is a sub*set* (not necessarily a sub*algebra*) of $\boldsymbol{RC}(T)$. For instance, let $\boldsymbol{C}(T)$ be the set of non-empty, connected, regular closed subsets of the space T. It is easy to see that, for $n \geq 3$, *Sat*$(\mathcal{RCC}\text{-}8, \boldsymbol{C}(\mathbb{R}^n))$ = *Sat*$(\mathcal{RCC}\text{-}8, \boldsymbol{RC}(\mathbb{R}^n))$ [47]. However, this is not true for $n = 1$ or $n = 2$. For the latter case, we consider an example similar to (4). Let $r_{i,j}$, $1 \leq i < j \leq 5$, be distinct variables other than r_i, $1 \leq i \leq 5$; and let φ be the formula

$$\bigwedge_{i \in \{j,k\}} \mathsf{NTPP}(r_i, r_{j,k}) \quad \wedge \quad \bigwedge_{\{i,j\} \cap \{k,l\} = \emptyset} \mathsf{DC}(r_{i,j}, r_{k,l}).$$

Clearly, φ is not satisfiable over $\boldsymbol{C}(\mathbb{R}^2)$, again, because, if it were, one could construct a plane drawing of K_5.

Another salient frame over which to interpret $\mathcal{RCC}$-8 is the collection $\boldsymbol{D}$ of subsets of $\mathbb{R}^2$ homeomorphic to the closed unit disc. This interpretation is noteworthy for the following reason. Fix some topological frame $(T, \boldsymbol{S})$, and let $R_1, \ldots, R_8$ be the relations over $\boldsymbol{S}$ expressed by the $\mathcal{RCC}$-8-primitives, as specified in Table 1. It is routine to show that $R_1, \ldots, R_8$ are mutually exclusive and jointly exhaustive: any ordered pair of elements of $\boldsymbol{S}$ belongs to exactly one of these relations. But now consider the various *relative products* $R_i \circ R_j$. If each of these relative products is the *union* of some subset of $\{R_1, \ldots, R_n\}$, then these relations generate a finite relation algebra.

THEOREM 20 ([38]). *The relations expressed by the $\mathcal{RCC}$-8-predicates, interpreted over the set $\boldsymbol{D}$ of disc-homeomorphs in $\mathbb{R}^2$, are the atoms of a relation algebra.*

Schaefer *et al.* [52] analyse the relationship between $Sat(\mathcal{RCC}\text{-}8, (\mathbb{R}^2, \boldsymbol{D}))$ and the problem of determining the weak realizability of topological graphs. Let $G = (V, E)$ be a graph and R a set of (unordered) pairs of elements of E. We say that (G, R) is *weakly realizable* if there is a drawing of G such that only the pairs of edges allowed to cross are those occurring in R. Schaefer *et al.* show that the problem of determining weak realizability of graphs is in NP. This is a remarkable result, as it is known that some weakly realizable topological graphs require drawings in which the number of crossing points is bounded below by an exponential function of the size of the graph [33]. Using the close relationship between weak realizability of topological graphs and satisfiability of $\mathcal{RCC}$-8-formulas by closed discs, Schaefer *et al.* obtain:

THEOREM 21 ([52]). *The problem $Sat(\mathcal{RCC}\text{-}8, (\mathbb{R}^2, \boldsymbol{D}))$ is* NP-*complete.*

7 n-ary contact relation

As formulas in $\mathcal{RCC}$-8 and $\mathcal{C}$ are built from $\mathcal{B}$-terms using the *binary* predicates $\tau_1 = \tau_2$ and $C(\tau_1, \tau_2)$, they are not capable of expressing certain relations involving three or more regions. An obvious way of extending the expressive power of $\mathcal{C}$ in this direction is to generalize the contact predicate and consider the extension $\mathcal{C}^m$ of $\mathcal{B}$ with arbitrary n-ary contact relations $C(\tau_1, \ldots, \tau_n)$, for $n > 1$: the $\mathcal{C}^m$-formulas are defined in the same way as $\mathcal{B}$-formulas, except that we have the additional clause

$$\varphi \quad ::= \quad \ldots \quad | \quad C(\tau_1, \ldots, \tau_n) \quad | \quad \ldots .$$

The definition of the truth-relation is extended as follows:

$$\mathfrak{M} \models C(\tau_1, \ldots, \tau_n) \quad \text{iff} \quad (\tau_1^{\mathfrak{M}})^- \cap \cdots \cap (\tau_n^{\mathfrak{M}})^- \neq \emptyset.$$

Clearly, $\mathcal{C}^m$ can also be regarded as a fragment of $\mathcal{S}4_u$; indeed, the n-ary contact relation is definable in $\mathcal{S}4_u$ as $C(\tau_1, \ldots, \tau_n) \equiv (\tau_1^- \cap \cdots \cap \tau_n^- \neq \mathbf{0})$.

