TWO-VARIABLE FIRST-ORDER LOGIC WITH EQUIVALENCE CLOSURE *

EMANUEL KIEROŃSKI, JAKUB MICHALISZYN †, IAN PRATT-HARTMANN ‡, AND LIDIA TENDERA §

Abstract. We consider the satisfiability and finite satisfiability problems for extensions of the two-variable fragment of first-order logic in which an equivalence closure operator can be applied to a fixed number of binary predicates. We show that the satisfiability problem for two-variable, first-order logic with equivalence closure applied to two binary predicates is in 2-NExpTime, and we obtain a matching lower bound by showing that the satisfiability problem for two-variable first-order logic in the presence of two equivalence relations is 2-NExpTime-hard. The logics in question lack the finite model property; however, we show that the same complexity bounds hold for the corresponding finite satisfiability problems. We further show that the satisfiability (=finite satisfiability) problem for the two-variable fragment of first-order logic with equivalence closure applied to a single binary predicate is NExpTime-complete.

Key words. computational complexity, decidability

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction.

We investigate extensions of the two-variable fragment of first-order logic in which certain distinguished binary predicates are declared to be equivalences, or in which an operation of 'equivalence closure' can be applied to these predicates. (The equivalence closure of a binary relation is the smallest equivalence that includes it.) Denoting the two-variable fragment of first-order logic with equality by FO^2, let EQ^2_k be the extension of FO^2 in which k distinguished binary predicates are interpreted as equivalences; and let EC^2_k be the extension of FO^2 in which we can take the equivalence closure of any of k distinguished binary predicates. We determine the computational complexity of the satisfiability and finite satisfiability problems for EQ^2_k and EC^2_k.

As is well-known, FO^2 enjoys the finite model property [25], and its satisfiability (= finite satisfiability) problem is NExpTime-complete [8]. It was shown in [20] that EQ^2_1 also has the finite model property, with satisfiability again NExpTime-complete. However, the same paper showed that the finite model property fails for EQ^2_2, and that its satisfiability problem is in 3-NExpTime. An identical upper bound for the finite satisfiability problem was later obtained in [22]. The best currently known corresponding lower bound for these problems is 2-ExpTime, obtained from the two-variable guarded fragment with equivalence relations [16] (discussed below). It was further shown in [20] that the satisfiability and finite satisfiability problems for EQ^2_3 are undecidable.

In this paper we show: (i) EC^2_1 retains the finite model property, and its satisfiability problem remains in NExpTime; (ii) the satisfiability and finite satisfiability problems for EC^2_2 are both in 2-NExpTime; (iii) the satisfiability and finite satisfiability problems for EQ^2_2 are both 2-NExpTime-hard. Taking into account the above-

*This is a full version of [19]. The authors would like to acknowledge the support of the Polish Ministry of Science and Higher Education grants N N206 37133, DEC-2011/03/N/ST6/00415. The third author would like to express his appreciation to the Institute of Computer Science, University of Wrocław, for its generous support of his sabbatical visit there in 2010–11.

†Institute of Computer Science, University of Wrocław, Wrocław, Poland

‡School of Computer Science, University of Manchester, Manchester, M13 9PL, United Kingdom

§Institute of Mathematics and Informatics, Opole University, Opole, Poland

mentioned results, this settles, for all $k \geq 1$, the complexity of satisfiability and finite satisfiability for both EC^2_k and EQ^2_k: all these problems are NExpTime-complete if $k = 1$, 2-NExpTime-complete if $k = 2$, and undecidable if $k \geq 3$. Thus, in this paper, we close the existing gap for EQ^2_2, and extend the complexity bounds for EQ^2_k to the more expressive logic EC^2_k, for $k = 1, 2$. Additionally, we show that the satisfiability and finite satisfiability problems for FO^2 with one equivalence and one transitive relation (without equality or any other binary relations) are both undecidable. This is a slight strengthening of a result announced in [22], which in turn sharpens an earlier result that FO^2 with two transitive relations is undecidable [13, 16].

The most significant of these new results is the upper complexity bound of 2-NExpTime for EC^2_2. Our strategy involves a non-deterministic reduction from the (finite) satisfiability problem for EC^2_2 to the problem of determining the existence of a (finite) edge-coloured bipartite graph subject to constraints on the numbers of edges of each colour incident to its vertices. This reduction runs in doubly-exponential time, and produces a set of constraints doubly-exponential in the size of the given EC^2_2-formula. We then show that this latter problem is in NPTime, by non-deterministic reduction to integer programming. Crucial to our argument is a 'Carathéodory-type' result on integer programming due to [5].

The logic FO^2 embeds, via the standard translation, multi-modal propositional logic, whose good algorithmic and model-theoretic behaviour is characteristically robust both with respect to extensions of its logical syntax (for example, by fixed point operations) and also with respect to restrictions on the class of structures over which it is interpreted (for example, in the form of conditions on the modal accessibility relations). Furthermore, many varieties of description logic [2]—now a standard paradigm in industrial applications—can be embedded in FO^2 or its various extensions.

In respect of robustness under syntactic extensions, FO^2 appears, by contrast, less attractive: with the notable exception of the counting extension [10, 27, 29], most of its syntactic extensions are undecidable [9, 11]. In respect of restrictions on the structures over which it is interpreted, however, the behaviour of FO^2 is more mixed, and to some extent less well-understood. The most salient such restrictions are those featuring (i) linear orders, (ii) transitive relations and (iii) equivalences. In the presence of a single linear order, the satisfiability and finite satisfiability problems for FO^2 remain NExpTime-complete [26]. For two linear orders, ExpSpace-completeness of finite satisfiability is shown, subject to certain restrictions on signatures, in [30]. (The case of unrestricted signatures, and decidability of the general satisfiability problem are currently open.) For three linear orders, both satisfiability and finite satisfiability are undecidable [17, 26]. Turning to transitive relations, the satisfiability problem for FO^2 in the presence of a single transitive relation has recently been shown to be in 2-NExpTime [33]. (The corresponding finite satisfiability problem is still open.) As mentioned above, both satisfiability and finite satisfiability of FO^2 are undecidable in the presence of two transitive relations. Restricting attention to interpretations involving equivalences yields the logics EQ^2_k, discussed in this paper.

Closely related to these logics are extensions of FO^2 in which the operations of *transitive closure* or *equivalence closure* can be applied to one or more binary predicates. Such operators can be used to express non-first-order notions such as *reachability* or *connectedness* in (directed or undirected) graphs—notions which arise naturally in a wide range of contexts, perhaps most notably in static program analysis. Fragments of first-order logic augmented with an operation of transitive closure for which decidability has been shown are actually rather rare. One case is the logic

$\exists\forall(\mathrm{DTC}^+[E])$, involving the deterministic transitive closure operator, which has an exponential-size model property [12]. Another is the logic obtained by extending the two-variable guarded fragment [1] with a transitive closure operator applied to binary symbols appearing only in guards; the satisfiability problem for this logic is 2-ExpTime-complete [24]. It has recently been shown that satisfiability of the fragment $\exists^*\forall^2$ with transitive closure of one binary relation is decidable in 2-NExpTime [18]. The decidability of satisfiability and finite satisfiability for FO^2 with transitive closure applied to a single binary relation are both still open. Adding equivalence closure operators to FO^2 yields the logics EC^2_k, discussed in this paper.

It is instructive to consider the relation of the above logics to the well-known *guarded fragment*—the subset of first-order logic in which all quantifiers are relativised by atoms [1]. By the *two-variable guarded fragment*, denoted GF^2, we understand the intersection of the guarded fragment with FO^2. It was shown in [7] that GF^2 has the finite model property, and that satisfiability is ExpTime-complete. As with FO^2, so too with GF^2, we can consider extensions in which certain distinguished binary predicates are required to denote transitive relations or equivalences, or in which corresponding closure operations can be applied to these predicates. The complexity bounds for such extensions of FO^2 and GF^2 are in many cases identical, a notable exception (mentioned above) being the case of two equivalences, which, for GF^2 yields a 2-ExpTime-complete logic [16], and for FO^2—as shown in this paper—a 2-NExpTime-complete logic. For GF^2, it also makes sense to study variants in which the distinguished predicates may appear only in guards [6]. In this case, GF^2 with *any* number of equivalences appearing only as guards remains NExpTime-complete [16], while GF^2 with *any* number of transitive relations appearing only as guards is 2-ExpTime-complete [32, 15]. Table 1.1 summarizes the above results.

The paper is organized as follows. In Sec. 2, we define the logics EC^2_k, in which the distinguished binary predicates $r_1, \ldots, r_k$ are paired with the corresponding predicates $r_1^\#, \ldots, r_k^\#$, representing their respective equivalence closures. In Sec. 3 we establish a 'Scott-type' normal form for EC^2_2, allowing us to restrict the nesting of quantifiers to depth two, and then show how this normal form can be transformed into so-called *reduced* normal form, producing a syntactically simpler formula at the cost of an exponential increase in size. In Sec. 4 we recall a *small substructure property* for FO^2 [20], allowing us to replace an arbitrary substructure in a model of some FO^2-formula φ with one whose size is exponentially bounded in the size of φ's signature. Then we prove a technical lemma, adjusting the above to our current purposes, which then will be used in the upper complexity bound for EC^2_2 obtained in Sec. 6. As a by-product, we obtain the finite model property for EC^2_1 along with a NExpTime upper bound on the complexity of satisfiability. In Sec. 5, we define two problems concerning bipartite graphs with coloured edges: the *graph existence* problem and *finite graph existence* problem. We show that both problems are in NPTime, by non-deterministic polynomial-time reduction to integer programming. (This is the most labour-intensive part of the entire proof.) Sec. 6 is then able to establish that the (finite) satisfiability problem for EC^2_2 is in 2-NExpTime via a non-deterministic doubly exponential-time reduction to the (finite) graph existence problem. Sec. 7 shows, using the familiar apparatus of tiling systems, that the satisfiability and finite satisfiability problems for EQ^2_2 are 2-NExpTime-hard. These matching bounds establish the 2-NExpTime-completeness of satisfiability and finite satisfiability for both EC^2_2 and EQ^2_2. In the last section we show that when, instead of EQ^2_2, we consider FO^2 with one equivalence and one transitive relation (or one equivalence and one partial

Logic	Special symbols	Number of special symbols in the signature		
		1	2	3 or more
GF^2 FMP ExpTime [7]	Transitivity	2-ExpTime*) [16]	undecidable [16, 13]	undecidable [6]
	Linear order	NExpTime**) [26]	ExpSpace**) ***) [30]	undecidable [26, 17]
	Equivalence	FMP, NExpTime [20]	2-ExpTime*) [16]	undecidable [20]
FO^2 FMP [25] NExpTime [8]	Transitivity	in 2-NExpTime*) [33]	undecidable [16, 13]	undecidable [11]
	Linear order	NExpTime [26]	ExpSpace***) [30]	undecidable [26, 17]
	Equivalence	FMP, NExpTime [20]	in 3-NExpTime [20, 22] **2-NExpTime** this paper	undecidable [20]
	Equivalence Closure	**FMP, NExpTime** this paper	**2-NExpTime** this paper	undecidable [20]

Table 1.1: Overview of two variable logics over special classes of structures. FMP stands for Finite Model Property. Unless indicated otherwise, the complexity bounds are tight. Key to symbols: *) only general satisfiability; **) follows from the results on FO^2, as any pair of elements is guarded by a linear order; ***) only finite satisfiability and subject to certain restrictions on signatures.

order), both the satisfiability and finite satisfiability problems become undecidable, even when we do not allow equality in the logic. Sections 7 and 8 (containing lower bounds) can be read immediately after the definitions of our logics from Section 2, independently of the intervening material.

2. Preliminaries. We employ standard terminology and notational from model theory throughout this paper (see, e.g. [4]). In particular, we refer to structures using Gothic capital letters, and their domains using the corresponding Roman capitals. We denote by FO^2 the two-variable fragment of first-order logic (with equality), without loss of generality restricting attention to signatures of unary and binary predicates. We denote by EC^2_k the set of FO^2-formulas over any signature $\tau = \tau_0 \cup \{r_1, \dots, r_k\} \cup \{r^\#_1, \dots, r^\#_k\}$, where τ_0 is an arbitrary set containing unary and binary predicates, and $r_1, \dots, r_k$, $r^\#_1, \dots, r^\#_k$ are distinguished binary predicates. In the sequel, any signature τ is assumed to be of the above form (for some appropriate value of k). We denote by EQ^2_k the set of EC^2_k-formulas in which the predicates $r^\#_1, \dots, r^\#_k$ do not occur.

The semantics for EC^2_k are as for FO^2, subject to the restriction that $r^\#_i$ is always interpreted as the *equivalence closure* of r_i. More precisely: we consider only structures $\mathfrak{A}$ in which, for all i ($1 \leq i \leq k$), $(r_i^\#)^\mathfrak{A}$ is the smallest reflexive, symmetric and transitive relation including $r_i^\mathfrak{A}$. The semantics for EQ_k^2 are likewise as for FO^2, but subject to the restriction that r_i is always interpreted as an equivalence. Where a structure is clear from context, we may equivocate between predicates and their extensions, writing, for example, r_i and $r_i^\#$ in place of the technically correct $r_i^\mathfrak{A}$ and $(r_i^\#)^\mathfrak{A}$. To see that EC_k^2 is more expressive than its sub-fragment EQ_k^2, observe that the EC_1^2-formula $\forall x \forall y . r_1^\#(x, y)$ expresses graph connectivity. As this property can be shown not to be expressible in first-order logic (using a standard compactness argument, e.g. cf. Proposition 3.1 in [23]), it follows that it cannot be expressed in any of the logics EQ_k^2.

Let $\mathfrak{A}$ be a structure over τ. We say that there is an r_i-*edge* between a and $a' \in A$ if $\mathfrak{A} \models r_i[a, a']$ or $\mathfrak{A} \models r_i[a', a]$. Distinct elements $a, a' \in A$ are r_i-*connected* if there exists a sequence $a = a_0, a_1, \ldots, a_{k-1}, a_k = a'$ in A such that for all j ($0 \leq j < k$) there is an r_i-edge between a_j and a_{j+1}. Such a sequence is called an r_i-*path* from a to a'. Thus, $\mathfrak{A} \models r_i^\#[a, a']$ if and only if a and a' are r_i-connected. A subset B of A is called r_i-*connected* if every pair of distinct elements of B is r_i-connected. Maximal r_i-connected subsets of A are equivalence classes of $r_i^\#$, and are called $r_i^\#$-*classes*. We also say that elements $a, a' \in A$ are *in free position in* $\mathfrak{A}$ if they are not r_i-connected, for any $i \in \{1, \ldots, k\}$. Similarly, subsets B and B' of A are *in free position in* $\mathfrak{A}$ if every two elements $b \in B$ and $b' \in B'$ are in free position in $\mathfrak{A}$.

We mostly work with the logic EC_2^2. In any structure $\mathfrak{A}$, the relation $r_1^\# \cap r_2^\#$ is also an equivalence, and we refer to its equivalence classes, simply, as *intersections*. Thus, an intersection is a maximal set that is both r_1- and r_2-connected. When discussing induced substructures, a subtlety arises regarding the interpretation of the closure operations. If $B \subseteq A$, we take it that, in the structure $\mathfrak{B}$ induced by B, the interpretation of $r_i^\#$ is given by simple restriction: $(r_i^\#)^\mathfrak{B} = (r_i^\#)^\mathfrak{A} \cap B^2$. This means that, while $(r_i^\#)^\mathfrak{B}$ is certainly an equivalence including $r_i^\mathfrak{B}$, it may not be the smallest, since, for some $a, a' \in B$, an r_i-path connecting a and a' in $\mathfrak{A}$ may contain elements which are not members of B. (Such a situation may arise even when B is an intersection.) To reduce notational clutter, we use the (possibly decorated) letter $\mathfrak{A}$ to denote 'full' structures in which we are guaranteed that $(r_i^\#)^\mathfrak{A}$ is the equivalence closure of $r_i^\mathfrak{A}$. For structures denoted by other letters, $\mathfrak{B}$, $\mathfrak{C}$, ... (again, possibly decorated), no such guarantee applies. Typically, but not always, these latter structures will be induced substructures. Also, when the domain of some structure $\mathfrak{A}$ consists of several disjoint sets, we often emphasize the fact by writing $A = B \,\dot{\cup}\, C$, etc.

An (atomic) 1-*type* (over a given signature) is a maximal satisfiable set of atoms or negated atoms with free variable x. Similarly, an (atomic) 2-*type* is a maximal satisfiable set of atoms and negated atoms with free variables x, y. Note that the numbers of 1-types and 2-types are bounded exponentially in the size of the signature. We often identify a type with the conjunction of all its elements.

For a given τ-structure $\mathfrak{A}$, we denote by $\text{tp}^\mathfrak{A}(a)$ the 1-type *realized* by a, i.e. the 1-type α such that $\mathfrak{A} \models \alpha[a]$. Similarly, for distinct $a, b \in A$, we denote by $\text{tp}^\mathfrak{A}(a, b)$ the 2-type *realized* by the pair a, b, i.e. the 2-type β such that $\mathfrak{A} \models \beta[a, b]$. We denote by $\boldsymbol{\alpha}[\mathfrak{A}]$ the set of all 1-types realized in $\mathfrak{A}$, and by $\boldsymbol{\beta}[\mathfrak{A}]$ the set of all 2-types realized in $\mathfrak{A}$. For $S \subseteq A$, we denote by $\boldsymbol{\alpha}[S]$ the set of all 1-types realized in S, and similarly for $\boldsymbol{\beta}[S]$. For $S_1, S_2 \subseteq A$, we denote by $\boldsymbol{\beta}[S_1, S_2]$ the set of all 2-types $\text{tp}^\mathfrak{A}(a_1, a_2)$ with $a_i \in S_i$, for $i = 1, 2$; we write $\boldsymbol{\beta}[a, S_2]$ in preference to $\boldsymbol{\beta}[\{a\}, S_2]$.

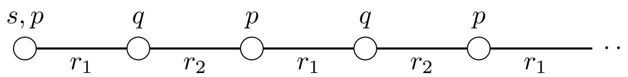

Figure 2.1: Model of an EQ^2_2-formula forcing infinitely many equivalence classes.

We conclude this section with an illustration of the expressive power of the logic EQ^2_2. Specifically, we exhibit a satisfiable formula in all of whose models the equivalences r_1 and r_2 both have infinitely many equivalence classes. This demonstrates the failure of the finite model property for both EQ^2_2 and EC^2_2. (Recall that, by contrast, FO^2 has the finite model property.) Let p, q and s be unary predicates in the signature τ_0. The EQ^2_2-formula

$$\forall x \forall y (r_2(x,y) \wedge p(x) \wedge p(y) \rightarrow x = y)$$

states that each r_2-class contains at most one element satisfying p. Thus, we can evidently write an an EQ^2_2-formula φ expressing the following conditions:

(i) some element satisfies both s and p;
(ii) every element satisfying p is r_1-equivalent to one satisfying q; every element satisfying q is r_2-equivalent to one satisfying p;
(iii) p and q are disjoint and each r_2-class contains at most one element satisfying p and one satisfying q; analogously for r_1-classes; the r_2-class of any element of s is trivial (a singleton);

The structure illustrated in Fig. 2.1 satisfies φ. Conversely, every model of φ contains an infinite chain of this form: choose some element of $s \cap p$ by (i); one then finds new elements in q and p along r_1- and r_2-links in an alternating fashion by appeal to condition (ii); these always have to be fresh elements, i.e., distinct from previous elements in the chain, on pain of violating (iii). A slightly more elaborate construction shows that EQ^2_2 can even force equivalence-classes to be infinite. The interested reader is referred to [16, 21] for more examples.

3. Normal Forms. In the sequel, we take the (possibly decorated) letter p to range over unary predicates, and the (possibly decorated) letter θ to range over quantifier-free (but not necessarily equality-free) FO^2-formulas. If φ is a formula, we write $\neg^0\varphi$ for φ and $\neg^1\varphi$ for $\neg\varphi$. A *normal form* EC^2_2-formula is a sentence

$$\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \psi_{11}, \tag{3.1}$$

where χ is of the form $\forall x \forall y.\theta$ and, for $s,t \in \{0,1\}$, ψ_{st} is a conjunction $\bigwedge_{i \in I} \forall x (p_i(x) \rightarrow \exists y (\neg^s r_1^{\#}(x,y) \wedge \neg^t r_2^{\#}(x,y) \wedge \theta_i))$ (with index set I depending on s and t).