The extra expressive power of $\mathcal{C}^m$ as compared to $\mathcal{RCC}$-8 and $\mathcal{C}$ can be illustrated by the following formula

$$C(r_1, r_2, r_3) \quad \wedge \bigwedge_{1 \leq i < j \leq 3} (r_i \cdot r_j = \mathbf{0}) \tag{12}$$

which says that boundaries of three regular closed sets r_1, r_2 and r_3 meet somewhere but the three sets have no common interior points. In particular, in order to satisfy it in Aleksandrov spaces, one requires a partial order of width 3 (see Fig. 10 a)) unlike for $\mathcal{C}$, where partial orders of width 2 (disjoint unions of forks) were enough. A model over $\boldsymbol{RC}(\mathbb{R}^2)$ satisfying (12) is depicted in Fig. 10 b): it interprets each r_i as a third of the disc; then the centre of the disc is the point in $r_1 \cap r_2 \cap r_3$ (i.e., a witness for $C(r_1, r_2, r_3)$).

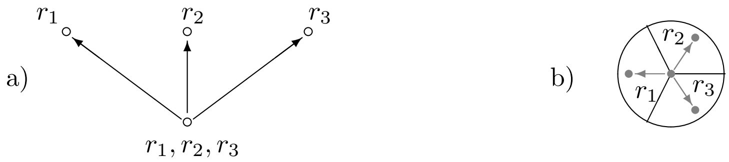

Figure 10. A 3-broom satisfying (12) and a satisfying model over $\boldsymbol{RC}(\mathbb{R}^2)$.

Despite the extra expressiveness, $\mathcal{C}^m$ is of the same complexity as $\mathcal{C}$:

THEOREM 22 ([23]). *Sat*$(\mathcal{C}^m, \text{Reg})$ *is* NP-*complete.*

This follows from the fact that every satisfiable $\mathcal{C}^m$-formula φ can be satisfied in an Aleksandrov model that is a disjoint union of linearly many (in $|\varphi|$) n-brooms, where n is the maximum arity of contact predicate in φ (each n-broom is a partial order $(\{z, x_1, \dots, x_n\}, R)$ with zRx_i, for all i).

It is also of interest to note that, unlike $\mathcal{C}$, $\mathcal{C}^m$ distinguishes between $\mathbb{R}$ and $\mathbb{R}^2$: formula (12) is clearly satisfiable in $\mathbb{R}^2$ but not in $\mathbb{R}$ (cf. Theorem 16).

THEOREM 23. *For* $n > 1$, *Sat*$(\mathcal{C}^m, \boldsymbol{RC}(\mathbb{R})) \neq$ *Sat*$(\mathcal{C}^m, \boldsymbol{RC}(\mathbb{R}^n))$ *and Sat*$(\mathcal{C}^m, \boldsymbol{RC}(\mathbb{R}^n)) =$ *Sat*$(\mathcal{C}^m, \text{ConReg})$. *All these sets are* PSpace-*complete.*

The following complexity results for the extensions $\mathcal{C}^m c$ and $\mathcal{C}^m cc$ of $\mathcal{C}^m$ are immediate consequences of the results considered earlier in the paper:

COROLLARY 24. *Sat*$(\mathcal{C}^m c, \text{Reg})$, *Sat*$(\mathcal{C}^m c, \text{ConReg})$ *are* ExpTime-*complete; Sat*$(\mathcal{C}^m cc, \text{Reg})$, *Sat*$(\mathcal{C}^m cc, \text{ConReg})$ *are* NExpTime-*complete.*

Finally, we note that every satisfiable $\mathcal{C}^m cc$-formula can be satisfied in an Aleksandrov model based on a frame of depth 1:

THEOREM 25. *For every finite Aleksandrov model* $\mathfrak{M}$ *induced by a quasi-order* (W, R), *there is an* $\text{Alek}^{\leq 1}$ *model* $\mathfrak{M}'$ *induced by* (W, R'), $R' \subseteq R$, *such that, for each* $\mathcal{B}$*-term* τ,

- $\tau^{\mathfrak{M}} = \tau^{\mathfrak{M}'}$;
- $\tau^{\mathfrak{M}}$ *and* $\tau^{\mathfrak{M}'}$ *have the same number of connected components.*

It follows that if $\mathfrak{M} \models \varphi$ *then* $\mathfrak{M}' \models \varphi$, *for every* $\mathcal{C}^m cc$*-formula* φ.