Lemma 3.1. *Let φ be an EC^2_2-formula over a signature τ. We can compute, in polynomial time, a normal-form EC^2_2-formula φ' over a signature τ' such that φ and φ' are satisfiable over the same domains, and τ' consists of τ together with some additional unary predicates.*

Proof. It was shown in [31] that we may compute, in polynomial time, an FO^2-formula $\varphi'' = \forall x \forall y.\chi \wedge \bigwedge_{i \in I} \forall x \exists y.\theta_i$, with the following properties: (i) $\varphi'' \models \varphi$; (ii) any model $\mathfrak{A} \models \varphi$ may be expanded to a model $\mathfrak{A}' \models \varphi''$ by interpreting additional unary predicates. Having computed φ'', take fresh unary predicates $p_{i,s,t}$, for all $i \in I$ and all $s,t \in \{0,1\}$; now let φ' be the result of replacing each conjunct $\forall x \exists y.\theta_i$ in φ''

by

$$\bigwedge_{s,t\in\{0,1\}} \forall x \left(p_{i,s,t}(x) \rightarrow \exists y(\neg^s r_1^\#(x,y) \wedge \neg^t r_2^\#(x,y) \wedge \theta_i)\right),$$

and adding the corresponding conjunct $\forall x(\bigvee_{s,t\in\{0,1\}} p_{i,s,t}(x))$. Reorganizing conjuncts and indices if necessary, φ' has the properties required by the lemma. □

The normal form (3.1) is an elaboration of the normal form for FO^2 presented in [31]. The four conjuncts ψ_{st} allow us to separate out the role of 2-types involving different combinations of the distinguished relations $r_1^\#$ and $r_2^\#$. However, it turns out that slightly simpler formulas suffice for this purpose. A *reduced normal form* EC_2^2-formula is a sentence

$$\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \omega, \tag{3.2}$$

where χ and the ψ_{st} are as in (3.1), and ω is a conjunction $\bigwedge_{i\in I} \exists x.p_i(x)$ for some index set I. Formulas in reduced normal form lack the ψ_{11} conjunct, and feature instead the conjunct ω, whose satisfaction depends only on the set of realized 1-types. As all conjuncts in the formulas ψ_{00}, ψ_{10}, and ψ_{01} are *guarded*, eliminating the (non-guarded) conjunct ψ_{11} simplifies the process of model construction. The following lemma shows that the reduced normal form is general enough for our purposes.

Lemma 3.2. *Given any EC_2^2-formula φ over a signature τ, we can compute, in exponential time, an EC_2^2-formula φ' in reduced normal form over a signature τ', such that: (i) $|\tau'|$ is bounded polynomially in $|\varphi|$; and (ii) φ and φ' are satisfiable over the same domains of cardinality greater than $f(|\varphi|)$ for a fixed exponential function f.*

The rest of this section is devoted to proving Lemma 3.2. We first fix a normal-form EC_2^2-sentence, φ, as in (3.1), over a signature τ. Write

$$\psi_{11} = \bigwedge_{i\in I} \forall x(p_i(x) \rightarrow \exists y(\neg r_1^\#(x,y) \wedge \neg r_2^\#(x,y) \wedge \theta_i(x,y))) \tag{3.3}$$

where $I = \{1,\ldots,m\}$. The following terminology will be useful. If $\mathfrak{A} \models \varphi$ and $a \in A$, then any element $b \in A$ such that $\mathfrak{A} \models \neg r_1^\#[a,b] \wedge \neg r_2^\#[a,b] \wedge \theta_i[a,b]$ is called an *ith free witness* (or simply a *free witness*) for a (*in* $\mathfrak{A}$). Such an ith free witness certainly exists if $\mathfrak{A} \models p_i[a]$.

Lemma 3.3. *Suppose $\mathfrak{A} \models \varphi$, where φ is a normal-form EC_2^2-formula* (3.1) *over τ, with ψ_{11} as in* (3.3)*, and $m = |I|$. Then there is a τ-structure $\mathfrak{A}' \models \varphi$ over the same domain, A, with the following property: there exists $B \subseteq A$, of cardinality at most $Z = 2m(m+2)(3m+5)(1+m+m^2)2^{|\tau|}$ such that, if any $a \in A$ has an ith free witness (for any $1 \leq i \leq m$), then a has an ith free witness in B.*

Proof. If $\alpha \in \boldsymbol{\alpha}[\mathfrak{A}]$, let A_α be the set of elements of A realizing the 1-type α in $\mathfrak{A}$. Our strategy is to define, for each $\alpha \in \boldsymbol{\alpha}[\mathfrak{A}]$, a subset $B_\alpha \subseteq A_\alpha$ of cardinality at most $2m(m+2)(3m+5)$, and to show that, for every $\ell \leq m$ and every $a \in A$, if a has ℓ distinct free witnesses in A_α, then a is in free position with respect to at least ℓ elements of B_α.

Fixing α, denote by s_i the restriction of $r_i^\#$ to A_α. Thus, s_1, s_2 and $s_1 \cap s_2$ are equivalence relations on A_α: in the remainder of this proof, we refer to the equivalence classes of $s_1 \cap s_2$ as *intersections*, since no confusion will result. We call an s_i-equivalence class comprising more than one intersection an *s_i-clique*; we call an intersection which is both an s_1-class and an s_2-class a *loner*; and we use the term *unit* to refer to either an s_1-clique or an s_2-clique or a loner. Thus, the collection of units forms a cover of A_α. Evidently: an s_1- and an s_2-clique have at most one intersection in common; no two different s_i-cliques have any elements (and so intersections) in common; and no s_i-clique includes any loner. If $a \in A$ is $r_i^\#$-related to any element in an intersection, I, then it is $r_i^\#$-related to every element in I: we simply say that a is *$r_i^\#$-related to I*. The following facts are again obvious: if a is $r_i^\#$-related to (any element of) any intersection in an s_i-clique, then a is $r_i^\#$-related to every intersection in that s_i-clique; if distinct units C and C' are s_i-equivalence classes, then a cannot be simultaneously $r_i^\#$-related to an intersection in C and also to an intersection in C'; and a is $r_1^\#$-related to at most one intersection in any s_2-clique, whence there is at least one intersection in that s_2-clique to which a is not $r_1^\#$-related (and similarly with indices exchanged).

To define B_α, select $2(m+2)$ distinct units in $\mathfrak{A}$. (If $\mathfrak{A}$ has fewer units, select them all). Each selected unit C thus contains at most $2(m+2)$ intersections belonging to any other selected unit: select all of these intersections, and, in addition, select $(m+1)$ further intersections in C if possible. (If this is not possible, then C contains fewer than $3m+5$ intersections in total, so select them all). Finally, in any selected intersection I, select up to m elements. (If I contains fewer than m elements, select them all). The set B_α of selected elements in selected intersections in selected units satisfies $|B_\alpha| \leq 2m(m+2)(3m+5)$.

We show that, for every $a \in A$, if a has $\ell \leq m$ distinct free witnesses in A_α, then a is in free position with respect to at least ℓ elements of B_α. Observe first that, if A_α has $2(m+2)$ or more units, then there are $m+2$ selected s_i-cliques or loners for some $i \in \{1,2\}$. Say, $i=1$. Then, fix $a \in A$. At least $m+1$ of these $m+2$ selected units are such that a is not $r_1^\#$-related to them, and at least m of these $m+1$ are not loners to which a is $r_2^\#$-related. Each of these m remaining units therefore contains at least one intersection to which a is in free position. And since distinct s_1-cliques are disjoint, we may choose one element from each, thus obtaining $m \geq \ell$ elements of B_α in free position with respect to a. Henceforth, then, we assume that A_α has fewer than $2(m+2)$ units; and therefore that all units are selected. Again, fix $a \in A$, and suppose first that $a \in A$ has free witnesses in some non-selected intersection. Then that intersection lies in a unit, C, containing at least $m+1$ selected intersections not belonging to any other unit. Without loss of generality, suppose C is an s_1-clique. Then a cannot be $r_1^\#$-related to any intersection in C, and can be $r_2^\#$-related to at most one intersection in C, whence we may find at least m selected intersections in C standing in free position to a. Since distinct intersections are disjoint, we may choose one element from each of these intersections, again obtaining $m \geq \ell$ elements of B_α in free position with respect to a. On the other hand, if all of a's free witnesses lie in selected intersections, then we can obviously replace any non-selected free witness by one of the m selected elements in the same intersection, thus obtaining ℓ elements of B_α in free position with respect to a.

By carrying out this procedure for every 1-type α, we obtain a collection of at most $2m(m+2)(3m+5)|\boldsymbol{\alpha}[\mathfrak{A}]|$ potential free witnesses. Call this set B_1; let B_2 be a set containing the required free witnesses for all elements of B_1; let B_3 be a set containing the required free witnesses for all elements of B_2; and let $B = B_1 \cup B_2 \cup B_3$. Thus, $|B| \leq Z$. We now change the binary predicates of $\mathfrak{A}$ to obtain a structure $\mathfrak{A}'$ as follows. Fix any $a \in A \setminus (B_1 \cup B_2)$. For all i ($1 \leq i \leq m$), if a has an ith free witness, then pick one such witness; and let the (distinct) elements obtained in this way be, in some order, $b_1, \ldots, b_\ell$. Now let $b'_1, \ldots, b'_\ell$ be distinct elements of B_1 in free position with respect to a, with $\mathrm{tp}^{\mathfrak{A}'}[b'_h] = \mathrm{tp}^{\mathfrak{A}}[b_h]$ for all h $(1 \leq h \leq \ell)$. By construction of B_1, this is clearly possible. Now set

$$\mathrm{tp}^{\mathfrak{A}'}[a, b'_h] = \mathrm{tp}^{\mathfrak{A}}[a, b_h] \tag{3.4}$$

for all h $(1 \leq h \leq \ell)$. If $b \in B_1$, then any required free witnesses for b lie in B_2, and so cannot have been disturbed by the re-assignments (3.4) (because $a \notin B_1 \cup B_2$). If $b \in (B_2 \setminus B_1)$, then b cannot be the element a in any instance of (3.4) (because $a \notin B_2$), and equally cannot be the element b_h, (because $b_h \in B_1$). Thus, required witnesses for elements of $B_1 \cup B_2$ are unaffected by the changes in (3.4), and are, by definition in $B_2 \cup B_3 \subseteq B$. That is: in the construction of $\mathfrak{A}'$, all elements of $B_1 \cup B_2$ retain their former i-witnesses in B, while all elements of $B \setminus (B_1 \cup B_2)$ acquire (possibly new) i-witnesses in $B_1 \subseteq B$. Furthermore $\boldsymbol{\beta}[\mathfrak{A}'] \subseteq \boldsymbol{\beta}[\mathfrak{A}]$. It follows that we have $\mathfrak{A}' \models \varphi$, so that $\mathfrak{A}'$ and B are as required. □

Now we can carry out the main task of this section, namely to prove Lemma 3.2.

Proof. [Lemma 3.2] Let φ be as in (3.1), and τ the signature of φ. As before, we write $\psi_{11} = \bigwedge_I \forall x(p_i(x) \rightarrow \exists y(\neg r_1^{\#}(x,y) \wedge \neg r_2^{\#}(x,y) \wedge \theta_i(x,y)))$, where $I = \{1, \ldots, m\}$. We proceed to eliminate the conjuncts of ψ_{11}. Let Z be as in Lemma 3.3, and write $z = \lceil \log(Z+1) \rceil$ (so that z is bounded by a fixed polynomial function of $|\varphi|$). Now take mz new unary predicates $p_{i,1}, \ldots, p_{i,z}$ $(1 \leq i \leq m)$, and a further z unary predicates $q_1, \ldots, q_z$. For all j $(0 \leq j < Z)$, denote by $\bar{p}_{i,j}(x)$ the formula $\neg^{j[1]} p_{i,1}(x) \wedge \cdots \wedge \neg^{j[z]} p_{i,z}(x)$, where $j[h]$ is the hth digit in the z-bit representation of j; define $\bar{q}_j$ similarly, for all j $(0 \leq j \leq Z)$. As an aid to intuition, when $j < Z$, read $\bar{p}_{i,j}(x)$ as "the ith free witness for x is the jth element of a special set" and read $\bar{q}_j(x)$ as "x is the jth element of the special set"; read $\bar{q}_Z(x)$ as "x is not in the special set". The following sentence states that, for all i $(1 \leq i \leq m)$, every element satisfies $\bar{p}_{i,j}(x)$ for some j $(0 \leq j < Z)$:

$$\chi_a = \forall x \bigwedge_{i=1}^{m} \bigvee_{j=0}^{Z-1} p_{i,j}(x).$$

The following sentence states that, for any pair of elements satisfying, respectively, $\bar{p}_{i,j}$ and $\bar{q}_j$, the second is an ith free witness for the first (if such a free witness exists):

$$\chi_b = \forall x \forall y \bigwedge_{i=1}^{m} \bigwedge_{j=0}^{Z-1} ((p_i(x) \wedge p_{i,j}(x) \wedge q_j(y)) \rightarrow (\neg r_1^{\#}(x,y) \wedge \neg r_2^{\#}(x,y) \wedge \theta_i)).$$

Let $\chi' = \chi_a \wedge \chi_b \wedge \chi$. Observe that all quantification in χ' is universal. Finally, the following sentence states that, for all j $(0 \leq j < Z)$, there is an element satisfying $\bar{q}_j(x)$:

$$\omega = \textstyle\bigwedge_{j=0}^{Z-1} \exists x \bar{q}_j(x).$$

Note that $|\chi'|$ and $|\omega|$ are bounded by an exponential function of $|\varphi|$. We claim that φ and $\varphi' = \chi' \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \omega$ are satisfiable over the same domains of cardinality at least Z. On the one hand, φ' evidently entails ψ_{11}, and hence φ. On the other hand, suppose $\mathfrak{A} \models \varphi$, with $|A| \geq Z$. Let $\mathfrak{A}'$ and the set B have the properties guaranteed by Lemma 3.3, and let $\{b_0, \ldots, b_{Z-1}\} \subseteq A'$ include B. We expand $\mathfrak{A}'$ to a structure $\mathfrak{A}''$ interpreting the predicates $p_{i,h}$ and q_h as follows: for all i $(1 \leq i \leq m)$ and $a \in A$, if the ith free witness for a exists and is equal to b_j, ensure $\mathfrak{A}'' \models \bar{p}_{i,j}[a]$; for all j

$(0 \le j \le Z-1)$, ensure $\mathfrak{A}'' \models \bar{q}_j[b_j]$ (note that for this we need $b_0, \ldots, b_{Z-1}$ to be distinct); for all $a \notin \{b_0, \ldots, b_{Z-1}\}$, ensure $\mathfrak{A}'' \models \bar{q}_Z[a]$. It is then easy to see that $\mathfrak{A}'' \models \chi' \wedge \omega$. □

4. Small Intersection Property for EC_2^2**.** In [21] (Proposition 4), it was proved that, for any structure $\mathfrak{A}$ with substructure $\mathfrak{B}$, one may replace $\mathfrak{B}$ by an 'equivalent' structure $\mathfrak{B}'$ of bounded size, in such a way as to preserve certain relations between various parts of $\mathfrak{A}$:

Lemma 4.1. *Let $\mathfrak{A}$ be a structure interpreting a signature of unary and binary predicates, let B be a subset of A such that $\boldsymbol{\alpha}[B] = \{\alpha\}$ for some 1-type α, and let $C = A \setminus B$. Then there is a τ-structure $\mathfrak{A}'$ with domain $A' = B' \dot\cup C$ for some set B' of size bounded by $3|\boldsymbol{\beta}[\mathfrak{A}]|^3$, such that:*

(i) $\mathfrak{A}' \restriction C = \mathfrak{A} \restriction C$;
(ii) $\boldsymbol{\alpha}[B'] = \boldsymbol{\alpha}[B] = \{\alpha\}$, *whence* $\boldsymbol{\alpha}[\mathfrak{A}'] = \boldsymbol{\alpha}[\mathfrak{A}]$;
(iii) $\boldsymbol{\beta}[B'] = \boldsymbol{\beta}[B]$ *and* $\boldsymbol{\beta}[B', C] = \boldsymbol{\beta}[B, C]$, *whence* $\boldsymbol{\beta}[\mathfrak{A}'] = \boldsymbol{\beta}[\mathfrak{A}]$;
(iv) for each $b' \in B'$ *there is some* $b \in B$ *with* $\boldsymbol{\beta}[b', A'] \supseteq \boldsymbol{\beta}[b, A]$;
(v) for each $a \in C$: $\boldsymbol{\beta}[a, B'] \supseteq \boldsymbol{\beta}[a, B]$.
(vi) for each $b' \in B'$ *we have* $\boldsymbol{\beta}[b', B'] = \boldsymbol{\beta}[B]$.

Conditions (i)-(vi) of the above Lemma ensure that any prenex ∀∀- or ∀∃-formula of FO^2 satisfied in $\mathfrak{A}$ is also satisfied in $\mathfrak{A}'$. This result was used in [21] to show that in models of EQ_1^2-sentences equivalence classes can be replaced by classes of bounded size. (Actually, we have modified the published result in [21] slightly: the restriction that the elements of B all have the same 1-type in $\mathfrak{A}$, as well as Condition (vi) and the size-bound on B', were absent from the original. However, these modifications require no change to the original proof.)

It is important to stress that the structures considered in Lemma 4.1 make no special provision regarding the predicates $r_1^\#, r_2^\#, \ldots$. In particular, even if $r_i^\#$ is interpreted as the equivalence closure of r_i in $\mathfrak{A}$, there is no guarantee that this will be so in $\mathfrak{A}'$. The main task of this section is to prove a variant of Lemma 4.1 in which this requirement can be imposed. Since, as we saw at the end of Sec. 2, EQ_2^2-sentences can force models to have infinitely many equivalence classes, and indeed to have infinite equivalence classes, this task is non-trivial.

This variant will be then used to prove the following lemma, were, as usual in this paper, $r_i^\#$ is always required to be interpreted as the transitive closure of r_i:

Lemma 4.2. *Let φ be a satisfiable EC_2^2-sentence in normal or in reduced normal form, over a signature τ. Then there exists a model of φ in which the size of each intersection is bounded by $K(|\tau|)$, for a fixed exponential function K.*

We begin with the advertised variant of Lemma 4.1 allowing us to bound the size of a fragment of an intersection consisting of realizations of a single 1-type.

Lemma 4.3. *Let $\mathfrak{A}$ be a τ-structure, D_1 be an $r_1^\#$-class, D_2 be an $r_2^\#$-class, α be a 1-type, and B be the set of all the elements of 1-type α from the intersection $D_1 \cap D_2$. Then there is a τ-structure $\mathfrak{A}''$ with domain $A'' = B'' \dot\cup C$, where $C = A \setminus B$ and B'' is some set of realizations of α with $|B''| \le 45|\boldsymbol{\beta}[\mathfrak{A}]|^6$, such that:*

(i) $\mathfrak{A}'' \restriction C = \mathfrak{A} \restriction C$;
(ii) $\boldsymbol{\alpha}[B''] = \boldsymbol{\alpha}[B] = \{\alpha\}$, *whence* $\boldsymbol{\alpha}[\mathfrak{A}''] = \boldsymbol{\alpha}[\mathfrak{A}]$;
(iii) $\boldsymbol{\beta}[B''] = \boldsymbol{\beta}[B]$ *and* $\boldsymbol{\beta}[B'', C] = \boldsymbol{\beta}[B, C]$, *whence* $\boldsymbol{\beta}[\mathfrak{A}''] = \boldsymbol{\beta}[\mathfrak{A}]$;
(iv) for each $b'' \in B''$, *there is some* $b \in B$ *with* $\boldsymbol{\beta}[b'', A''] \supseteq \boldsymbol{\beta}[b, A]$;
(v) for each $a \in C$, $\boldsymbol{\beta}[a, B''] \supseteq \boldsymbol{\beta}[a, B]$;
(vi) $B'' \cup (D_1 \setminus B)$ *is an* $r_1^\#$*-class and* $B'' \cup (D_2 \setminus B)$ *an* $r_2^\#$*-class.*

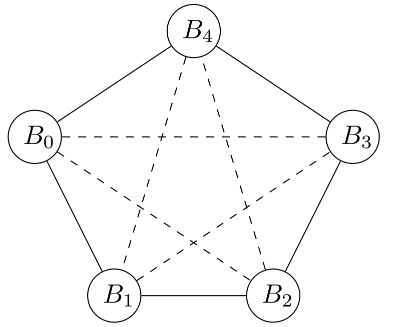

Figure 4.1: Making B'' r_1- and r_2-connected. A solid (dashed) line between B_i and B_j means that each element from B_i has an r_1-edge (r_2-edge) to each element from B_j.