8 Conclusion

We conclude the paper with a table summarizing the complexity results considered above as well as the open problems. Merged cells in the table mean that the corresponding logics coincide, ExP stands for ExpTime, and NExp stands for NExpTime.

	Reg	ConReg	$\boldsymbol{RC}(\mathbb{R}^n)$ $n > 2$	$\boldsymbol{RC}(\mathbb{R}^2)$	$\boldsymbol{RC}(\mathbb{R})$
$\mathcal{RCC}$-8	NP Thm. 18				
$\mathcal{RCC}$-8c				?	≤PSpace,≥NP
$\mathcal{RCC}$-8cc				?	≤PSpace,≥NP
$\mathcal{B}$	NP				
$\mathcal{B}c$	Exp	Exp	?	?	≤PSpace,≥NP
$\mathcal{B}cc$	NExp	NExp	?	?	≤PSpace,≥NP
$\mathcal{C}$	NP Thm. 15	PSpace Thm. 16			
$\mathcal{C}c$	Exp Thm. 17	Exp Thm. 17	≥Exp Thm. 17	≥Exp Thm. 17	PSpace
$\mathcal{C}cc$	NExp Thm. 17	NExp Thm. 17	≥NExp Thm. 17	≥NExp Thm. 17	PSpace
$\mathcal{C}^m$	NP Thm. 22	PSpace Thm. 23		PSpace Thm. 23	PSpace
$\mathcal{C}^m c$	Exp	Exp	≥Exp	≥Exp	PSpace
$\mathcal{C}^m cc$	NExp	NExp	≥NExp	≥NExp	PSpace

	All	Con	$\mathbb{R}^n$, $n > 2$	$\mathbb{R}^2$	$\mathbb{R}$
$\mathcal{S}4_u$	PSpace Thm. 2	PSpace Thm. 3			
$\mathcal{S}4_u c$	Exp Thm. 6	Exp Thm. 10	≥Exp	≥Exp	PSpace
$\mathcal{S}4_u cc$	NExp Thm. 7	NExp Thm. 10	≥NExp	≥NExp	PSpace Thm. 9

Acknowledgements

The work on this paper was partially supported by the U.K. EPSRC research grants EP/E034942/1 and EP/E035248/1. We are grateful to Dimiter Vakarelov for comments and discussions.

BIBLIOGRAPHY

[1] M. Aiello and J. van Benthem. A modal walk through space. *J. of Applied Non-Classical Logics*, 12(3–4):319–364, 2002.

[2] C. Areces, P. Blackburn, and M. Marx. The computational complexity of hybrid temporal logics. *Logic J. of the IGPL*, 8:653–679, 2000.

[3] Ph. Balbiani, T. Tinchev, and D. Vakarelov. Modal logics for region-based theories of space. *Fundamenta Informaticae*, 81:29–82, 2007.

[4] M. Ben-Or, D. Kozen, and J. Reif. The complexity of elementary algebra and geometry. *J. of Computer and System Sciences*, 32:251–264, 1986.

[5] B. Bennett. Spatial reasoning with propositional logic. In *Proceedings of KR*, pages 51–62. Morgan Kaufmann, 1994.

[6] B. Bennett and I. Düntsch. Axioms, algebras and topology. In M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 99–160. Springer, 2007.

[7] L. Biacino and G. Gerla. Connection structures. *Notre Dame J. of Formal Logic*, 32:242–247, 1991.

[8] N. Bourbaki. *General Topology, Part 1.* Hermann, Paris and Addison-Wesley, 1966.

[9] D. Cantone and V. Cutello. Decision algorithms for elementary topology I. Topological syllogistics with set and map constructs, connectedness and cardinailty composition. *Communications on Pure and Applied Mathematics*, XLVII:1197–1217, 1994.

[10] B. L. Clarke. A calculus of individuals based on 'connection'. *Notre Dame J. of Formal Logic*, 23:204–218, 1981.

[11] B. L. Clarke. Individuals and points. *Notre Dame J. of Formal Logic*, 26:61–75, 1985.

[12] A. G. Cohn and J. Renz. Qualitative spatial representation and reasoning. In F. van Hermelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 551–596. Elsevier, 2008.

[13] E. Davis. The expressivity of quantifying over regions. *J. of Logic and Computation*, 16:891–916, 2006.

[14] G. De Giacomo. *Decidability of Class-Based Knowledge Representation Formalisms.* PhD thesis, Università degli Studi di Roma 'La Sapienza', 1995.