Proof. If $|B| \leq 1$, then we simply put $B'' = B$ and we are done. Otherwise, our first step is a simple application of Lemma 4.1. Let p_1, p_2 be fresh unary predicates. Let $\bar{\mathfrak{A}}$ be the expansion of $\mathfrak{A}$ obtained by setting p_1, p_2 true for all elements of D_1, resp. D_2. Let the result of the application of Lemma 4.1 to $\bar{\mathfrak{A}}$ and the substructure induced by B be a structure $\bar{\mathfrak{A}}'$, in which B' is the replacement of B. By $\mathfrak{A}'$ we denote the restriction of $\bar{\mathfrak{A}}'$ to the original signature, i.e. the structure obtained from $\bar{\mathfrak{A}}'$ by dropping the interpretations of p_1 and p_2. Thus, $\mathfrak{A}'$ is a structure with domain $C \cup B'$ and $|B'|$ is exponentially bounded in the signature.

Let $D_i' = B' \cup (D_i \setminus B)$ $(i = 1, 2)$. By the second equality from part (iii) of Lemma 4.1 and by our strategy of marking elements of D_i with the auxiliary predicate p_i, it follows that any pair of elements from D_i' is joined by $r_i^{\#}$. However it is not guaranteed that D_i' is r_i-connected, and we need to repair this defect. To do so, we employ an additional combinatorial construction, yielding a structure $\mathfrak{A}''$ whose domain is $C \dot{\cup} B''$. The restrictions of the structures $\mathfrak{A}$, $\mathfrak{A}'$, and $\mathfrak{A}''$ to C are equal. We denote $D_i'' = B'' \cup (D_i \setminus B)$ $(i = 1, 2)$. The main goal of the construction of $\mathfrak{A}''$ is to make B'' r_1- and r_2-connected, which, due to part (v) of Lemma 4.1, will also make D_1'' r_1-connected, and D_2'' r_2-connected. We consider three cases. We first present the constructions required in all cases and after that we prove correctness of each of them.

Case 1: There is a pair of distinct elements $s, t \in B$ such that $\mathfrak{A} \models r_1[s, t]$, and there is a pair of distinct elements $u, w \in B$ such that $\mathfrak{A} \models r_2[u, w]$.

We build B'' from five pairwise disjoint sets $B_0, \ldots, B_4$. In $\mathfrak{A}''$, we define the substructures $\mathfrak{B}_i$ as copies of $\mathfrak{B}'$, and we make the substructures induced by $C \cup B_i$ isomorphic to $\mathfrak{A}'$. It remains to set the connections (i.e. 2-types) among the $\mathfrak{B}_i$'s. For $i = 0, \ldots, 4$, and for every pair of elements $b_1 \in B_i$, $b_2 \in B_{(i+1) \bmod 5}$, set $\mathrm{tp}^{\mathfrak{A}''}(b_1, b_2) := \mathrm{tp}^{\mathfrak{A}}(s, t)$. For every pair of elements $b_1 \in B_i$, $b_2 \in B_{(i+2) \bmod 5}$, set $\mathrm{tp}^{\mathfrak{A}''}(b_1, b_2) := \mathrm{tp}^{\mathfrak{A}}(u, w)$. See Fig. 4.1. Note that this fully defines $\mathfrak{A}''$.

Case 2: For every pair of distinct elements $s, t \in B$ we have $\mathfrak{A} \models \neg r_1[s, t] \wedge \neg r_2[s, t]$.

Let $S_i^1, \ldots, S_i^{k_i}$ $(i = 1, 2)$, be the partition of D_i' in $\mathfrak{A}'$ into maximal r_i-connected subsets. Let us first observe that each S_i^k contains at least one element from B'. Indeed, $S_i^k \setminus B'$ is a subset of D_i, from which there are no r_i-edges to $D_i \setminus (B \cup (S_i^k \setminus B'))$ in $\mathfrak{A}$, since otherwise, such an edge would be retained in $\mathfrak{A}'$ and S_i^k would not be

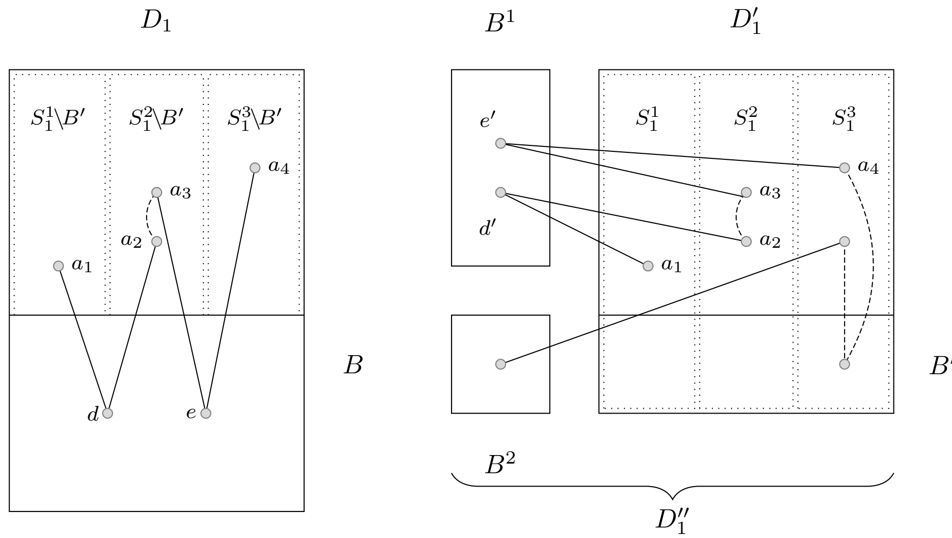

Figure 4.2: Making D_1'' r_1-connected in Case 2, by means of B^1. Note that $D_1' \setminus B' = D_1 \setminus B$. Solid lines represent direct r_1-connections, dashed lines represent r_1-paths. Elements a_1 and a_4 are not necessarily r_1-connected in $\mathfrak{A}'$ but they become r_1-connected in $\mathfrak{A}''$ by a path going through d' and e'.

maximal. Thus, since D_i is r_i-connected in $\mathfrak{A}$, there must be an element $a \in S_i^k \setminus B'$, with an r_i-edge to some $b \in B$ in $\mathfrak{A}$. Now, property (v) of Lemma 4.1 guarantees that there exists $b' \in B'$ with $\mathrm{tp}^{\mathfrak{A}'}(a, b') = \mathrm{tp}^{\mathfrak{A}}(a, b)$, so b' has an r_i-edge to a, and thus $b' \in S_i^k$. This observation implies that the number of maximal r_i-connected subsets of D_i' in $\mathfrak{A}'$ is bounded by $|B'|$, i.e. exponentially in the signature $(i = 1, 2)$.

We build B'' from B' and two sets B^1 and B^2 containing new elements of type α constructed as described below. We define $\mathfrak{A}''\restriction C \cup B'$ to be equal to $\mathfrak{A}'$. We say that S_i^k and S_i^l are *connected by B through an element $d \in B$* in $\mathfrak{A}$ if and only if there are $a_1 \in S_i^k \setminus B'$, $a_2 \in S_i^l \setminus B'$, such that a_1, d, a_2 is an r_i-path in $\mathfrak{A}$ (see D_1 in Fig. 4.2). For S_i^k and S_i^l connected by B through some element, we choose one such connecting element d and add a fresh element d' to B^i. For every $c \in C$, we set $\mathrm{tp}^{\mathfrak{A}''}(d', c) := \mathrm{tp}^{\mathfrak{A}}(d, c)$. The 2-types between d' and B' are set in such a way that $\boldsymbol{\beta}[d, B] = \boldsymbol{\beta}[d', B']$; by part (vi) of Lemma 4.1 we always have enough elements in B' to secure this property (recall also that B contains at least two realizations of α, so we have some patterns which can be used for setting the connections between d' and B'). The 2-types inside $B^1 \cup B^2$ are set as arbitrary 2-types used in $\mathfrak{B}$.

Case 3: There exists a pair of distinct elements $s, t \in B$ such that $\mathfrak{A} \models r_1[s, t]$, but for all pairs of distinct elements $u, v \in B$, we have $\mathfrak{A} \models \neg r_2[u, v]$. (Or symmetrically, exchanging r_1 and r_2.)

This construction is a combination of the previous two. We build B'' from three disjoint sets B_0, B_1, B^2 of realizations of α. The role of the sets B_0 and B_1 is similar to the role of the sets $B_0, \ldots, B_4$ from Case 1, while the role of B^2 is similar to the role of B^2 from Case 2.

In $\mathfrak{A}''$ we define the substructures $\mathfrak{B}_i$ as copies of $\mathfrak{B}'$ and we make the substructures induced by $C \cup B_i$ $(i = 0, 1)$ isomorphic to $\mathfrak{A}'$. For every pair of elements

$b_1 \in B_0$, $b_2 \in B_1$ we set $\mathrm{tp}^{\mathfrak{A}''}(b_1, b_2) := \mathrm{tp}^{\mathfrak{A}}(s,t)$.

Let $S_2^1, \ldots, S_2^{k_2}$ be the partition of D_2' in $\mathfrak{A}'$ into maximal r_2-connected subsets. As in Case 2, each S_2^k contains at least one element from B'. This implies that the number of r_2-connected subsets of D_2' is again bounded by $|B'|$. Recall that S_2^k and S_2^l are connected by B through $d \in B$ if there are $a_1 \in S_i^k \setminus B'$, $a_2 \in S_2^l \setminus B'$ such that a_1, d, a_2 is an r_2-path in $\mathfrak{A}$. If S_2^k and S_2^l are connected by B through some element, we choose one such connecting element $d \in B$, and add a fresh element d' to B^2. For every $c \in C$, we set $\mathrm{tp}^{\mathfrak{A}''}(d', c) := \mathrm{tp}^{\mathfrak{A}}(d, c)$. The 2-types between d' and B_i $(i = 0, 1)$ are set in such a way that $\boldsymbol{\beta}[d, B] = \boldsymbol{\beta}[d', B_i]$. The 2-types inside B^2 are set as arbitrary 2-types used in $\mathfrak{B}$.

Finally, for every pair of elements $b_1 \in B^2$, $b_2 \in B_0 \cup B_1$ we set $\mathrm{tp}^{\mathfrak{A}''}(b_1, b_2) := \mathrm{tp}^{\mathfrak{A}}(s,t)$. This makes B^2 r_1-connected to the remaining part of D_1''.

Now we argue that $\mathfrak{A}''$ and B'' are as required. It should be clear that properties (i)-(v) are fulfilled and that the size of B'' is not greater than $5|B'|^2$, which, by the bound on B' from Lemma 4.1 is not greater than $45|\boldsymbol{\beta}[\mathfrak{A}]|^6$. Now we show that property (vi) also holds.

Case 1: First, note that our strategy of connecting B_i-s ensures that $B'' = B_0 \cup \ldots \cup B_4$ is both r_1- and r_2-connected. We show now that, for any i and $a \in D_i \setminus B$ $(= D_i'' \setminus B'' = D_i' \setminus B')$ there is an r_i-path in $\mathfrak{A}''$ between a and some element $b'' \in B''$. As D_i is r_i-connected there must be a path in $\mathfrak{A}$ from a to some $b \in B$. Let $a = a_0, \ldots, a_k = b$ be such a path, with $a_j \notin B$ for all $j < k$. Obviously, a_0 and a_{k-1} are r_i-connected in $\mathfrak{A}''$ as both are members of C, and the structure of C is copied to $\mathfrak{A}''$. We show that a_{k-1} is connected to some element in B''. Indeed, property (v) of Lemma 4.1 guarantees that there is an r_i-edge between a_{k-1} and some element b' of 1-type $\alpha \cup \{p_1(x), p_2(x)\}$ in $\bar{\mathfrak{A}}'$, and property (i) of the same lemma guarantees that there are no such elements outside B'. By our construction, in $\mathfrak{A}''$ there is also an edge between a_{k-1} and b'' - the copy of b' in B_0. Therefore, D_i'' is r_i-connected for $i \in \{1, 2\}$. By property (iii) of Lemma 4.1, there are no r_i-connections from B' to elements that do not satisfy p_i (i.e. elements from $C \setminus D_i$), and therefore D_i'' is a maximal r_i-connected set.

Case 2: Recall that $D_i'' = B'' \cup (D_i \setminus B)$ and $B'' = B' \cup B^1 \cup B^2$, so $D_i'' = (B' \cup (D_i \setminus B)) \cup B^1 \cup B^2 = D_i' \cup B^1 \cup B^2$. Let us first observe that D_i' is r_i-connected $(i = 1, 2)$ in $\mathfrak{A}''$. If $a, b \in S_i^l$ for some l then a, b are r_i-connected by the definition of S_i^l. If $a \in S_i^l$, $b \in S_i^k$ and S_i^l, S_i^k are connected by B through some d then, by our construction, there is an r_i-path a', d', b' for some $a' \in S_i^l$, $b' \in S_i^k$ and $d' \in B^i$. This path can be extended by a path from a to a' and a path from b' to b. Thus a and b are r_i-connected in $\mathfrak{A}''$. This argument can be inductively extended to cover the case of arbitrary a, b: without loss of generality, we assume that $a, b \notin B'$ (since any element from B' must have an r_i-edge to $D_i' \setminus B'$ by part (iv) of Lemma 4.1, D_i is r_i-connected, and there are no r_i-edges inside B). In $\mathfrak{A}$ there is an r_i-path from a to b. This path can be split into fragments consisting of elements belonging to some $S_i^k \setminus B$ and a single element from B (with the exception of the last fragment which does not contain an element from B). The S_i^k-s which are neighbours in this path are thus connected by B. This guarantees an R_i-path from a to b in $\mathfrak{A}''$. The set B^i is r_i-connected to D_i' since, by our construction, any element from B^i has r_i-edges to at least two elements from D_i'. It remains to show that B^2 is r_1-connected to the remaining part of D_1'', and, symmetrically that B^1 is r_2-connected to the remaining part of D_2''. Consider the case of B^2 and r_1-connections. Let b'' be an element from B^2. The element b'' was added to B^2 as a copy of some element b from B. In particular its connections to $D'_1 \setminus B'$ in $\mathfrak{A}''$ were copied from $\mathfrak{A}$ (recall that $D'_2 \setminus B'' = D'_1 \setminus B' = D_1 \setminus B$). As there are no r_1-edges inside B, and B is r_1-connected, there must be an edge from b to some element of $D_1 \setminus B$ in $\mathfrak{A}$. Thus there is an r_1-edge from b'' to $D'_1 \setminus B'$ in $\mathfrak{A}''$. Analogously for B^1 and r_2-connections.

Case 3: Here the proof is a combination of the arguments from the two previous cases. Consider the case in which B contains an r_1-edge but has no r_2-edges (the symmetric case can be treated analogously). First, note that our strategy of connecting elements ensures that $B_0 \cup B_1$ is r_1-connected. Exactly as in Case 1 we can show that any element of $D'_1 \setminus B'$ is r_1-connected to B_0. The final step of our construction ensures that also B^2 is r_1-connected to B_0. This shows that D''_1 is r_1-connected. The argument that D''_2 is r_2-connected goes as in Case 2: first see that $(D'_2 \setminus B') \cup B^2$ is r_2-connected, and then note that every element from $B_0 \cup B_1$ must have an r_2-edge to the remaining part of D''_2. □

Now we are ready to prove Lemma 4.2.

Proof. [Lemma 4.2] We first argue that the structure obtained as an application of Lemma 4.3 satisfies the same normal form formulas over τ as the original structure. Let $\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \psi_{11}$ be a formula in normal form over τ, as in (3.1). Supposing φ to be satisfiable, let $\mathfrak{A} \models \varphi$, let $B \subseteq A$ be a maximal set that is r_1- and r_2-connected and such that $\boldsymbol{\alpha}[B] = \{\alpha\}$ is a singleton set, let D_i be the $r_i^\#$-class including B ($i = 1, 2$), let $C = A \setminus B$, and let $\mathfrak{A}''$ be the structure (with domain $A'' = B'' \dot\cup C$) obtained by applying Lemma 4.3.

Formula χ is satisfied in $\mathfrak{A}''$ thanks to property (iii) of Lemma 4.3. For any $c \in C$, properties (i) and (v) guarantee that c has all required witnesses. For any $b \in B''$, the same thing is guaranteed by property (iv).

Now, to find a small replacement of a whole intersection, we apply Lemma 4.3 iteratively to all 1-types realized in this intersection. Property (vi) guarantees that the obtained substructure is a maximal r_1- and r_2-connected set, so indeed it is an intersection in the new model.

The proof of the Löwenheim-Skolem theorem (every satisfiable formula is satisfiable in a countable model) can easily be extended to EC^2; thus we may restrict our attention to countable structures. Let $I_1, I_2, \ldots$ be a (possibly infinite) sequence of all intersections in a $\mathfrak{A}$, let $\mathfrak{A}_0 = \mathfrak{A}$, and let $\mathfrak{A}_{j+1}$ be the structure $\mathfrak{A}_j$ modified by replacing intersection I_{j+1} by its small replacement I'_{j+1} as described above. We define the limit structure $\mathfrak{A}_\infty$ with the domain $I'_1 \cup I'_2, \ldots$ such that for all $k < l$ the connections (i.e. 2-types) between I'_k and I'_l are defined in the same way as in $\mathfrak{A}_l$. It is easy to see that $\mathfrak{A}_\infty$ satisfies φ and all intersections in $\mathfrak{A}_\infty$ are bounded exponentially in $|\tau|$.

The described construction works also for formulas in reduced normal form because the conjunct ω is satisfied due to property (ii) of Lemma 4.3. □

A Note on EC^2_1. We can now easily get the following *exponential classes property* for EC^2_1.

LEMMA 4.4. *Let φ be a satisfiable (reduced) normal form EC^2_1 formula. Then φ is satisfiable in a model in which all $r_1^\#$-classes are bounded exponentially.*

Proof. Consider the EC^2_2 formula $\varphi' = \varphi \wedge \forall x \forall y . r_2(x, y)$. Clearly, it is satisfiable (take a model of φ and interpret r_2 as the total relation). We apply Lemma 4.2 to φ' obtaining a structure $\mathfrak{A}'$ with small intersections. After dropping the interpretation of r_2 in $\mathfrak{A}'$ we get a structure $\mathfrak{A}$ which is a model of φ. It has appropriately bounded $r_1^{\#}$-classes as they correspond to intersections of $\mathfrak{A}'$. □

Lemma 4.4 generalizes the small classes property for FO^2 with one equivalence relation from [21]. We can now repeat the construction from [21] (p. 738, 4.1.2. *Few classes*) to show:

THEOREM 4.5. *Let φ be a satisfiable EC_1^2 formula. Then φ is satisfiable in a model of at most exponential size. Thus the satisfiability problem (= finite satisfiability problem) is* NEXPTIME-*complete.*

5. The Graph Existence Problem. Let $\mathfrak{A}$ be any countable EC_2^2-structure over some fixed signature, all of whose intersections are subject to some fixed size bound. Then there is a finite collection Δ of isomorphism types of intersections that $\mathfrak{A}$ can possibly realize. Now let U be the set of $r_1^{\#}$-classes occurring in $\mathfrak{A}$, and V, the set of $r_2^{\#}$-classes. Of course, each $r_1^{\#}$-class $u \in U$ is a union of intersections, and similarly for each $r_2^{\#}$-class $v \in V$. As we observed in the proof of Lemma 3.3, $\mathfrak{A}$ may contain 'loners'—that is, intersections which are both $r_1^{\#}$-classes and $r_2^{\#}$-classes, and which are thus elements of both U and V. Since, in the sequel, we shall want to regard U and V as disjoint sets, we count loners twice: once as an element of U and once as an element of V. (Technically, we need to create isomorphic copies of intersections to represent the elements of V; however, to avoid presentational clutter, we continue to speak of elements of V as intersections from $\mathfrak{A}$ without qualification.) Now we may construct a (possibly infinite) bipartite graph on the vertex sets U and V by taking (u, v) to be an edge just in case u and v share some intersection. In fact, since any $r_1^{\#}$-class $u \in U$ may share at most one intersection with any $r_2^{\#}$-class $v \in V$, we may take the edge (u, v) to be *coloured* by the isomorphism type of the intersection in question, i.e. by some colour $\delta \in \Delta$. In this section, we define two problems concerning bipartite graphs with coloured edges, and show (Theorem 5.10) that they are NPTIME-complete. We use this fact in Sec. 6 to establish our upper complexity bounds for EC_2^2.

We make extensive use of results on linear programming and integer programming. A *linear equation* (*inequality*) is always an expression $t_1 = t_2$ ($t_1 \geq t_2$) where t_1 and t_2 are linear terms with coefficients in $\mathbb{N}$. Given a system $\mathcal{E}$ of linear equations and inequalities, we take the *size* of $\mathcal{E}$, denoted $\|\mathcal{E}\|$, to be the total number of bits required to write $\mathcal{E}$ in standard notation; notice that $\|\mathcal{E}\|$ may be much larger than $|\mathcal{E}|$, the number of equations and inequalities in $\mathcal{E}$. The problem *linear programming* is as follows:

GIVEN: a system $\mathcal{E}$ of linear equations and inequalities.
OUTPUT: Yes, if $\mathcal{E}$ has a solution over $\mathbb{Q}$; No, otherwise.

The problem *integer programming* is as follows:

GIVEN: a system $\mathcal{E}$ of linear equations and inequalities.
OUTPUT: Yes, if $\mathcal{E}$ has a solution over $\mathbb{N}$; No otherwise.

Denote by $\mathbb{N}^*$ the set $\mathbb{N} \cup \{\aleph_0\}$. We interpret the arithmetic operations $+$ and $\cdot$ as well as the ordering $<$ over $\mathbb{N}^*$ as expected. Specifically: $\aleph_0 + n = \aleph_0 + \aleph_0 = \aleph_0$ for all $n \in \mathbb{N}$; $\aleph_0 \cdot 0 = 0$, and $\aleph_0 \cdot m = \aleph_0 \cdot \aleph_0 = \aleph_0$ for all non-zero $m \in \mathbb{N}$; and $n < \aleph_0$ for all $n \in \mathbb{N}$. The problem *extended integer programming* is as follows:

Given: a system $\mathcal{E}$ of linear equations and inequalities.
Output: Yes, if $\mathcal{E}$ has a solution over $\mathbb{N}^*$; No, otherwise.

Thus, for example, the system $\mathcal{E}$ given by

$$x_1 \geq x_2 + 1 \qquad\qquad x_2 \geq x_1 + 1$$

has no solution over $\mathbb{N}$—or indeed over $\mathbb{Q}$—but does have a solution over $\mathbb{N}^*$, namely $x_1 = x_2 = \aleph_0$. Observe that the coefficients in $\mathcal{E}$ are, in all cases, required to be in $\mathbb{N}$.

The following results on linear and integer programming are well-known.

Proposition 5.1 ([14], Theorem 1). *The problem linear programming is in* PTime.

Proposition 5.2 ([5], Theorem 1). *Let $\mathcal{E}$ be a system of linear equations and inequalities with coefficients in $\mathbb{N}$, and let $k > 0$. If each coefficient in $\mathcal{E}$ has at most k bits, and $\mathcal{E}$ has a solution over $\mathbb{N}$, then it has a solution over $\mathbb{N}$ in which the number of non-zero values is bounded by $p(k|\mathcal{E}|)$, where p is a fixed polynomial.*

Proposition 5.3 ([3], Theorem 2). *Let $\mathcal{E}$ be a system of linear equations and inequalities with coefficients in $\mathbb{N}$. If $\mathcal{E}$ has a solution over $\mathbb{N}$, then it has a solution over $\mathbb{N}$ in which all values are bounded by $2^{p(\|\mathcal{E}\|)}$, where p is a fixed polynomial. Hence, integer programming is in* NPTime.

Proposition 5.2 is a *Carathéodory-type* result for integer programming: if an integer vector is in the positive integral cone of some large set of integer vectors, then it is in the positive integral cone of a small subset of them. We may extend both Proposition 5.2 and Proposition 5.3 to solutions over $\mathbb{N}^*$ in the obvious way:

Corollary 5.4. *Let $\mathcal{E}$ be a system of linear equations and inequalities with coefficients in $\mathbb{N}$, and let $k > 0$. If each coefficient in $\mathcal{E}$ has at most k bits, and $\mathcal{E}$ has a solution over $\mathbb{N}^*$, then it has a solution over $\mathbb{N}^*$ in which the number of non-zero values is bounded by $p(k|\mathcal{E}|)$, where p is a fixed polynomial.*

Proof. Fix some solution $\bar{a}$ over $\mathbb{N}^*$, let $\mathcal{E}'$ be the collection of all equations and inequalities in $\mathcal{E}$ whose left- and right-hand sides are finite under this solution, and let $\mathcal{E}'' = \mathcal{E} \setminus \mathcal{E}'$. Thus, ignoring terms with zero-coefficients, $\mathcal{E}'$ features no variables whose value in $\bar{a}$ is infinite. Choose a solution $\bar{b}$ of $\mathcal{E}'$ over $\mathbb{N}$ with at most $p'(k|\mathcal{E}'|)$ non-zero values, where p' is the polynomial guaranteed by Proposition 5.2. Now choose, for each element of $\mathcal{E}''$, at most two variables such that making these infinite is sufficient to render the left- or right-hand sides infinite, as determined by $\bar{a}$. Make all other variables zero. We thus obtain a solution with at most $p'(k|\mathcal{E}'|) + 2|\mathcal{E}''|$ non-zero values. □

Corollary 5.5. *Let $\mathcal{E}$ be a system of linear equations and inequalities with coefficients in $\mathbb{N}$. If $\mathcal{E}$ has a solution over $\mathbb{N}^*$, then it has a solution over $\mathbb{N}$ in which all finite values are bounded by $2^{p(\|\mathcal{E}\|)}$, where p is a fixed polynomial. Hence, extended integer programming is in* NPTime.

Proof. Similar to proof of Corollary 5.4. □

5.1. Bipartite graph existence. Let Δ be a finite, non-empty set. A Δ-*graph* is a triple $H = (U, V, \mathbf{E}_\Delta)$, where U, V are disjoint, countable (possibly finite, or even empty) sets, and $\mathbf{E}_\Delta$ is a collection of pairwise disjoint subsets $E_\delta \subseteq U \times V$, indexed by the elements of Δ. We call the elements of $W = U \cup V$ *vertices*, and the elements of E_δ, δ-*edges*; and we say that H is *finite* if $U \cup V$ is finite. It helps to think of $\mathbf{E}_\Delta$ as the result of colouring the edges of the bipartite graph (U, V, E), where $E = \bigcup_{\delta \in \Delta} E_\delta$

is a set of edges from U to V, using the colours in Δ. For any $w \in W$, we define the function $\mathrm{ord}_w^H : \Delta \to \mathbb{N}^*$, called the *order* of w, by

$$\mathrm{ord}_u^H(\delta) = |\{v \in W : (u, v) \in E_\delta\} \qquad (u \in U)$$
$$\mathrm{ord}_v^H(\delta) = |\{u \in W : (u, v) \in E_\delta\} \qquad (v \in V).$$

Thus, ord_w^H tells us, for each colour δ, how many δ-edges w is incident to in H. We now proceed to define the problem BGE ("bipartite graph existence"). A *BGE-instance* is a quadruple $\mathcal{P} = (\Delta, \Delta_0, F, G)$, where Δ is a finite, non-empty set, $\Delta_0 \subseteq \Delta$, and F and G are sets of functions $\Delta \to \mathbb{N}$. A *solution* of $\mathcal{P}$ is a Δ-graph $H = (U, V, \mathbf{E}_\Delta)$ such that:

$$\text{for all } \delta \in \Delta_0,\ E_\delta \text{ is non-empty;} \tag{G1}$$
$$\text{for all } u \in U,\ \mathrm{ord}_u^H \in F; \tag{G2}$$
$$\text{for all } v \in V,\ \mathrm{ord}_v^H \in G. \tag{G3}$$

The problem *BGE* is as follows:

GIVEN: a BGE-instance $\mathcal{P}$.
OUTPUT: Yes, if $\mathcal{P}$ has a solution; No, otherwise.

The problem *finite BGE* is as follows:

GIVEN: a BGE-instance $\mathcal{P}$.
OUTPUT: Yes, if $\mathcal{P}$ has a finite solution; No, otherwise.

That is: given $\Delta_0 \subseteq \Delta$ and sets of order-functions F, G over Δ, we wish to know whether there exists a (finite) Δ-graph $(U, V, \mathbf{E}_\Delta)$ in which the vertices in U realize only those order-functions in F, the vertices in V realize only those order-functions in G, and each of the colours in Δ_0 is represented by at least one edge. Notice that, even though the bipartite graphs in question may be infinite, the orders in F and G are assumed to have finite values.

Before proceeding, we obtain a simple complexity bound for BGE. This result illustrates the basic approach taken in the sequel, while avoiding much of the distracting detail.

LEMMA 5.6. *Let F, G be finite sets of functions $\Delta \to \mathbb{N}$, and suppose there exist natural numbers x_f (for all $f \in F$) and y_g (for all $g \in G$) such that, for all $\delta \in \Delta$,*

$$\sum_{f \in F} f(\delta) \cdot x_f = \sum_{g \in G} g(\delta) \cdot y_g.$$

Then there exists a finite Δ-graph $(U, V, \mathbf{E}_\Delta)$ and a positive integer k such that: (i) for all functions $f : \Delta \to \mathbb{N}$, the number of vertices in U with order f is given by

$$\begin{cases} k \cdot x_f & \text{if } f \in F \\ 0 & \text{otherwise}; \end{cases}$$

(ii) for all functions $g : \Delta \to \mathbb{N}$, the number of vertices in V with order g is given by

$$\begin{cases} k \cdot y_g & \text{if } g \in G \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We proceed by induction on the quantity

$$Q = \sum_{\delta \in \Delta} \sum_{f \in F} f(\delta) \cdot x_f = \sum_{\delta \in \Delta} \sum_{g \in G} g(\delta) \cdot y_g.$$

Suppose first $Q = 0$. Denoting by $\mathbf{0}$ the function uniformly mapping every element of Δ to 0, and bearing in mind that $Q = 0$, we see that $f \in F$ and $x_f > 0$ implies $f = \mathbf{0}$; and similarly, $g \in G$ and $y_g > 0$ implies $g = \mathbf{0}$. If $\mathbf{0} \notin F$, define $x_{\mathbf{0}} = 0$; and if $\mathbf{0} \notin G$, define $y_{\mathbf{0}} = 0$. Let U, V be disjoint sets of cardinalities $x_{\mathbf{0}}$ and $y_{\mathbf{0}}$, respectively; and set $E_\delta = \emptyset$ for all $\delta \in \Delta$. Thus, in the Δ-graph $H = (U, V, \mathbf{E}_\Delta)$, every vertex has order $\mathbf{0}$. It is then immediate that H and $k = 1$ satisfy the requirements of the lemma.

Suppose, now $Q > 0$. Thus, $x_f > 0$, $f(\delta) > 0$, $y_g > 0$ and $g(\delta) > 0$ for some $f \in F$, $g \in G$ and $\delta \in \Delta$. Let f_0, g_0 and δ_0 be such any values. We may think each number x_f as giving the multiplicity of f in a multiset of functions $\Delta \to \mathbb{N}$; and similarly for the numbers y_g. We proceed by taking one instance of f_0, and decrementing its value at δ_0; likewise, we take one instance of g_0, and decrement its value at δ_0. Formally, define

$$f'(\delta) = \begin{cases} f_0(\delta) - 1 & \text{if } \delta = \delta_0 \\ f_0(\delta) & \text{otherwise.} \end{cases} \qquad g'(\delta) = \begin{cases} g_0(\delta) - 1 & \text{if } \delta = \delta_0 \\ g_0(\delta) & \text{otherwise.} \end{cases}$$

If $f' \notin F$, set $x_{f'} = 0$, and if $g' \notin G$, set $y_{g'} = 0$. Let $F' = F \cup \{f'\}$ and $G' = G \cup \{g'\}$. Now let

$$x'_f = \begin{cases} x_f - 1 & \text{if } f = f_0 \\ x_f + 1 & \text{if } f = f' \\ x_f & \text{otherwise} \end{cases} \qquad y'_g = \begin{cases} y_g - 1 & \text{if } g = g_0 \\ y_g + 1 & \text{if } g = g' \\ y_g & \text{otherwise} \end{cases}$$

Thus,

$$\sum_{f \in F'} f(\delta_0) \cdot x'_f = \sum_{f \in F} f(\delta_0) \cdot x'_f - 1$$
$$\sum_{g \in G'} g(\delta_0) \cdot y'_f = \sum_{g \in G} g(\delta_0) \cdot y'_f - 1.$$

Since we have merely decremented the value of one instance of f_0 at the point δ_0, and done the same for one instance of g_0, it is obvious that, for all $\delta \in \Delta$,

$$\sum_{f \in F'} f(\delta) \cdot x'_f = \sum_{g \in G'} g(\delta) \cdot y'_f;$$

and, moreover, $\sum_{\delta \in \Delta} \sum_{f \in F'} f(\delta) \cdot x'_f = Q - 1$.

By inductive hypothesis, let the finite Δ-graph $H' = (U', V', \mathbf{E}'_\Delta)$ and the positive integer k' satisfy the lemma for the sets of functions F' and G', and the various natural numbers x'_f and y'_g. Note that U' contains $k' \cdot x'_{f'} \geq k'$ vertices having order f'; let U'_0 be a subset of these with cardinality k'. Similarly, V' contains $k' \cdot y'_{g'} \geq k'$ vertices having order g'; let V'_0 be a subset of these with cardinality k'. Take an isomorphic copy $H'' = (U'', V'', \mathbf{E}''_\Delta)$ of H', and let U''_0 and V''_0 be the copies of U'_0 and V'_0 under this isomorphism. Let $H^* = (U, V, \mathbf{E}^*_\Delta)$ be the disjoint union of H' and H'' (i.e. $U = U' \cup U''$, $V = V' \cup V''$ and $E^*_\delta = E'_\delta \cup E''_\delta$ for all $\delta \in \Delta$). Finally, let H be obtained from H^* by adding δ_0-coloured edges so as to pair up the the vertices of U_0' and V_0'', and by adding δ_0-coloured edges so as to pair up the the vertices of U_0'' and V_0'. Note that these edges cannot occur in H^*. For all $u \in U_0' \cup U_0''$, we have $\text{ord}_u^{H^*} = f'$ and $\text{ord}_u^H = f_0$; similarly, for all $v \in V_0' \cup V_0''$, we have $\text{ord}_v^{H^*} = g'$ and $\text{ord}_g^H = g_0$. Let $k = 2k'$. Continuing to write $x_{f'} = 0$ if $f' \notin F$, consider any $f \in F'$. By inductive hypothesis, there are exactly $k' \cdot x_f'$ vertices $u \in U'$ such that $\text{ord}_u^{H'} = f$. Now let us calculate the number of vertices $u \in U$ such that $\text{ord}_u^H = f$.

For $f = f_0$, we must count all the vertices having order f_0 in H' and H'' together with all the vertices of U_0' and U_0''. This yields $2k' \cdot x_{f_0}' + 2k' = k \cdot x_{f_0}$ vertices.

For $f = f'$, we must count all the vertices having order f' in H' and H'', but ignoring the vertices of U_0' and U_0''. This yields $2k' \cdot x_{f'}' - 2k' = k \cdot x_{f'}$ vertices.

For all other $f \in F'$, we simply count the number of vertices of U' and U'' together having order f. This yields $2k' \cdot x_f' = k \cdot x_f$ vertices.

Thus, for all $f \in F$, the number of vertices in $u \in U$ such that $\text{ord}_u^H = f$ is $k \cdot x_f$ as required. A similar argument establishes the symmetric condition for the vertices in V. □

Proposition 5.7. *The problems BGE and finite BGE are in* PTime.

Proof. We reduce finite BGE to linear programming. Consider any BGE-instance $\mathcal{P} = (\Delta, \Delta_0, F, G)$. We claim that $\mathcal{P}$ has a finite solution if and only if the system of equations and inequalities

$$\sum_{f \in F} f(\delta) \cdot x_f = \sum_{g \in G} g(\delta) \cdot y_f \qquad (\delta \in \Delta) \qquad (5.1)$$

$$\sum_{f \in F} f(\delta) \cdot x_f > 0 \qquad (\delta \in \Delta_0) \qquad (5.2)$$

involving the variables $\{x_f\}_{f \in F}$ and $\{x_g\}_{g \in G}$, has a solution over $\mathbb{N}$. For the only-if direction, suppose $(U, V, \mathbf{E}_\Delta)$ is a finite solution of $\mathcal{P}$. For all $f \in F$, let x_f be the number of elements of U having order f; and for all $g \in G$, let y_g be the number of elements of V having order g. Then the number of δ-coloured edges is given by both the right- and left-hand side of (5.1), thus securing (5.1) and (5.2). The if-direction follows from Lemma 5.6. Evidently, if the system (5.1) and (5.2) has a solution over the non-negative rationals, then it has a solution over $\mathbb{N}$, and vice versa. The theorem then follows from Proposition 5.1.

For the general (non-finite) case, we reduce BGE to the satisfiability problem for propositional Horn clauses. (One might try to solve the above equation system over $\mathbb{N}^*$ but satisfiability over $\mathbb{Q}^*$ is not obviously in PTime.) For $f \in F$, let X_f be a proposition letter, which we may informally read as "There are no vertices in U having order f. Similarly, for $g \in G$, let Y_g be a proposition letter, which we may informally read as "There are no vertices in V having order g." Consider the set Γ of propositional Horn-clauses

$$\left\{\left(\bigwedge_{g\in G: g(\delta)>0} Y_g\right) \to X_f \mid f \in F, \delta \in \Delta \text{ s.t. } f(\delta) > 0\right\} \tag{5.3}$$

$$\left\{\left(\bigwedge_{f\in F: f(\delta)>0} X_f\right) \to Y_g \mid g \in G, \delta \in \Delta \text{ s.t. } g(\delta) > 0\right\} \tag{5.4}$$

$$\left\{\left(\bigwedge_{f\in F: f(\delta)>0} X_f\right) \to \bot \mid \delta \in \Delta_0\right\}. \tag{5.5}$$

Intuitively, (5.3) says "For all $\delta \in \Delta$, if no vertices in V are incident on any δ-edges, then neither are any vertices in U;" (5.4) expresses the reverse implication; and (5.5) says "For all $\delta \in \Delta_0$, some vertices in U are incident on some δ-edges." Suppose Γ is satisfiable. For each $f \in F$ such that X_f is false, take an infinite set U_f, and for each $g \in G$ such that Y_g is false, take an infinite set V_g. Let $U = \bigcup_{f\in F} U_f$ and $V = \bigcup_{g\in G} V_g$. For each $f \in F$, each $u \in U_f$, and each $\delta \in \Delta$, attach $f(\delta)$ δ-labelled edges to u; and similarly for the elements of V, using the functions $g \in G$. By (5.3), if a δ-labelled edge is attached to some vertex (hence infinitely many vertices) of U, then a δ-labelled edge is attached to some vertex (hence infinitely many vertices) of V. And by (5.4), the same holds with U and V transposed. Hence these edges can easily be matched up to form an infinite bipartite graph. By (5.5), there exist δ-labelled edges for every $\delta \in \Delta_0$. Hence $\mathcal{P}$ is a positive instance of BGE. Conversely, if $\mathcal{P}$ is a positive instance of BGE, let $H = (U, V, \mathbf{E}_\Delta)$ be a solution. Now interpret the variables X_f and Y_g as indicated above. It is obvious that (5.3)–(5.5) hold. Thus, Γ is satisfiable. This completes the reduction. □

We note in passing that there exists a sequence $\{\mathcal{P}_n\}_{n\geq 1}$ of positive instances of finite BGE such that the size of $\mathcal{P}_n = (\Delta_n, \Delta_0, F_n, G_n)$ is bounded by a polynomial function of n, but such that the smallest solution has size approximately 2^n. Specifically, we set $\Delta_n = \{\delta_0, \ldots, \delta_{2n-1}\}$, $\Delta_0 = \{\delta_0\}$, $F_n = \{f_0, \ldots, f_n\}$ and $G_n = \{g_0, \ldots, g_{n-1}\}$, where, taking addition in subscripts δ_{2i+1} modulo $2n$:

$$f_0(\delta) = \begin{cases} 1 & \text{if } \delta = \delta_0 \\ 0 & \text{otherwise.} \end{cases} \qquad f_i(\delta) = \begin{cases} 1 & \text{if } \delta = \delta_{2i-1} \\ 2 & \text{if } \delta = \delta_{2i} \\ 0 & \text{otherwise;} \end{cases} \quad (0 < i < n)$$

$$f_n(\delta) = \begin{cases} 1 & \text{if } \delta = \delta_{2n-1} \\ 0 & \text{otherwise.} \end{cases} \qquad g_i(\delta) = \begin{cases} 1 & \text{if } \delta = \delta_{2i} \\ 1 & \text{if } \delta = \delta_{2i+1} \\ 0 & \text{otherwise} \end{cases} \quad (0 \leq i < n).$$

The reader may easily convince himself that $\mathcal{P}_n$ has a finite solution, and that in any solution, at least 2^i distinct vertices on the right-hand side are incident on δ_{2i}-edges, for all i $(0 \leq i < n)$. Thus, the finite BGE-solutions themselves cannot serve as witnesses for membership in NPTime.