[15] T. de Laguna. Point, line and surface as sets of solids. *The Journal of Philosophy*, 19:449–6, 1922.

[16] G. Dimov and D. Vakarelov. Contact algebras and region-based theory of space: A proximity approach, I. *Fundamenta Informaticae*, 74(2–3):209–249, 2006.

[17] G. Dimov and D. Vakarelov. Contact algebras and region-based theory of space: A proximity approach, II. *Fundamenta Informaticae*, 74(2–3):251–282, 2006.

[18] C. Dornheim. Undecidability of plane polygonal mereotopology. In A. Cohn, L. Schubert, and S. Shapiro, editors, *Proceedings of KR*, pages 342–353. Morgan Kaufmann, 1998.

[19] I. Düntsch and M. Winter. A representation theorem for Boolean contact algebras. *Theoretical Computer Science*, 347:498–512, 2005.

[20] M. Egenhofer and R. Franzosa. Point-set topological spatial relations. *International J. of Geographical Information Systems*, 5:161–174, 1991.

[21] M. Egenhofer and J. Herring. Categorizing topological relationships between regions, lines and point in geographic databases. Technical report, University of Maine, 1991.

[22] M. Fischer and M. Rabin. Super-exponential complexity of Presburger arithmetic. *Complexity of Computation, SIAM-AMS Proceedings*, 7:27–41, 1974.

[23] D. Gabelaia, R. Kontchakov, A. Kurucz, F. Wolter, and M. Zakharyaschev. Combining spatial and temporal logics: expressiveness vs. complexity. *J. of Artificial Intelligence Research (JAIR)*, 23:167–243, 2005.

[24] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933.

[25] R. Goldblatt. Mathematical modal logic: A view of its evolution. In D. M. Gabbay and J. Woods, editors, *Handbook of the History of Logic, Volume 7: Logic and the Modalities in the Twentieth Century*, pages 1–98. Elsevier, 2006.

[26] V. Goranko and S. Passy. Using the universal modality: gains and questions. *J. of Logic and Computation*, 2:5–30, 1992.

[27] A. Griffiths. *Computational Properties Of Spatial Logics In The Real Plane.* PhD thesis, Depart. of Computer Science, University of Manchester, 2008. Forthcoming.

[28] A. Grzegorczyk. Undecidability of some topological theories. *Fundamenta Mathematicae*, 38:137–152, 1951.

[29] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic.* MIT Press, 2000.

[30] D. Hilbert. *Grundlagen der Geometrie.* B.G. Teubner, Leipzig and Berlin, 1909.

[31] R. Kontchakov, I. Pratt-Hartmann, F. Wolter, and M. Zakharyaschev. On the computational complexity of spatial logics with connectedness constraints. Submitted, 2008.

[32] S. Koppelberg. *Handbook of Boolean Algebras*, volume 1. North-Holland, 1989.

[33] J. Kratochvíl and J. Matoušek. String graphs requiring exponential representations. *J. of Combinatorial Theory, Series B*, 53:1–4, 1991.

[34] G. Kuper, L. Libkin, and J. Paredaens, editors. *Constraint Databases*. Springer, 2000.

[35] O. Kutz, H. Sturm, N.-Y. Suzuki, F. Wolter, and M. Zakharyaschev. Logics of metric spaces. *ACM Transactions on Computational Logic*, 4:260–294, 2003.

[36] R. Ladner. The computational complexity of provability in systems of modal logic. *SIAM J. on Computing*, 6:467–480, 1977.

[37] C. Lewis and C. Langford. *Symbolic Logic*. Appleton-Century-Crofts, New York, 1932.

[38] S. Li and M. Ying. Extensionality of the RCC8 composition table. *Fundamenta Informaticae*, 55(3–4):363–385, 2003.

[39] C. Lutz and F. Wolter. Modal logics of topological relations. *Logical Methods in Computer Science*, 2, 2006.

[40] J.C.C. McKinsey and A. Tarski. The algebra of topology. *Annals of Mathematics*, 45:141–191, 1944.

[41] M.H.A. Newman. *Elements of the Topology of Plane Sets of Points*. Cambridge, 1964.

[42] W. Nutt. On the translation of qualitative spatial reasoning problems into modal logics. In W. Burgard, T. Christaller, and A. Cremers, editors, *Proceedings of KI*, volume 1701 of *LNCS*, pages 113–124. Springer, 1999.

[43] I. Orlov. The calculus of compatibility of propositions. *Mathematics of the USSR, Sbornik*, 35:263–286, 1928. (In Russian).

[44] I. Pratt-Hartmann. A topological constraint language with component counting. *J. of Applied Non-Classical Logics*, 12:441–467, 2002.