5.2. Skew edges. Recall our motivation for introducing edge-coloured bipartite graphs: we intend the left-hand vertices to represent $r_1^{\#}$-classes in some EC_2^2-structure, the right-hand vertices to represent $r_2^{\#}$-classes, and the variously coloured edges to represent intersections having various isomorphism types. In general, EC_2^2-formulas can impose restrictions on pairs of intersections which belong neither to the same $r_1^\#$-class nor to the same $r_2^\#$-class. Thus, for example, the formula $\forall xy((p(x) \wedge q(y)) \rightarrow (r_1^\#(x,y) \vee r_2^\#(x,y)))$ says that there cannot be such a pair of intersections, one with an element satisfying p and the other with an element satisfying q. And we need some way of representing these restrictions in terms of the corresponding edge-coloured bipartite graph. To this end, we call a pair of edges (u_1, v_1) and (u_2, v_2) in a bipartite graph *skew* if $u_1 \neq u_2$ and $v_1 \neq v_2$. We now proceed to define the problem BGES ("bipartite graph existence with skew restrictions"). A *BGES-instance* is a quintuple $\mathcal{P} = (\Delta, \Delta_0, F, G, X)$, where Δ, Δ_0, F and G are as before, and X is a symmetric relation on Δ. A *solution* of $\mathcal{P}$ is a bipartite Δ-graph $H = (U, V, \mathbf{E}_\Delta)$ satisfying (G1)–(G3) above, as well as

$$\text{if } e \in E_\delta \text{ and } e' \in E_{\delta'} \text{ with } e,\, e' \text{ skew, then } (\delta, \delta') \in X. \tag{G4}$$

The problem *BGES* is as follows:

GIVEN: a BGES-instance $\mathcal{P}$.
OUTPUT: Yes, if $\mathcal{P}$ has a solution; No, otherwise.

The problem *finite BGES* is defined analogously. Thus, (finite) BGES is just like (finite) BGE, but with X specifying the allowed colours of skew edge-pairs.

To establish a lower complexity bound for BGES and finite BGES, we proceed by reduction from the well-known NPTIME-hard problem 3-SAT: given a set of propositional clauses each of which contains at most three literals, determine whether there exists a truth-valuation making all clauses simultaneously true.

LEMMA 5.8. *The problems BGES and finite BGES are* NPTIME*-hard.*

Proof. Let $\varphi = \bigwedge_{C \in \mathcal{C}} C$ be an instance of 3-SAT, where each C is a disjunction of literals over variables from a set $\mathcal{V}$. For a given literal l, let $v(l)$ denote the variable of this literal and $s(l) = 1$ if l is positive, and $s(l) = 0$ otherwise. We define a BGES-instance $\mathcal{P}_\varphi = (\Delta, \Delta_0, F, G, X)$, of size polynomial in $|\varphi|$, such that: (i) if φ is satisfiable, then $\mathcal{P}_\varphi$ has a finite solution; and (ii) if $\mathcal{P}_\varphi$ has a solution, then φ is satisfiable. (In fact, $\mathcal{P}_\varphi$ will have no infinite solutions.) Let $\Delta_0 := \{\diamond\}$, where $\diamond$ is a fresh symbol, $\Delta := \{\langle C, v(l), s(l)\rangle : l \text{ is a disjunct of } C, C \in \mathcal{C}\} \cup \mathcal{C} \cup \mathcal{V} \cup \Delta_0$, and let $X := \Delta^2 \setminus \{(\diamond, \diamond)\}$. It remains to define F and G. We take F to consist of a function $f_\diamond$, together with a function f_p^s for each $p \in \mathcal{V}$, $s \in \{0, 1\}$. Likewise, G consists of a function $g_\diamond$ together with a function g_C^W for each $C \in \mathcal{C}$ and every nonempty subset W of literals of C (notice that since there are at most three disjuncts in each clause, for each clause there are at most seven such subsets). All these functions have domain Δ and co-domain $\{0, 1\}$, and are defined as follows:

$$
\begin{aligned}
f_\diamond(\delta) &= 1 \text{ iff } \delta \in \{\diamond\} \cup \mathcal{C}\\
f_p^s(\delta) &= 1 \text{ iff } \delta \in \{p\} \cup \{\langle C, p, s\rangle : \text{for some literal } l \text{ of } C, v(l) = p \wedge s(l) = s\}\\
g_\diamond(\delta) &= 1 \text{ iff } \delta \in \{\diamond\} \cup \mathcal{V}\\
g_C^W(\delta) &= 1 \text{ iff } \delta \in \{C\} \cup \{\langle C, v(l), s(l)\rangle : l \text{ is a literal in } W\}.
\end{aligned}
$$

This completes the reduction. Clearly, it can be performed in polynomial time.

(i) Assume that φ is satisfiable, and let σ be a truth-valuation which makes φ true. We construct a finite solution for $\mathcal{P}_\varphi$ of the form $H = (U, V, E_\Delta)$, where $U = \{u_\diamond\} \cup \mathcal{V}$, $V = \{v_\diamond\} \cup \mathcal{C}$, $E_\diamond = \{(u_\diamond, v_\diamond)\}$, $E_p = \{(p, v_\diamond)\}$ for $p \in \mathcal{V}$, $E_C = \{(u_\diamond, C)\}$ for $C \in \mathcal{C}$,

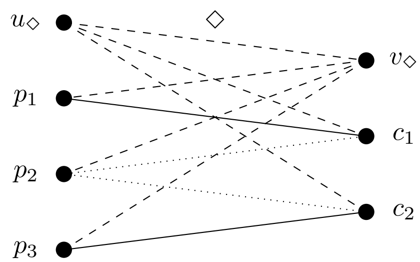

Figure 5.1: The intended solution of $\mathcal{P}_\varphi$ for $\varphi = c_1 \wedge c_2$, where $c_1 = p_1 \vee \neg p_2 \vee \neg p_3$ and $c_2 = \neg p_2 \vee p_3$, and a valuation σ given by $\sigma(p_1) = \sigma(p_3) = 1$ and $\sigma(p_2) = 0$. Dashed lines represent $\mathcal{V}$-edges and $\mathcal{C}$-edges, solid lines represent $\langle p_1, c_1, 1\rangle$-edges and $\langle p_3, c_2, 1\rangle$-edges and dotted lines represent $\langle p_2, c_1, 0\rangle$-edges and $\langle p_2, c_2, 0\rangle$-edges.

$(p, C) \in E_{\langle C', p', 0\rangle}$ if and only if $C = C', p = p'$, $\neg p$ is a literal in C and $\sigma(p) = 0$, and $(p, C) \in E_{\langle C', p', 1\rangle}$ if and only if $C = C', p = p'$, p is a literal in C and $\sigma(p) = 1$. Observe that $\mathrm{ord}^H_{u_\Diamond} = f_\Diamond$, $\mathrm{ord}^H_{v_\Diamond} = g_\Diamond$, for all $p \in \mathcal{V}$ we have $\mathrm{ord}^H_p = f_p^{\sigma(p)}$, and for all $C \in \mathcal{C}$ we have $\mathrm{ord}^H_C = g^W_C$, where W consists of those literals of C which are made true by σ. So H is indeed a solution for $\mathcal{P}_\varphi$. See Fig. 5.1, which illustrates an intended solution for an example φ.

(ii) Let $H = (U, V, E_\Delta)$ be a solution of $\mathcal{P}_\varphi$. We argue that φ is satisfiable. Observe first threat $|E_\Diamond| = 1$. Indeed, Δ_0 guarantees that $|E_\Diamond|$ is greater than 0, no function $f \in F \cup G$ satisfies $f(\Diamond) > 1$ and $\Diamond$-edges cannot be skew. Let $(u_\Diamond, v_\Diamond)$ be the only edge in $E_\Diamond$. Note now that the only possible order function of $u_\Diamond$ is $f_\Diamond$, and the only possible order function of $v_\Diamond$ is $g_\Diamond$. It is not hard to see that for each $p \in \mathcal{V}$, $|E_p| = 1$. This is because $g_\Diamond$ is the only order function in G that allows p-edges, and there is precisely one vertex in V, namely $v_\Diamond$, that has order $g_\Diamond$. Since each $u \in U \setminus \{u_\Diamond\}$ has to be connected to precisely one such edge (because of the definition of F) it follows that $|U| = |\mathcal{V}| + 1$. We denote by u_p the vertex of U that is incident to the p-edge. Similarly, for each $C \in \mathcal{C}$, $|E_C| = 1$, and $|V| = |\mathcal{C}| + 1$. We denote by u_C the vertex of V that is incident to the C-edge.

Now we are ready to define the valuation σ that satisfies φ. For each variable p, we set $\sigma(p) = 1$ if for some C, $E_{\langle C,p,1\rangle}$ is not empty, and $\sigma(p) = 0$ if for some C, $E_{\langle C,p,0\rangle}$ is not empty or for all C, s, $E_{\langle C,p,s\rangle}$ are empty. Note that this definition is sound — for any p, the only vertex of U that can be incident on any edge with colour $\langle C, p, s\rangle$ is u_p (because that vertex must also be incident to the p-coloured edge), so the order function of u_p is either f^0_p or f^1_p. Thus u_p is incident only to edges whose colour is of the form $\langle C, p, 0\rangle$ or $\langle C, p, 1\rangle$, resp. We show that σ indeed satisfies φ. Let $C \in \mathcal{C}$ be a clause. Since u_C is incident to a C-edge, the order function of u_C must be of the form g^W_C for some non-empty W. Let l be a literal from W. Clearly, C is incident to a $\langle C, v(l), s(l)\rangle$-edge, and therefore $E_{\langle C,v(l),s(l)\rangle}$ is not empty, so $\sigma(l) = 1$ and C is satisfied. □

In Sec. 5.3, we shall obtain a matching NPTime upper bound for BGES. We end this section with a simple observation on skew edges.

Lemma 5.9. *Suppose $H = (U, V, \mathbf{E}_\Delta)$ is a Δ-graph. If $\delta \in \Delta$, then H has a pair of skew edges e, e' in E_δ if and only if both the following conditions hold:*

(*i*) *there is more than one $u \in U$ incident on a δ-edge;*

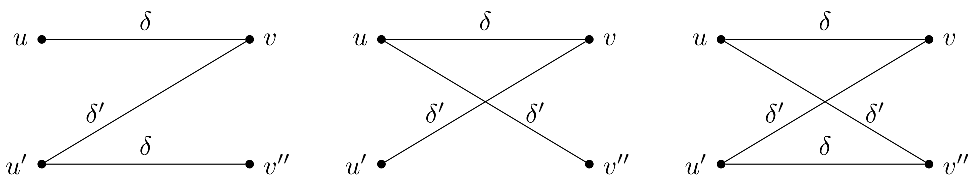

Figure 5.2: The three configurations in final condition of Lemma 5.9: in each configuration, no other δ- or δ'-edges occur.

(ii) there is more than one $v \in V$ incident on a δ-edge.
Further, if $\delta' \in \Delta$ is distinct from δ, then H has a pair of skew edges $e \in E_\delta$ and $e' \in E_{\delta'}$ if and only if all the following conditions hold:
(iii) there are δ-edges and δ'-edges;
(iv) there is more than one $u \in U$ incident on either a δ- or a δ'-edge;
(v) there is more than one $v \in V$ incident on either a δ- or a δ'-edge;
(vi) the edge-sets E_δ and $E_{\delta'}$ are not isomorphic to any of the three configurations shown in Fig. 5.2.

Proof. For the first statement, it is obvious that, if $e, e' \in E_\delta$ are skew, then Conditions (i) and (ii) hold. Suppose, conversely, Conditions (i) and (ii) hold. By Condition (i), let $e = (u, v)$, $e' = (u', v')$ be edges in E_δ with $u \neq u'$. If $v \neq v'$, these edges are skew and we are done; so assume $v = v'$. By Condition (ii), let $e'' = (u'', v'')$ be an edge in E_δ with $v \neq v''$. Then e'' is skew to at least one of e and e'.

For the second statement, it is obvious that, if e and e' are skew with $e \in E_\delta$ and $e' \in E_{\delta'}$, then Conditions (iii)–(v) obtain; and, for Condition (vi), a quick check confirms that, if E_δ and $E_{\delta'}$ are as in Fig. 5.2, then $e \in E_\delta$ and $e' \in E_{\delta'}$ cannot be skew. For the converse, suppose Conditions (iii)–(vi) obtain, but H contains no δ-edge skew to any δ'-edge. By Conditions (iii) and (iv) , we can find $(u, v) \in E_\delta$ and $(u', v') \in E_{\delta'}$ with $u \neq u'$. Since these are not skew, $v = v'$. By Condition (v), we can find $v'' \in V$, distinct from v, lying on either a δ-edge or a δ'-edge. But then, if (u'', v'') is a δ-edge, $u'' = u'$; and if it is a δ'-edge, $u'' = u$. And clearly, no other δ- or δ'-edges are possible. Hence, E_δ and $E_{\delta'}$ are exactly as depicted in one of the arrangements of Fig. 5.2, contradicting Condition (vi). □

We see from Lemma 5.9 that skew restrictions can introduce *upper* bounds on the number of occurrences of vertices of certain orders. (Thus, for example, if $(\delta, \delta) \notin X$, then in any graph satisfying (G4), one of Conditions (i) or (ii) in Lemma 5.9 must fail: in other words, either there is at most one vertex $u \in U$ with any order f such that $f(\delta) \geq 1$, or there is at most one vertex $v \in V$ with any order g such that $g(\delta) \geq 1$.) This means that we cannot in general take the union of two solutions to a BGES problem to form a larger solution. In number-theoretic terms, when we convert BGES instances into systems of equations over $\mathbb{N}$ (or $\mathbb{N}^*$), the resulting solution sets are—as we shall see—not preserved under multiplication by a constant. This observation explains the complexity-theoretic differences (assuming, of course, that NPTime ≠ PTime) between BGE and BGES.

5.3. Ceilings on orders. To apply the graph existence problem to the concerns of the present paper, we require one further complication. So far, we have taken the sets F and G in any BGES-instance to specify the allowed orders of vertices *exactly.* We now consider the case where these orders are known only up to a certain *ceiling*,

M. Specifically, for $M \geq 0$, we define $\lfloor n \rfloor_M = \min(n, M)$, and if f is any function with range $\mathbb{N}$, we denote by $\lfloor f \rfloor_M$ the composition $\lfloor \cdot \rfloor_M \circ f$ (i.e., $\lfloor f \rfloor_M$ is the result of applying f and 'capping' at M). We proceed to define the problem BGESC (bipartite graph existence with skew constraints and ceiling). A *BGESC-instance* is a sextuple $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$, where Δ, Δ_0, X are as before, M is a positive integer, and F and G are sets of functions $\Delta \to [0, M]$. A *solution* of $\mathcal{P}$ is a bipartite Δ-graph $H = (U, V, \mathbf{E}_\Delta)$ satisfying the following variants of conditions (G1)–(G4):

$$\text{for all } \delta \in \Delta_0,\ E_\delta \text{ is non-empty;} \tag{G1}$$

$$\text{for all } u \in U,\ \lfloor \mathrm{ord}^H_u \rfloor_M \in F; \tag{G2$'$}$$

$$\text{for all } v \in V,\ \lfloor \mathrm{ord}^H_v \rfloor_M \in G; \tag{G3$'$}$$

$$\text{if } e \in E_\delta \text{ and } e' \in E_{\delta'},\ \text{with } e,\ e' \text{ skew, then } (\delta, \delta') \in X. \tag{G4}$$

The problem *BGESC* is defined as follows.

Given: a BGESC-instance $\mathcal{P}$.
Output: Yes, if $\mathcal{P}$ has a solution; No otherwise.

The problem *finite BGESC* is defined analogously. Thus (finite) BGESC is just like (finite) BGES, but with M specifying the bound past which we do not bother counting orders. By Lemma 5.8, these problems are certainly NPTime-hard. (Just take M to be the maximum value of any function in $F \cup G$ plus one.)

The following definition will be used later in this section: we introduce it here because of its obvious connection to the problem BGESC. Let $H = (U, V, \mathbf{E}_\Delta)$ and $H' = (U', V', \mathbf{E}'_\Delta)$ be Δ-graphs. We write $H \approx_M H'$, and say that H and H' are *M-approximations of* each other if $U = U'$, $V = V'$ and, for all $w \in U \cup V$, $\lfloor \mathrm{ord}^H_w \rfloor_M = \lfloor \mathrm{ord}^{H'}_w \rfloor_M$. Thus, a BGESC problem-instance $(\Delta, \Delta_0, M, F, G, X)$ requires us to determine the existence of an M-approximation to some solution of the corresponding BGES problem-instance $(\Delta, \Delta_0, F, G, X)$.

The main result of this section is:

Theorem 5.10. BGESC *and finite* BGESC *are* NPTime*-complete.*

Theorem 5.10, as well as being interesting in its own right, allows us to prove that the satisfiability and finite satisfiability problems for EC^2_2 are in 2-NExpTime, as we shall see in Sec. 6.

The remainder of this section is devoted to a proof of the membership part of Theorem 5.10. We mention here that readers interested primarily in decidability, rather than computational complexity, may simply reduce (finite) BGESC to the (finite) satisfiability problem for $\mathcal{C}^2$—the two-variable fragment of first-order logic with counting quantifiers. The reduction is straightforward, and we outline it only in general terms. For each $f \in F$, let p_f be a unary predicate, for each $g \in G$, let q_g be a unary predicate, and for each $\delta \in \Delta$, let r_δ be a binary predicate. We think of $p_f(x)$ as saying "x is left-hand node with order f", and similarly for q_g; and we think of $r_\delta(x, y)$ as saying "(x, y) is a δ-edge." Given a BGESC-instance $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$, we can write $\mathcal{C}^2$-formulas expressing obvious constraints under these interpretations, for example:

$$\forall x \left(p_f(x) \to \exists_{=f(\delta)} y . r_\delta(x, y)\right) \qquad \text{if } f(\delta) < M$$
$$\forall x \left(p_f(x) \to \exists_{\geq M} y . r_\delta(x, y)\right) \qquad \text{otherwise,}$$

and similarly for the q_g. Using this signature, Conditions (i)–(v) in Lemma 5.9 can evidently be expressed using $\mathcal{C}^2$-formulas. To see that the same holds for Condition (vi), consider the first graph in Fig. 5.2. We can rule out the possibility that the edges of E_δ and $E_{\delta'}$ have this configuration using the $\mathcal{C}^2$-formula

$$\neg\big(\exists x \exists y (r_\delta(x,y) \wedge \exists x (r_{\delta'}(y,x) \wedge \exists y . r_\delta(x,y))) \wedge \\ \exists_{=2} x \exists y (r_\delta(x,y), \vee r_{\delta'}(x,y)) \wedge \exists_{=2} y \exists x (r_\delta(y,x), \vee r_{\delta'}(y,x))\big),$$

(We assume obvious formulas stating the disjointness of the edge-colours and ensuring the division of vertices into left- and right-hand sides.) The other two graphs can be ruled out similarly. Thus, we may write a $\mathcal{C}^2$-formula $\varphi_\mathcal{P}$ such that $\varphi_\mathcal{P}$ is (finitely) satisfiable if and only if $\mathcal{P}$ is a positive instance of (finite) BGESC. And the (finite) satisfiability problem for $\mathcal{C}^2$ is known to be decidable [10, 27]. Unfortunately, both problems are NExpTime-complete [29], and so do not yield tight a complexity bound for (finite) BGESC. So we still have work to do below.