[45] I. Pratt-Hartmann. First-order mereotopology. In M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 13–97. Springer, 2007.

[46] D. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In B. Nebel, C. Rich, and W. Swartout, editors, *Proceedings of KR*, pages 165–176. Morgan Kaufmann, 1992.

[47] J. Renz. A canonical model of the region connection calculus. In A. Cohn, L. Schubert, and S. Shapiro, editors, *Proceedings of KR*, pages 330–341. Morgan Kaufmann, 1998.

[48] J. Renz and B. Nebel. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. In *IJCAI*, pages 522–527, 1997.

[49] J. Renz and B. Nebel. Qualitative spatial reasoning using constraint calculi. In M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 161–216. Springer, 2007.

[50] M. Reynolds. The complexity of the temporal logic over the reals. Manuscript; available at `http://www.csse.uwa.edu.au/ mark/research/Online/CORT.htm`, 2008.

[51] P. Roeper. Region-based topology. *J. of Philosophical Logic*, 26:251–309, 1997.

[52] M. Schaefer, E. Sedgwick, and D. Štefankovič. Recognizing string graphs in NP. *J. of Computer and System Sciences*, 67:365–380, 2003.

[53] V. Shehtman. "Everywhere" and "Here". *J. of Applied Non-Classical Logics*, 9, 1999.

[54] M. Sheremet, D. Tishkovsky, F. Wolter, and M. Zakharyaschev. From topology to metric: modal logic and quantification in metric spaces. In G. Governatori, I. Hodkinson, and Y. Venema, editors, *Advances in Modal Logic 6*, pages 429–448, 2006.

[55] T. Smith and K. Park. An algebraic approach to spatial reasoning. *International J. of Geographical Information Systems*, 6:177–192, 1992.

[56] J. Stell. Boolean connection algebras: A new approach to the Region Connection Calculus. *Artificial Intelligence*, 122:111–136, 2000.

[57] D. Sustretov. Topological semantics and decidability. arXiv:math/0703106v3 [math.LO], February 2008.

[58] A. Tarski. On the calculus of relations. *J. of Symbolic Logic*, 6(3):73–89, 1941.

[59] A. Tarski. What is Elementary Geometry? In *The Axiomatic Method, with Special Reference to Geometry and Physics*, pages 16–29. North-Holland, 1959.

[60] B. ten Cate, D. Gabelaia, and D. Sustretov. Modal languages for topology: expressivity and definability. arXiv:math/0610357v2 [math.LO], October 2006.

[61] D. Vakarelov. Region-based theory of space: algebras of regions, representation theory, and logics. In D. M. Gabbay, S. Goncharov, and M. Zakharyaschev, editors, *Mathematical Problems from Applied Logic II*, pages 267–348. Springer, 2007.

[62] J. van Benthem and G. Bezhanishvili. Modal logics of space. In M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 217–298. Springer, 2007.

[63] L. van den Dries. *Tame Topology and O-Minimal Structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge, 1998.

[64] P. van Emde Boas. The convenience of tilings. In A. Sorbi, editor, *Complexity, Logic and Recursion Theory*, volume 187 of *Lecture Notes in Pure and Applied Mathematics*, pages 331–363. Marcel Dekker Inc., 1997.

[65] A. N. Whitehead. *Process and Reality*. New York: The MacMillan Company, 1929.

[66] F. Wolter and M. Zakharyaschev. Spatial reasoning in RCC-8 with Boolean region terms. In W. Horn, editor, *Proceedings of ECAI*, pages 244–248. IOS Press, 2000.

[67] F. Wolter and M. Zakharyaschev. Qualitative spatio-temporal representation and reasoning: a computational perspective. In G. Lakemeyer and B. Nebel, editors, *Exploring AI in the New Millenium*, pages 175–216. Morgan Kaufmann, 2002.

[68] F. Wolter and M. Zakharyaschev. A logic for metric and topology. *J. of Symbolic Logic*, 70:795–828, 2005.

Roman Kontchakov
School of Computer Science and Information Systems
Birkbeck College
London WC1E 7HX
UK
roman@dcs.bbk.ac.uk

Ian Pratt-Hartmann
Department of Computer Science
Manchester University
Manchester M13 9PL
UK
ipratt@cs.man.ac.uk

F. Wolter
Department of Computer Science
University of Liverpool
Liverpool L69 3BX
UK
frank@csc.liv.ac.uk

Michael Zakharyaschev
School of Computer Science and Information Systems
Birkbeck College
London WC1E 7HX
UK
michael@dcs.bbk.ac.uk