To establish Theorem 5.10, however—and, in particular, to cope with the variant conditions (G2′) and (G3′)—we require a still more intricate version of BGESC. Define a *directed* Δ*-graph* to be a quintuple $H = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$, where U and V are countable (possibly finite, possibly empty) disjoint sets, and $\mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta$ and $\mathbf{E}^\circ_\Delta$ are families of sets E^+_δ, E^-_δ and E°_δ, all of which (taken together) form a collection of pairwise disjoint subsets of $U \times V$. We always write $E_\delta = E^+_\delta \cup E^-_\delta \cup E^\circ_\delta$ for any $\delta \in \Delta$. It helps to think of H as the result of giving the edges of the underlying (un-directed) Δ-graph $\bar{H} = (U, V, \mathbf{E}_\Delta)$ one of three orientations: *left-to-right* (i.e., U*-to-*V) (E^+_δ), *right-to-left* (E^-_δ) or *bi-directional* (E°_δ). For $u \in U$ and $v \in V$, we define the functions $\deg^H_u : \Delta \to \mathbb{N}^*$ and $\deg^H_v : \Delta \to \mathbb{N}^*$ by

$$\deg^H_u(\delta) = |\{v \in V : (u,v) \in E^+_\delta \cup E^\circ_\delta\}|$$
$$\deg^H_v(\delta) = |\{u \in U : (u,v) \in E^-_\delta \cup E^\circ_\delta\}|,$$

and we define the functions $\mathrm{Deg}^H_u : \Delta \to (\mathbb{N}^*)^2$ and $\mathrm{Deg}_v : \Delta \to (\mathbb{N}^*)^2$ by

$$\mathrm{Deg}^H_u(\delta) = (|\{v \in V : (u,v) \in E^+_\delta\}|, |\{v \in V : (u,v) \in E^\circ_\delta\}|)$$
$$\mathrm{Deg}^H_v(\delta) = (|\{u \in U : (u,v) \in E^-_\delta\}|, |\{u \in U : (u,v) \in E^\circ_\delta\}|).$$

Thus, for any vertex w, $\deg^H_w(\delta)$ (pronounced: "δ-degree of w") counts the number of uni- or bi-directional δ-edges emanating from w, ignoring incoming edges. The pair $\mathrm{Deg}^H_w(\delta)$ simply splits $\deg^H_w(\delta)$ into the uni- and bi-directional components. We require the following notation in the sequel. If (m,n) is a pair of elements of $\mathbb{N}^*$, we write $(m,n)|_1 = m$ and $(m,n)|_2 = n$. Thus: $\deg^H_w(\delta) = \mathrm{Deg}^H_w(\delta)|_1 + \mathrm{Deg}^H_w(\delta)|_2$.

Let H be a directed Δ-graph and M a positive integer. We say that H is M*-bounded* if $\deg^H_w(\delta) \leq M$ for all $w \in U \cup V$ and all $\delta \in \Delta$. We say that H is M*-proper* if, for all $u \in U$, $v \in V$ and $\delta \in \Delta$: (i) $(u,v) \in E^+_\delta$ implies $\deg^H_v(\delta) \geq M$; and (ii) $(u,v) \in E^-_\delta$ implies $\deg^H_u(\delta) \geq M$.

It is possible to transform Δ-graphs into directed Δ-graphs by appropriately labelling their edges.

Lemma 5.11. *Suppose M is a positive integer and $H = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ an M-bounded, M-proper, directed Δ-graph; and define the collection $\mathbf{E}_\Delta$ by setting $E_\delta = E^+_\delta \cup E^-_\delta \cup E^\circ_\delta$ for all $\delta \in \Delta$. Then the Δ-graph $\bar{H} = (U, V, \mathbf{E}_\Delta)$ satisfies $\deg^H_w = \lfloor \mathrm{ord}^{\bar{H}}_w \rfloor_M$, for all $w \in U \cup V$. Moreover, given a Δ-graph H', and positive integer M, we can find an M-bounded, M-proper, directed Δ-graph H such that $\bar{H} \approx_M H'$.*

Proof. The first statement is immediate from the fact that H is M-proper. For if $u \in U$ participates in any edges of E^-_δ, we have $\mathrm{ord}^{\bar{H}}_u(\delta) \geq \deg^H_w(\delta) = M$; otherwise, $\mathrm{ord}^{\bar{H}}_u(\delta) = \deg^H_w(\delta)$. Similarly for the vertices of V. For the second statement, suppose $H' = (U, V, \mathbf{E}_\Delta)$. We construct $\mathbf{E}^+_\Delta$, $\mathbf{E}^-_\Delta$, $\mathbf{E}^\circ_\Delta$ as follows. For each $u \in U$ and each $\delta \in \Delta$, select edges $(u, v) \in E_\delta$ until either M edges have been selected or no more can be found; mark each selected edge (u, v) with an arrow from u to v. For each $v \in V$ and each $\delta \in \Delta$, select edges $(u, v) \in E_\delta$ until either M edges have been selected or no more can be found; mark each selected edge (u, v) with an arrow from v to u. For each $\delta \in \Delta$, let E^+_δ be the set of $(u, v) \in E_\delta$ with an arrow from u to v, but no arrow from v to u; let E^-_δ be the set of $(u, v) \in E_\delta$ with an arrow from v to u, but no arrow from u to v; let E°_δ be the set of $(u, v) \in E_\delta$ with an arrow from u to v and also an arrow from v to u. Discard any edges in E_δ with no arrows at all. By construction, H is M-bounded. To see that it is M-proper, consider first any $u \in U$ and $\delta \in \Delta$. If u is incident on at most M edges in E_δ, then a left-to-right arrow will be placed on all of these edges, and so u will be incident on no edges of $\mathbf{E}^-_\Delta$. If, on the other hand, u is incident on at more than M edges in E_δ, then a left-to-right arrow will be placed on M of these, whence $\deg^H_w(\delta) = M$. A symmetric argument applies to any $v \in V$. Similar reasoning shows that $\bar{H} \approx_M H'$. □

Let Γ and Δ be non-empty sets. A Γ-*partitioned, directed* Δ-*graph* is a quintuple $H = (\mathbf{U}_\Gamma, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$, where $\mathbf{U}_\Gamma$ is a collection of pairwise disjoint sets U_γ such that, setting $U = \bigcup_{\gamma \in \Gamma} U_\gamma$, the quintuple $\dot{H} = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ is a directed Δ-graph. When dealing with Γ-partitioned, directed Δ-graphs, we always use the notation $U = \bigcup_{\gamma \in \Gamma} U_\gamma$; and we continue to use the notation $E_\delta = E^+_\delta \cup E^-_\delta \cup E^\circ_\delta$. We define the functions $\deg^H_w$ and Deg^H_w as above; additionally, we define the functions $\mathrm{DEG}^H_v : \Gamma \times \Delta \to (\mathbb{N}^*)^2$ for $v \in V$ by:

$$\mathrm{DEG}^H_v(\gamma, \delta) = (|\{u \in U_\gamma : (u, v) \in E^-_\delta\}|, |\{u \in U_\gamma : (u, v) \in E^\circ_\delta\}|).$$

It helps to think of H as the result of partitioning the left-vertices of the underlying directed Δ-graph, $\dot{H}$, into (possibly empty) cells U_γ, indexed by the elements of Γ. Note that the right-vertices V are not partitioned in this way. The function $\mathrm{DEG}^H_v(\gamma, \delta)$ thus specifies how the right-to-left and bi-directional edges incident to v distribute over the partition $\mathbf{U}_\Gamma$: in particular, $\mathrm{Deg}^H_v(\delta)|_i = \sum_{\gamma \in \Gamma} \mathrm{DEG}^H_v(\gamma, \delta)|_i$ for $i = 1, 2$. For M a positive integer, we call H M-*bounded* (M-*proper*) if the underlying unpartitioned directed Δ-graph, H' is. We call a function $r : \Gamma \times \Delta \to \mathbb{N}^2$ *unitary* if

$$\sum_{\delta \in \Delta} (r(\gamma, \delta))|_2 \leq 1 \qquad \text{for all } \gamma \in \Gamma,$$

and we call H *unitary* if, for all $v \in V$, DEG_v is unitary. Thus, H is unitary just in case, for each γ, no vertex in V is linked via bi-directional edges (regardless of colour) to more than one vertex in U_γ.

It is possible to transform directed Δ-graphs into unitary, partitioned, directed Δ-graphs by appropriately labelling their left-vertices.

Lemma 5.12. *Suppose M is a positive integer, and $H = (\mathbf{U}_\Gamma, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ a Γ-partitioned, directed Δ-graph; and let $U = \bigcup_{\gamma \in \Gamma} U_\gamma$. Then the directed Δ-graph $\dot{H} = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ satisfies $\mathrm{Deg}^{\dot{H}}_w = \mathrm{Deg}^H_w$ for all $w \in U \cup V$; hence, H is M-bounded if and only if $\dot{H}$ is, and also M-proper if and only if $\dot{H}$ is. Moreover, given an M-bounded, directed Δ-graph H', we can find a set Γ with $|\Gamma| \leq M^2|\Delta|^2$, and a* *unitary,* Γ*-partitioned, directed* Δ*-graph* $H = (\mathbf{U}_\Gamma, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ *such that* $H' = \dot{H}$.

Proof. The first statement follows from the definition of Deg^H_w. For the second statement, suppose $H' = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ is given: we must define the partition $\mathbf{U}_\Gamma$ of U. To do so, simply consider the graph $G = (U, E)$ where $(u, u') \in E$ just in case u and u' are distinct and there exists $v \in V$ and $\delta, \delta' \in \Delta$ with $(u, v) \in E^\circ_\delta$ and $(u', v) \in E^\circ_{\delta'}$. Note that G is simply an ordinary (undirected) graph here. Then the degree of a vertex of G—i.e. the number of edges on which that vertex is incident—is bounded by $M|\Delta|(M|\Delta| - 1) < M^2|\Delta|^2$. Hence, the vertices of G can be coloured with $M^2|\Delta|^2$ colours so that no two vertices joined by an edge have the same colour. Let Γ be the set of colours used, and let U_γ be the set of vertices of colour γ, for $\gamma \in \Gamma$. This guarantees that H is unitary. □

We now proceed to define the problem PDBGE ("partitioned, directed bipartite graph existence"). A *PDBGE-instance* is a septuple $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P, R, X)$, where Γ, Δ, Δ_0, M and X are as before, P is a set of functions $\Delta \to [0, M]^2$, and R is a set of unitary functions $\Gamma \times \Delta \to [0, M]^2$. A *solution* of $\mathcal{Q}$ is an M-bounded, M-proper, Γ-partitioned, directed Δ-graph $H = (\mathbf{U}_\Gamma, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ such that:

$$\text{for all } \delta \in \Delta_0,\ E_\delta \text{ is non-empty;} \tag{D1}$$

$$\text{for all } u \in U,\ \mathrm{Deg}^H_u \in P; \tag{D2}$$

$$\text{for all } v \in V,\ \mathrm{DEG}^H_v \in R. \tag{D3}$$

$$\text{if } e \in E_\delta \text{ and } e' \in E_{\delta'}, \text{ with } e, e' \text{ skew, then } (\delta, \delta') \in X. \tag{D4}$$

The problem *PDBGE* is defined as follows.

GIVEN: a PDBGE-instance $\mathcal{P}$.
OUTPUT: Yes, if $\mathcal{P}$ has a solution; No otherwise.

Thus, PDBGE is a variant of BGESC in which the left-hand vertices have colours (chosen from Γ), and the edges have orientations (left-to-right, right-to-left or bi-directional). The problem *finite PDBGE* is defined analogously. We proceed to establish membership of (finite) PDBGE in NPTIME. Two simple, combinatorial results will prove useful in this enterprise.

LEMMA 5.13. *Let* $\ell, m, n \geq 0$*, let* Z *be a set, and let* $Z_0, Z_1, \ldots Z_n$ *be subsets of* Z *with* $|Z_0| = \ell$*. Then there exists* Z^* *such that:* $Z_0 \subseteq Z^* \subseteq Z$*;* $|Z^*| \leq \ell + mn$*; and for all* i $(1 \leq i \leq n)$*, either* $Z_i \subseteq Z^*$ *or* $|Z_i \cap (Z \setminus Z^*)| > m$.

Proof. Begin by setting $Z^* = Z_0$. As long as there is any Z_i such that $1 \leq |Z_i \cap (Z \setminus Z^*)| \leq m$, add all the elements of Z_i to Z^*. This process must terminate after at most n rounds, each involving the addition of at most m elements. □

LEMMA 5.14. *Let* $m, n \geq 1$*, let* Z *be a set, and let* $Z_1, \ldots, Z_n$ *be subsets of* Z *with* $|Z_i| \geq m(n+1)$*. Then we can partition* Z *into sets* Z^+ *and* Z^- *such that, for all* i $(1 \leq i \leq n)$*,* $|Z_i \cap Z^+| \geq m$ *and* $|Z_i \cap Z^-| \geq m$.

Proof. For each i $(1 \leq i \leq n)$, select m elements of Z_i for inclusion in Z^+. Let Z^- be the set of elements not selected in this process. By construction, $|Z_i \cap Z^+| \geq m$, and, furthermore, $|Z^+| \leq nm$. That $|Z_i \cap Z^-| \geq m$ then follows from the fact that $|Z_i| \geq m(n+1)$. □

LEMMA 5.15. *The problems* PDBGE *and* finite PDBGE *are in* NPTIME.

Proof. We deal first with the case PDBGE; the result for finite PDBGE will follow by a simple adaptation. The proof consists of four stages. In Stage 1, we take any PDBGE-instance $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P, R, X)$, and construct a certain data-structure, $\mathcal{D}$, which we refer to as a quasi-certificate. In Stage 2, we derive a collection of conditions which $\mathcal{D}$ must satisfy, on the assumption that $\mathcal{Q}$ has a solution. These conditions, numbered (5.6)–(5.23) in the proof, constitute a Boolean combination of linear equations and inequalities in the variables $x_{\gamma,p}$ and y_r (with γ, p and r ranging over specified index sets). We show how satisfying values for these variables can be read off from any solution of $\mathcal{Q}$. In Stage 3, we reverse this process, showing that, given quasi-certificate $\mathcal{D}$, satisfying (5.6)–(5.23), we can construct a solution of $\mathcal{Q}$. Thus, the original PDBGE-instance $\mathcal{Q}$ has been transformed into the problem of determining the solvability of a system of linear equations and inequalities. We characterize the size of this system of conditions rather carefully: in particular, we show that the total number of equations and inequalities involved is polynomial in the quantities $|\Gamma|$, $|\Delta|$ and M, as indeed are all the constant terms involved; however, the number of variables, and therefore the total size of the system of conditions, need not be so bounded. In Stage 4, we use facts about integer programming (specifically, Proposition 5.2) to show that the existence of some $\mathcal{D}$ satisfying (5.6)–(5.23) can be checked in time polynomially bounded as a function of $|\Gamma|$, $|\Delta|$ and M.

The following imagery will be helpful in the sequel. Let $H = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ be a directed Δ-graph. If $u \in U$ and $\delta \in \Delta$, we speak of any δ-edge in $e \in E^+_\delta \cup E^\circ_\delta$ such that u is incident on e as being 'sent' by u. Likewise, if $v \in V$, we speak of any δ-edge in $e \in E^-_\delta \cup E^\circ_\delta$ such that v is incident on e as being 'sent' by v. (Thus, left-to-right edges are sent by their left-vertices, right-to-left edges by their right-vertices, and bi-directional edges by both of their vertices.) If H is M-bounded, a vertex can send at most M δ-edges; and if H is M-proper, a vertex can 'receive' a δ-edge only if it sends at least M δ-edges. That is, vertices which send fewer than M δ-edges are disqualified from receiving any uni-directional δ-edges at all. Accordingly, where H and M are clear from context, and H is M-proper, we call a vertex w of H *δ-receptive* if $\deg^H_w(\delta) \geq M$, regardless of whether w actually receives any δ-edges.

Stage 1: Let a PDBGE-instance $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P, R, X)$ be given. We first assume that there is a solution of $\mathcal{Q}$, and we use that solution to construct a *quasi-certificate*

$$\mathcal{D} = (\mathbf{U}^*_\Gamma, \mathbf{U}^+_\Gamma, V^*, V^+, \mathbf{L}^+_\Delta, \mathbf{L}^-_\Delta, \mathbf{L}^\circ_\Delta, \mathbf{p}_{U^+}, \mathbf{r}_{V^+}).$$

Here, the components $\mathbf{U}^*_\Gamma$ and $\mathbf{U}^+_\Gamma$ are collections of sets satisfying $U^*_\gamma \subseteq U^+_\gamma$ for all $\gamma \in \Gamma$; V^* and V^+ are sets satisfying $V^* \subseteq V^+$. Furthermore, writing $U = \bigcup_\Gamma U_\gamma$, and similarly for U^* and U^+, the components $\mathbf{L}^+_\Delta$, $\mathbf{L}^-_\Delta$ and $\mathbf{L}^\circ_\Delta$ are subsets of $(U^+ \times V^*) \cup (U^* \times V^+)$ such that $(\mathbf{U}^+_\Gamma, V^+, \mathbf{L}^+_\Delta, \mathbf{L}^-_\Delta, \mathbf{L}^\circ_\Delta)$ is a Γ-partitioned, directed Δ-graph. Finally, the component $\mathbf{p}_{U^+}$ is a collection of functions in P indexed by the elements of U^+; and the component $\mathbf{r}_{V^+}$ is a collection of functions in R indexed by the elements of V^+.

Suppose $H = (\mathbf{U}_\Gamma, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ is a solution of $\mathcal{Q}$; as usual, we write $U = \bigcup_{\gamma \in \Gamma} U_\gamma$, and $E_\delta = E^+_\delta \cup E^-_\delta \cup E^\circ_\delta$, for $\delta \in \Delta$. We begin the construction of $\mathcal{D}$ by defining the collection of sets $\mathbf{U}^*_\Gamma$ and the set V^*. As a preliminary, say that $\delta \in \Delta$ is *left-special* if at most two vertices $u \in U$ satisfy $\deg^H_u(\delta) > 0$, and say that $u \in U$ is *special* if $\deg^H_u(\delta) > 0$ for some left-special δ. Let U'_γ be the set of special vertices in U_γ, for all $\gamma \in \Gamma$, and let V' be the set of special elements of V, defined analogously. Evidently, $|U'_\gamma| \leq 2|\Delta|$, and $|V'| \leq 2|\Delta|$. Fix $\gamma \in \Gamma$ and, for $\delta \in \Delta$, define $U_{\gamma,\delta}$ to be the set of δ-receptive vertices in U_γ. We apply Lemma 5.13 with $m = M|\Delta|(|\Delta| + 1)$, $n = |\Delta|$, $Z = U_\gamma$, $Z_0 = U'_\gamma$, and $Z_1, \ldots, Z_n$ a list of the sets $U_{\gamma,\delta}$, for $\delta \in \Delta$. Then there exists a set of vertices U^*_γ such that: $U'_\gamma \subseteq U^*_\gamma \subseteq U_\gamma$; $|U^*_\gamma| \leq 2|\Delta| + M|\Delta|^2(|\Delta| + 1)$; and, for all $\delta \in \Delta$, if any vertices $u \in U_\gamma \setminus U^*_\gamma$ are δ-receptive, then at least $M|\Delta|(|\Delta| + 1)$ are. Similarly, there exists V^* such that: $V' \subseteq V^* \subseteq V$; $|V^*| \leq 2|\Delta| + M|\Delta|^2(|\Delta| + 1)$; and if any vertices $v \in V \setminus V^*$ are δ-receptive, then at least $M|\Delta|(|\Delta| + 1)$ are.

We now define the collection of sets $\mathbf{U}^+_\Gamma$ and the set V^+. To reduce notational clutter, let us write $u \to v$ if $(u, v) \in E^+_\delta \cup E^\circ_\delta$ for some $\delta \in \Delta$ and $u \leftarrow v$ if $(u, v) \in E^-_\delta \cup E^\circ_\delta$ for some $\delta \in \Delta$. Now let

$$U^+_\gamma = U^*_\gamma \cup \{u \in U_\gamma \mid u \leftarrow v \text{ for some } v \in V^*\}$$
$$V^+ = V^* \cup \{v \in V \mid u \to v \text{ for some } u \in U^*\}.$$

Thus, U^+_γ adds to U^*_γ those elements of U_γ reachable via either a right-to-left or a bi-directional edge from V^*, while V^+ adds to V^* those elements of V reachable via either a left-to-right or a bi-directional edge from U^*. Since H is M-bounded, each of the sets U^+_γ or V^+ has cardinality at most $(2|\Delta| + M|\Delta|^2(|\Delta| + 1))(M|\Delta| + 1)$.

The next step in the construction of $\mathcal{D}$ is to define the collections of edge-sets $\mathbf{L}^+_\Delta$, $\mathbf{L}^-_\Delta$ and $\mathbf{L}^\circ_\Delta$. Let Ω denote the set of pairs $(U^+ \times V^*) \cup (U^* \times V^+)$; and for all $\delta \in \Delta$, let $L^+_\delta = (E^+_\delta) \cap \Omega$, $L^-_\delta = (E^-_\delta) \cap \Omega$ and $L^\circ_\delta = (E^\circ_\delta) \cap \Omega$. Then $H^- = (\mathbf{U}^+_\Gamma, V^+, \mathbf{L}^+_\Delta, \mathbf{L}^-_\Delta, \mathbf{L}^\circ_\Delta)$ is an M-bounded, Γ-partitioned, directed Δ-graph (though it need not be M-proper). The motivation for defining H^- is that it is polynomially bounded in M and $|\Delta|$, and that the vertices in U^* and V^* have the same degrees in H^- as they have in H.

The final components of our quasi-certificate $\mathcal{D}$ are the collections of functions $\mathbf{p}_{U^+}$ and $\mathbf{r}_{V^+}$, where $p_u \in P$ for all $u \in U^+$, and $r_v \in R$ for all $v \in V^+$. To define these functions, we simply set

$$p_u = \mathrm{Deg}^H_u$$
$$r_v = \mathrm{DEG}^H_v$$

for $u \in U^+$ and $v \in V^+$: since H is a solution of $\mathcal{Q}$, we have $p_u \in P$ and $r_v \in R$ as required.

This completes the construction of the quasi-certificate $\mathcal{D}$.

Stage 2: In this stage, we derive some properties of $\mathcal{D}$. If (m, n) and (m', n') are pairs of natural numbers, we write $(m, n) \succeq (m', n')$ if $m \geq m'$ and $n \geq n'$. Evidently:

$$\bigwedge_{u \in U^+} \bigwedge_{\delta \in \Delta} \left(p_u(\delta) \succeq \mathrm{Deg}^{H^-}_u(\delta) \right) \tag{5.6}$$

$$\bigwedge_{v \in V^+} \bigwedge_{\gamma \in \Gamma} \bigwedge_{\delta \in \Delta} \left(r_v(\gamma, \delta) \succeq \mathrm{DEG}^{H^-}_v(\gamma, \delta) \right). \tag{5.7}$$

On the other hand, by construction of the sets U^+ and V^+, we have

$$\bigwedge_{u \in U^*} \bigwedge_{\delta \in \Delta} \left(p_u(\delta) = \mathrm{Deg}^{H^-}_u(\delta) \right) \tag{5.8}$$

$$\bigwedge_{v \in V^*} \bigwedge_{\gamma \in \Gamma} \bigwedge_{\delta \in \Delta} \left(r_v(\gamma, \delta) = \mathrm{DEG}^{H^-}_v(\gamma, \delta) \right). \tag{5.9}$$

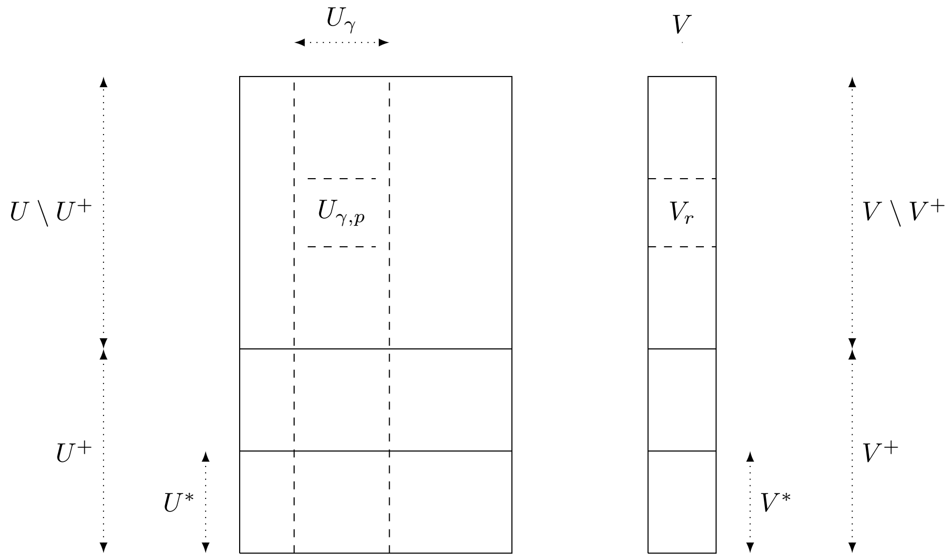

Figure 5.3: The partitioning of $(U_\gamma \setminus U^+)$ into the collection $\{U_{\gamma,p} \mid p \in \mathbf{P}\}$, and of $(V \setminus V^+)$ into the collection $\{V_r \mid r \in \mathbf{R}\}$.

Now, let $\mathbf{P}$ be the set of functions $p : \Delta \to [0, M]^2$, and $\mathbf{R}$ the set of unitary functions $r : \Gamma \times \Delta \to [0, M]^2$. Thus, $P \subseteq \mathbf{P}$ and $R \subseteq \mathbf{R}$. We remark, however, that $\mathbf{P}$ and $\mathbf{R}$ are large sets—not polynomially bounded in $|\Delta|$. For all $\gamma \in \Gamma$ and all $p \in \mathbf{P}$, let $x_{\gamma,p}$ be a new symbol; and for all $r \in \mathbf{R}$, let y_r be a new symbol. Formally, these symbols are *variables* ranging over $\mathbb{N}^*$. Informally, we have a particular valuation in mind: $x_{\gamma,p}$ is the cardinality of the set $U_{\gamma,p} = \{u \in U_\gamma \setminus U^+ \mid \mathrm{Deg}^H_u = p\}$, and y_r is the cardinality of the set $V_r = \{v \in V \setminus V^+ \mid \mathrm{DEG}^H_v = r\}$. Note that the (possibly empty) sets $U_{\gamma,p}$ and V_r partition $U_\gamma \setminus U^+$ and $V \setminus V^+$, respectively, as illustrated in Fig. 5.3. Since H is a solution of $\mathcal{Q}$, we know that $U_{\gamma,p} = \emptyset$ whenever $p \notin P$; similarly, $V_r = \emptyset$ whenever $r \notin R$. That is, under the suggested valuation, the following equations hold:

$$\bigwedge_{\gamma \in \Gamma} \sum_{p \in \mathbf{P} \setminus P} x_{\gamma,p} = 0 \tag{5.10}$$

$$\sum_{r \in \mathbf{R} \setminus R} y_r = 0. \tag{5.11}$$

Our suggested valuation satisfies further conditions. We examine first those arising from the bi-directional edges in H. Fixing $\gamma \in \Gamma$ and $\delta \in \Delta$, the expression $\sum_{u \in U^+_\gamma} \big(p_u(\delta)|_2\big) + \sum_{p \in \mathbf{P}} \big(p(\delta)|_2\big) x_{\gamma,p}$ records the total number of edges in E°_δ incident on the vertices of U_γ; similarly, $\sum_{v \in V^+} \big(r_v(\gamma, \delta)|_2\big) + \sum_{r \in \mathbf{R}} \big(r(\gamma, \delta)|_2\big) y_r$ records the total number of edges in $E^\circ_\delta \cap (U_\gamma \times V)$ incident on the vertices of V. Since these must be equal, we have the condition:

$$\bigwedge_{\gamma\in\Gamma}\bigwedge_{\delta\in\Delta}\left(\sum_{u\in U^+_\gamma}\big(p_u(\delta)|_2\big)+\sum_{p\in\mathbf{P}}\big(p(\delta)|_2\big)x_{\gamma,p}=\right.$$
$$\left.\sum_{v\in V^+}\big(r_v(\gamma,\delta)|_2\big)+\sum_{r\in\mathbf{R}}\big(r(\gamma,\delta)|_2\big)y_r\right). \quad (5.12)$$

We next examine those conditions on $\mathcal{D}$ arising from the uni-directional edges in H. The following notation, which loosely alludes to the construction of $\bar{H}$ and $\dot{H}$ in Lemmas 5.11 and 5.12, will help us do so. For $p\in\mathbf{P}$, define the function $\bar{p}:\Delta\to\mathbb{N}$ by $\bar{p}(\delta)=p(\delta)|_1+p(\delta)|_2$; for $r\in\mathbf{R}$, define the function $\bar{\dot{r}}:\Delta\to\mathbb{N}$ by $\bar{\dot{r}}(\delta)=\sum_\Gamma(r(\gamma,\delta)|_1+r(\gamma,\delta)|_2)$. Observe that, for the particular collections of functions $\mathbf{p}_{U^+}$ and $\mathbf{r}_{V^+}$ defined above, we have $\bar{p}_u=\deg^H_u$ for all $u\in U^+$, and $\bar{\dot{r}}_v=\deg^H_v$ for all $v\in V^+$. In particular, $u\in U^+$ is δ-receptive just in case $\bar{p}_u(\delta)=M$, and $v\in V^+$ is δ-receptive just in case $\bar{\dot{r}}_v(\delta)=M$.

Consider first the left-to-right edges incident on vertices in $U^+\setminus U^*$, as well as the right-to-left edges incident on vertices in $V^+\setminus V^*$. Any vertex $u\in U^+\setminus U^*$, lies on $p_u(\delta)|_1$ left-to-right δ-edges in H. We have two possibilities. If all of these edges link u to vertices in V^*, then $p_u(\delta)|_1=\mathrm{Deg}^{H^-}_u(\delta)|_1$. If, on the other hand, u is linked by a left-to-right δ-edge to at least one vertex in $V\setminus V^*$, then we have at least one δ-receptive vertex in $V\setminus V^*$, since H is M-proper. Corresponding remarks apply to $v\in V^+\setminus V^*$. The following term in the variables $x_{\gamma,p}$ specifies, under the valuation suggested above, the number of δ-receptive vertices of $U_\gamma\setminus U^*_\gamma$:

$$|\{u\in U^+_\gamma\setminus U^*_\gamma:\bar{p}_u(\delta)=M\}|+\sum\{x_{\gamma,p}:p\in\mathbf{P},\bar{p}(\delta)=M\}.$$

We abbreviate this term by $\overleftarrow{s}_\gamma(\delta)$; obviously, there is an analogous term, $\overrightarrow{t}(\delta)$, specifying the number of δ-receptive vertices in $V\setminus V^*$. Thus, we have:

$$\bigwedge_{u\in U^+}\bigwedge_{\delta\in\Delta}\left(p_u(\delta)|_1=(\mathrm{Deg}^{H^-}_u(\delta))|_1\vee\overrightarrow{t}(\delta)\geq 1\right) \quad (5.13)$$

$$\bigwedge_{v\in V^+}\bigwedge_{\gamma\in\Gamma}\bigwedge_{\delta\in\Delta}\left(r_v(\gamma,\delta)|_1=(\mathrm{DEG}^{H^-}_u(\gamma,\delta))|_1\vee\overleftarrow{s}_\gamma(\delta)\geq 1\right). \quad (5.14)$$

Consider now the left-to-right edges incident on vertices in $U\setminus U^+$, as well as the right-to-left edges incident on vertices in $V\setminus V^+$. Fix $\gamma\in\Gamma$ and $\delta\in\Delta$. If, for any $p\in\mathbf{P}$, $x_{\gamma,p}>0$, then there exists a vertex $u\in U_\gamma\setminus U^+$ lying on $p(\delta)|_1$ left-to-right δ-edges in H. We have two possibilities. If all of these edges link u to vertices in V^*, then the number of δ-receptive vertices of V^* must be at least $p(\delta)|_1$. If, on the other hand, u is linked by a left-to-right δ-edge to at least one vertex in $V\setminus V^*$, then we have at least one δ-receptive vertex in $V\setminus V^*$. Taking the constant n_δ to be the number of δ-receptive elements of V^* (which can be computed from $\mathbf{L}^-_\Delta$ and $\mathbf{L}^\circ_\Delta$), the number of elements of $U_\gamma\setminus U^+$ for which $p(\delta)|_1$ exceeds n_δ is given by the term $\sum\{x_{\gamma,p}:p\in\mathbf{P}\text{ s.t. }p(\delta)|_1>n_\delta\}$. Thus:

$$\bigwedge_{\gamma\in\Gamma}\bigwedge_{\delta\in\Delta}\left(\overrightarrow{t}(\delta)\geq 1\vee\sum\{x_{\gamma,p}:p\in\mathbf{P}\text{ s.t. }p(\delta)|_1>n_\delta\}=0\right). \quad (5.15)$$

Similarly, taking $m_{\gamma,\delta}$ to be the number of δ-receptive elements of U^*_γ (which can be computed from $\mathbf{L}^+_\Delta$ and $\mathbf{L}^\circ_\Delta$), we have:

$$\bigwedge_{\gamma\in\Gamma}\bigwedge_{\delta\in\Delta}\Big(\overleftarrow{s}_\gamma(\delta)\geq 1 \vee \sum\{y_r : r\in\mathbf{R} \text{ s.t. } r(\gamma,\delta)|_1 > m_{\gamma,\delta}\} = 0\Big). \tag{5.16}$$

We note in this connection that, by construction of the sets U^*_γ, if $U_\gamma\setminus U^*_\gamma$ contains *any* δ-receptive vertices, then it contains at least $M|\Delta|(|\Delta|+1)$; and similarly for $V\setminus V^*$. Thus, we have the conditions:

$$\bigwedge_{\gamma\in\Gamma}\bigwedge_{\delta\in\Delta}\Big(\overleftarrow{s}_\gamma(\delta)=0 \vee \overleftarrow{s}_\gamma(\delta)\geq M|\Delta|(|\Delta|+1)\Big) \tag{5.17}$$

$$\bigwedge_{\delta\in\Delta}\Big(\overrightarrow{t}(\delta)=0 \vee \overrightarrow{t}(\delta)\geq M|\Delta|(|\Delta|+1)\Big). \tag{5.18}$$

So far, we have made no use of the fact that, since H is a solution of $\mathcal{Q}$, then, for all $\delta\in\Delta_0$, E_δ is non-empty. To do so, we define some useful abbreviations, gathering additional conditions on $\mathcal{D}$ along the way. Observe first that the following constants specify the number of vertices in U^*_γ and $U^+_\gamma\setminus U^*_\gamma$, respectively, lying on some edge in $E^+_\delta\cup E^\circ_\delta$:

$$s^*_\gamma(\delta) = |\{u\in U^*_\gamma : \bar{p}_u(\delta)>0\}|$$
$$s^+_\gamma(\delta) = |\{u\in U^+_\gamma\setminus U^*_\gamma : \bar{p}_u(\delta)>0\}|.$$

But, since H is M-proper (and $M\geq 1$), these are the numbers of vertices in U^*_γ and $U^+_\gamma\setminus U^*_\gamma$, respectively, lying on some edge in E_δ. The following term in the variables $x_{\gamma,p}$ likewise specifies the number of vertices in $U_\gamma\setminus U^+_\gamma$ lying on some edge in E_δ:

$$s_\gamma(\delta) = \sum\{x_{\gamma,p} \mid p\in\mathbf{P}, \bar{p}(\delta)>0\}.$$

Analogous expressions, $t^*(\delta)$, $t^+(\delta)$ and $t(\delta)$, can be constructed to count how many vertices of V^*, $V^+\setminus V^*$ and $V\setminus V^+$, respectively, lie on some edge in E_δ. Hence, the terms

$$\hat{s}(\delta) = \sum_{\gamma\in\Gamma}(s^*_\gamma(\delta)+s^+_\gamma(\delta)+s_\gamma(\delta))$$
$$\hat{t}(\delta) = t^*(\delta)+t^+(\delta)+t(\delta)$$

denote the number of elements of U and V, respectively, lying on some edge in E_δ.

Recall that U^* is guaranteed to contain all special elements of U—i.e. elements with $\deg_u(\delta)>0$ for which at most one other element satisfies $\deg_u(\delta)>0$. Put another way: if at most two elements $u\in U$ satisfy $\deg_u(\delta)>0$, then no elements $u\in U\setminus U^*$ do:

$$\bigwedge_{\delta\in\Delta}\left(\sum_{\gamma\in\Gamma}(s^+_\gamma(\delta)+s_\gamma(\delta))=0 \vee \sum_{\gamma\in\Gamma}(s^*_\gamma(\delta)+s^+_\gamma(\delta)+s_\gamma(\delta))>2\right) \tag{5.19}$$

And similarly for V^*:

$$\bigwedge_{\delta\in\Delta}\Big((t^+(\delta)+t(\delta)=0)\vee(t^*(\delta)+t^+(\delta)+t(\delta)>2)\Big). \tag{5.20}$$

Now we can state the condition on $\mathcal{D}$ arising from the fact that, for all $\delta \in \Delta_0$, E_δ is non-empty. We simply write

$$\bigwedge_{\delta \in \Delta_0} \left(\hat{t}(\delta) > 0\right). \tag{5.21}$$

So far, we have made no use of the fact that, since H is a solution of $\mathcal{Q}$, if $e \in E_\delta$ and $e' \in E_{\delta'}$ are skew, then $(\delta, \delta') \in X$. Recalling the terms $\hat{s}(\delta)$ and $\hat{t}(\delta)$, we can evidently write an analogous linear term $\hat{s}(\delta, \delta')$, in the variables $x_{\gamma,p}$, specifying the number of vertices in U lying on *either* δ*- or* δ'-edges, with a corresponding term $\hat{t}(\delta, \delta')$ for V. Writing $L_\delta = L^+_\delta \cup L^-_\delta \cup L^\circ_\delta$ for $\delta \in \Delta$, and similarly for δ', Lemma 5.9 yields the following pair of conditions on $\mathcal{D}$:

$$\bigwedge_{(\delta,\delta) \in \Delta^2 \setminus X} (\hat{s}(\delta) \leq 1 \vee \hat{t}(\delta) \leq 1) \tag{5.22}$$

$$\bigwedge_{\substack{(\delta,\delta') \in \Delta^2 \setminus X \\ \delta \neq \delta'}} \left\{ \begin{array}{l} \left[\hat{s}(\delta) = 0\right] \vee \left[\hat{s}(\delta') = 0\right] \vee \left[\hat{s}(\delta, \delta') = 1\right] \vee \left[\hat{t}(\delta, \delta') = 1\right] \vee \\ \\ \left[\hat{s}(\delta, \delta') = \hat{t}(\delta, \delta') = 2 \text{ and } L_\delta,\ L_{\delta'} \text{ are as in Fig. 5.2}\right] \end{array} \right\}. \tag{5.23}$$

This completes the list of conditions on $\mathcal{D}$. We have shown that, if $\mathcal{Q}$ has a solution $H = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$, then there exists a quasi-certificate $\mathcal{D} = (\mathbf{U}^*_\Gamma, \mathbf{U}^+_\Gamma, V^*, V^+, \mathbf{L}^+_\Delta, \mathbf{L}^-_\Delta, \mathbf{L}^\circ_\Delta, \mathbf{p}_{U^+}, \mathbf{r}_{V^+})$, such that the conditions (5.6)–(5.23) can be satisfied by choosing appropriate values (over $\mathbb{N}^*$) for the variables $x_{\gamma,p}$ and y_r. A quick scan of these conditions (and of the abbreviations they contain) shows that—regarding the symbols $x_{\gamma,p}$ and y_r as variables, and all others as constants—they are all Boolean combinations of linear equations and inequalities. We have already observed that the cardinalities of the sets U^+_γ and V^+ are bounded by $(2|\Delta| + M|\Delta|^2(|\Delta| + 1))(M|\Delta| + 1)$; thus, by scanning the index sets over which any conjunctions or disjunctions occurring in (5.6)–(5.23), range, we see that the number of linear equations and inequalities involved is bounded by a polynomial function of the size of Γ, $|\Delta|$ and M. Finally, all constant terms—such as, for example, numbers n_δ or the function values $p_u(\delta)$ (for $u \in U^+$)—are also evidently bounded by a polynomial function of the Γ, $|\Delta|$ and M. We remark that the number of variables appearing in these conditions—and hence their total size—is not so bounded. This fact necessitates the reasoning in Stage 4, below.

Stage 3: Now suppose we have a quasi-certificate

$$\mathcal{D} = (\mathbf{U}^*_\Gamma, \mathbf{U}^+_\Gamma, V^*, V^+, \mathbf{L}^+_\Delta, \mathbf{L}^-_\Delta, \mathbf{L}^\circ_\Delta, \mathbf{p}_{U^+}, \mathbf{r}_{V^+})$$

where $\mathbf{U}^*_\Gamma$ and $\mathbf{U}^+_\Gamma$ are collections of sets satisfying $U^*_\gamma \subseteq U^+_\gamma$ for all $\gamma \in \Gamma$; V^* and V^+ are sets satisfying $V^* \subseteq V^+$; $\mathbf{L}^+_\Delta$, $\mathbf{L}^-_\Delta$ and $\mathbf{L}^\circ_\Delta$ are collections of edge-sets in $\Omega = (U^+ \times V^*) \cup (U^* \times V^+)$ such that $H^- = (\mathbf{U}^+_\Gamma, V^+, \mathbf{L}^+_\Delta,\ \mathbf{L}^-_\Delta,\ \mathbf{L}^\circ_\Delta)$ is a Γ-partitioned, directed Δ-graph; $\mathbf{p}_{U^+}$ is a collection of functions $p_u \in P$; and $\mathbf{r}_{V^+}$ is a collection of functions $r_v \in R$. And suppose Conditions (5.6)–(5.23) can be satisfied over $\mathbb{N}^*$. We show that there exists a solution $H = (U, V, \mathbf{E}^+_\Delta, \mathbf{E}^-_\Delta, \mathbf{E}^\circ_\Delta)$ of $\mathcal{Q}$. To this end, we henceforth take the $x_{\gamma,f}$ and y_g to be elements of $\mathbb{N}^*$ such that (5.6)–(5.23) hold. For all $\gamma \in \Gamma$ and all $p \in \mathbf{P}$, let $U_{\gamma,p}$ be a fresh set of cardinality $x_{\gamma,p}$;

let $U_\gamma = U^+_\gamma \cup \bigcup_{\mathbf{P}} U_{\gamma,p}$; and let $U = \bigcup_\Gamma U_\gamma$. For all $r \in \mathbf{R}$, let V_r be a fresh set of cardinality y_r; and let $V = V^+ \cup \bigcup_{\mathbf{R}} V_r$. As usual, we set $U^* = \bigcup_{\gamma \in \Gamma} U^*_\gamma$ and $U^+ = \bigcup_\Gamma U^+_\gamma$. When $u \in U_{\gamma,p}$, we take p_u to denote p, and when $v \in V_r$, we take r_v to denote r. In this way, the notation p_u makes sense for all $u \in U$, and the notation r_v makes sense for all $v \in V$. If $u \in U$ and $\delta \in \Delta$, we think of $p_u(\delta)$ (a pair of integers) as the 'desired' value of $\mathrm{Deg}^H_u(\delta)$ when H is finally constructed; and if $v \in V$, $\gamma \in \Gamma$ and $\delta \in \Delta$, we think of $r_v(\gamma, \delta)$ as the 'desired' value of $\mathrm{DEG}^H_v(\gamma, \delta)$ when H is finally constructed. Accordingly, we call $u \in U$ *δ-receptive* if $\bar{p}_u(\delta) = M$, and we call $v \in V$ *δ-receptive* if $\bar{\bar{r}}_v(\delta) = M$.

Our task is to define the collections of edge-sets $\mathbf{E}^+_\Delta$, $\mathbf{E}^-_\Delta$ and $\mathbf{E}^\circ_\Delta$. We begin by setting $\mathbf{E}^+_\Delta$, $\mathbf{E}^-_\Delta$ and $\mathbf{E}^\circ_\Delta$ on the pairs in Ω to coincide exactly with $\mathbf{L}^+_\Delta$, $\mathbf{L}^-_\Delta$ and $\mathbf{L}^\circ_\Delta$, respectively. In the sequel, if $u \in U^*$, we shall not add any edges (u, v) to any of the edge-sets E^+_δ or E°_δ; likewise, if $v \in V^*$, we shall not add any edges (u, v) to any of the edge sets E^-_δ or E°_δ. In this way, using (5.8) and (5.9), we ensure that $\mathrm{Deg}^H_u = p_u \in P$ for all $u \in U^*$, and $\mathrm{DEG}^H_v = r_v \in R$ for all $v \in V^*$. The remainder of the construction is concerned with extending the definition of $\mathbf{E}^+_\Delta$, $\mathbf{E}^-_\Delta$ and $\mathbf{E}^\circ_\Delta$ to the whole of $U \times V$.

We begin with the collection of bi-directional edge sets, $\mathbf{E}^\circ_\Delta$. Fix $\gamma \in \Gamma$. Now associate with each $u \in U_\gamma$ exactly $p_u(\delta)|_2$ bi-directional δ edges, and associate with each $v \in V$ exactly $r_v(\gamma, \delta)|_2$ bi-directional δ edges. We think of $u \in U_\gamma$ as having $p_u(\delta)|_2$ 'dangling' δ-edges which need to be paired up with dangling edges belonging to vertices in V; and we think of $v \in V$ as having $r_v(\gamma, \delta)|_2$ dangling δ-edges which need to be paired up with dangling edges belonging to vertices in U_γ. By (5.12), the total number of δ-edges left dangling by vertices in U_γ is the same as the total number left dangling by the vertices in V, and so we can put these dangling edges in a 1–1 correspondence; indeed, this may obviously be done consistently with the partial correspondence induced by L°_δ. We then simply take (u, v) to be in E°_δ just in case u and v are associated with dangling δ-edges that have been paired up in this process. (Note that E°_δ agrees with L°_δ on Ω.) For this assignment to make sense, we must check that vertices $u \in U_\gamma$ and $v \in V$ cannot be paired twice in this process. After all, if $u \in U_\gamma$ and $v \in V$ were both associated with one (dangling) δ-edge and one (dangling) δ'-edge, we could not use both dangling pairs to form two edges in the graph, since then E°_δ and $E^\circ_{\delta'}$ would not be disjoint. However, no such double pairings can arise, because r_v is, by assumption, unitary: v never 'wants' to be linked by more than one bi-directional edge (regardless of colour) to vertices in U_γ. (Indeed, this was the point of introducing the notion of *partitioned* directed Δ-graphs in the first place.) Carrying out this process for all $\gamma \in \Gamma$, we have set $\mathbf{E}^\circ_\Delta$ so as to ensure that

$$\begin{aligned} \mathrm{Deg}^H_u(\delta)|_2 &= p_u(\delta)|_2 \\ \mathrm{DEG}^H_v(\gamma, \delta)|_2 &= r_v(\gamma, \delta)|_2 \end{aligned}$$

for all $u \in U$ and $v \in V$.

We now turn to the uni-directional edges in H. As a prelude, we use Lemma 5.14 to partition the sets $V \setminus V^*$ and $U_\gamma \setminus U^*_\gamma$ (for $\gamma \in \Gamma$) into sets of 'positive' and 'negative' elements. Suppose $\delta \in \Delta$. Now, it follows from (5.18) that, if there are any δ-receptive elements of $V \setminus V^*$ at all, then there are at least $M|\Delta|(|\Delta| + 1)$ of them. By Lemma 5.14, therefore, putting $m = M|\Delta|$ and $n = |\Delta|$, we may divide $V \setminus V^*$ into sets of positive and negative elements such that, for all $\delta \in \Delta$, if there are any δ-receptive elements of $V \setminus V^*$, then there are at least $M|\Delta|$ positive such elements, and at least $M|\Delta|$ negative such elements. Similarly, by (5.17), we may divide each

$U_\gamma \setminus U^*_\gamma$ ($\gamma \in \Gamma$) into sets of positive and negative elements such that, for all $\delta \in \Delta$, if there are any δ-receptive elements of $U_\gamma \setminus U^*_\gamma$, then there are at least $M|\Delta|$ positive such elements, and at least $M|\Delta|$ negative such elements.

We are now ready to define the collection of left-to-right edge sets, $\mathbf{E}^+_\Delta$. We have already dealt with the elements of U^*_γ, so consider first any element $u \in U^+_\gamma \setminus U^*_\gamma$. For each $\delta \in \Delta$, we have two cases, depending on the condition $\mathrm{Deg}^{H^-}_u(\delta)|_1 = p_u(\delta)|_1$. If this condition holds, then, for each $v \in V$, we simply take (u, v) to be in E^+_δ just in case $(u, v) \in L^+_\delta$. This does not change any previously made assignments, and will result in the condition that $\mathrm{Deg}^H_u(\delta)|_1 = p_u(\delta)|_1$. If, on the other hand, $\mathrm{Deg}^{H^-}_u(\delta)|_1 < p_u(\delta)|_1$, then, by (5.13), $V \setminus V^*$ contains some δ-receptive elements, whence, as we have just argued, we can find $M|\Delta|$ such elements that are positive, and also $M|\Delta|$ that are negative. If u is positive (negative), we can therefore choose $p_u(\delta)|_1 - \mathrm{Deg}^{H^-}_u(\delta)|_1$ positive (negative) $v \in V \setminus V^*$ such that (u, v) has not so far been assigned to any edge, and simply make the assignment $(u, v) \in E^+_\delta$. It is obvious that, at the end of this process, $\mathrm{Deg}^H_u(\delta)|_1 = p_u(\delta)|_1$. Suppose, finally, $u \in U_\gamma \setminus U^+_\gamma$. Again, we have two cases, depending on the condition $n_\delta \geq p_u(\delta)|_1$. If this condition holds, then, for each $v \in V$, we can find $p_u(\delta)|_1$ δ-receptive $v \in V^*$, and simply make the assignment $(u, v) \in E^+_\delta$. Since $u \notin U^+_\gamma$, this cannot disturb any previously made assignments. If, on the other hand, $n_\delta < p_u(\delta)|_1$, then, by (5.15), $V \setminus V^*$ contains some δ-receptive elements, and hence at least $M|\Delta|$ positive such elements and at least $M|\Delta|$ negative such elements. Again, if u is positive (negative), we choose $p_u(\delta)|_1$ positive (negative) $v \in V \setminus V^*$ such that (u, v) has not so far been assigned to any edge, and simply make the assignment $(u, v) \in E^+_\delta$. When all these assignments have been made, we have $\mathrm{Deg}^H_u(\delta)|_1 = p_u(\delta)|_1$. At this point, E^+_δ has been completely defined for all $\delta \in \Delta$ in such a way that $\mathrm{Deg}^H_u(\delta)|_1 = p_u(\delta)|_1$ for all $u \in U$. Since the definition of $\mathbf{E}^\circ_\Delta$ has already secured $\mathrm{Deg}^H_u(\delta)|_2 = p_u(\delta)|_2$ for all $u \in U$, we have $\mathrm{Deg}^H_u(\delta) = p_u(\delta)$. If $u \in U^+$, the fact that $p_u \in P$ ensures that $\mathrm{Deg}^H_u \in P$; if $u \in U \setminus U^+$, the same conclusion follows from (5.10).

To define the collection of right-to-left edge sets, $\mathbf{E}^-_\Delta$, we proceed in an analogous way, relying on Condition (5.14) instead of (5.13), and on Condition (5.16) instead of (5.15). There is one small difference, however. If $v \in V \setminus V^*$ is positive (negative) we choose only *negative* (*positive*) elements of $U_\gamma \setminus U^*_\gamma$ to receive right-to-left edges from v. Thus, while left-to-right edges link positive Us to positive Vs and negative Us to negative Vs, right-to-left edges link positive Us to *negative* Vs and negative Us to *positive* Vs. This strategy prevents the assignments of right-to-left edges disturbing the earlier left-to-right assignments. At the end of this process, we have $\mathrm{DEG}^H_v = r_v$ for all $v \in V$. If $v \in V^+$, the fact that $r_v \in R$ ensures that $\mathrm{DEG}^H_v \in R$; if $v \in V \setminus V^+$, the same conclusion follows from (5.11).

That $E_\delta = E^+_\delta \cup E^-_\delta \cup E^\circ_\delta \neq \emptyset$ for all $\delta \in \Delta_0$ follows easily from Conditions (5.21); and that there exists no pair of skew edges $e \in E_\delta$, $e' \in E_{\delta'}$ with $(\delta, \delta') \notin X$ follows from Conditions (5.19), (5.20), (5.22) and (5.23), using Lemma 5.9. (Notice that Conditions (5.19) and (5.20) are needed to ensure that, if there are only two vertices in U lying on δ- or δ'-edges, and only two vertices in V lying on δ- or δ'-edges, then all the δ- and δ'-edges are accounted for by H^-.) Thus, H is a solution of $\mathcal{Q}$, as required.

Stage 4: To complete the proof, suppose that $\mathcal{D}$ exists and satisfies Conditions (5.6)–(5.23). These conditions are simply a Boolean combination (involving $\wedge$ and $\vee$) of easily checkable statements about $\mathcal{Q}$—let us call them $\mathcal{Q}$-statements—and linear equations and inequalities in the variables $x_{\gamma,p}$ and y_r. Select a single disjunct from each disjunction so that a simple conjunction results. Now verify the truth of all the $\mathcal{Q}$-statements in this conjunction (failing if any is false); and let $\mathcal{E}$ be the remaining conjunction of linear equations and inequalities. Thus, $m = |\mathcal{E}|$ is bounded by a polynomial function of M and Δ, and each coefficient in $\mathcal{E}$ certainly has at most k bits, where k is given by a polynomial function of M and Δ. By Corollary 5.4, if $\mathcal{E}$ has a solution over $\mathbb{N}^*$, then it has a solution in which at most polynomially many values are non-zero (as a function of km). The relevant set of non-zero values may be guessed and written down in polynomial time, and all other variables ignored. Thus, from Conditions (5.6)–(5.23), we can non-deterministically construct an equisatisfiable, polynomial-sized integer-programming problem. But Corollary 5.5 states that this problem is in NPTime.

To show that finite PDBGE is in NPTime, we reason in exactly the same way, but with $\mathbb{N}^*$ replaced by $\mathbb{N}$, and Corollaries 5.4 and 5.5 replaced by Propositions 5.2 and 5.3, respectively. The details of the proof are unaffected. □

The proof of Lemma 5.15 actually shows a little more:

Corollary 5.16. *Let $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P, R, X)$ be a (finite) PDBGE-instance. If $\mathcal{Q}$ has a solution, then we can find subsets $P_0 \subseteq P$ and $R_0 \subseteq R$, bounded by a polynomial function of $|\Gamma|$, $|\Delta|$ and M, such that the (finite) PDBGE-instance $(\Gamma, \Delta, \Delta_0, M, P_0, R_0, X)$ also has a solution.*

Proof. Let P_0 be the set of functions $p \in P$ for which either $p \in \mathbf{p}_{U^+}$ or $x_{\gamma,p}$ is non-zero (for some γ) in the proof of Lemma 5.15; and similarly let R_0 be the set of functions $r \in R$ for which either $r \in \mathbf{r}_{V^+}$ or y_r is non-zero. □

We are now able to establish Theorem 5.10, the main result of this section.

Proof. [Theorem 5.10] Let the (finite) BGESC-instance $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$ be given. We carry out the following procedure, where h is some fixed polynomial. Guess subsets $P_0 \subseteq \mathbf{P}$ and $R_0 \subseteq \mathbf{R}$ of cardinality at most $h(M|\Delta|)$, and determine whether

$$\dot{p} \in F \qquad \text{for all } p \in P_0 \tag{5.24}$$
$$\bar{\dot{r}} \in G \qquad \text{for all } r \in R_0 \tag{5.25}$$

failing if not. Now let Γ be a set of cardinality $M^2|\Delta|^2$, and run a non-deterministic polynomial time algorithm which succeeds just in case the (finite) PDBGE-instance $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P_0, Q_0, X)$ is positive, and report the result.

The above non-deterministic procedure obviously runs in polynomial time. We claim that, for suitable choice of the polynomial h, it has a successful run if and only if $\mathcal{P}$ is positive. For suppose the procedure has a successful run. Let the Γ-partitioned, directed Δ-graph H be a solution of $\mathcal{Q}$. Then the conditions (5.24) and (5.25), together with Lemmas 5.11 and 5.12 ensure that, setting $H' = \dot{H}$ and $H'' = \overline{H'}$, the Δ-graph H'' is a solution of $\mathcal{P}$. Conversely, suppose $\mathcal{P}$ is positive, and let the Δ-graph H'' be a solution of $\mathcal{P}$. By Lemma 5.11, there is an M-bounded, M-proper directed Δ-graph H' such that $\overline{H'} = H''$; and by Lemma 5.12, there exists a set Γ with $|\Gamma| \leq M^2|\Delta|^2$ and a unitary (M-bounded, M-proper) Γ-partitioned directed directed Δ-graph H such that $\dot{H} = H'$. Now define

$$P = \{p \in \mathbf{P} \mid \dot{p} \in F\}$$
$$R = \{r \in \mathbf{R} \mid \bar{\dot{r}} \in G\}$$

Thus, H is a solution of the (finite) PDBGE-instance $(\Gamma, \Delta, \Delta_0, M, P, R, X)$. Hence, for suitable choice of h, Corollary 5.16 ensures that we can find $P_0 \subseteq P$ and $R_0 \subseteq R$, with cardinalities bounded by $h(M, |\Delta|)$, such that H is a solution of the (finite) PDBGE-instance $\mathcal{Q} = (\Gamma, \Delta, \Delta_0, M, P_0, R_0, X)$. But then the above procedure has a successful run, as required. □

Using the same reasoning as for Corollary 5.16, we have:

Corollary 5.17. *If* $(\Delta, \Delta_0, M, F', G', X)$ *is a positive instance of* (*finite*) *BGESC, then there exist subsets* $F \subseteq F'$, $G \subseteq G'$, *both of cardinality bounded by a polynomial function* h_0 *of* $|\Delta|$ *and* M, *such that* $(\Delta, \Delta_0, M, F, G, X)$ *is also a positive instance of* (*finite*) *BGESC.*

6. Upper Bound for EC_2^2. The purpose of this section is to establish that the satisfiability and finite satisfiability problems for EC_2^2 are both in 2-NExpTime. We proceed by transforming a reduced normal-form EC_2^2-formula φ, non-deterministically, into a BGESC-instance, $\mathcal{P}$, and showing that φ is (finitely) satisfiable if and only if this transformation can be carried out in such a way that $\mathcal{P}$ is a positive instance of (finite) BGESC. Any solution of $\mathcal{P}$ is a bipartite graph in which the left-hand vertices represent $r_1^{\#}$-classes, the right-hand vertices represent $r_2^{\#}$-classes and the edges represent intersections; incidence of an edge on a vertex represents inclusion of the corresponding intersection in the corresponding $r_1^{\#}$- or $r_2^{\#}$-class. Owing to Lemma 4.2 we may restict our attention to intersections of exponentially bounded size. The main work in this reduction is performed in Sec. 6.2; Sec. 6.1 is devoted to establishing technical results allowing us to manipulate structures built from collections of intersections. We introduce some additional notation. If $\tau = \tau_0 \cup \{r_1, r_2\} \cup \{r_1^{\#}, r_2^{\#}\}$, we say that a τ-structure $\mathfrak{I}$ is a *pre-intersection* if for $i = 1, 2$, and for all $a, a' \in I$ we have $\mathfrak{I} \models r_i^{\#}[a, a']$ (but we do not require $(r_i^{\#})^{\mathfrak{I}}$ to be the equivalence closure of $r_i^{\mathfrak{I}}$). Obviously, if I is an intersection of $\mathfrak{A}$, then the induced substructure $\mathfrak{I}$ is a pre-intersection. By the *type* of a pre-intersection, we mean its isomorphism type.

Let Δ be a set of types of pre-intersections, and $f : \Delta \to \mathbb{N}^*$ a function not uniformly 0 on Δ. We write $\mathfrak{D} \approx [\![f]\!]_1$ to indicate that the structure $\mathfrak{D}$ is a single $r_1^{\#}$-class built out of exactly $f(\delta)$ pre-intersections of type δ, for each $\delta \in \Delta$. More precisely: (i) the domain of $\mathfrak{D}$ can be represented as $D = \bigcup\{D_{\delta,i} \mid \delta \in \Delta,\ 0 \le i < f(\delta)\}$; (ii) for all $\delta \in \Delta$ and all $i < f(\delta)$, $\mathfrak{D}\restriction D_{\delta,i}$ is a pre-intersection of type δ; (iii) every pair of elements of D is r_1-connected in $\mathfrak{D}$; (iv) $r_1^{\#}$ is the equivalence closure of r_1; (v) no elements from different sets $D_{\delta,i}$ are related by r_2. Note that a pair of elements belonging to a single pre-intersection is not required to be connected by an r_2-path in $\mathfrak{D}$ (in a model containing $\mathfrak{D}$ as an $r_1^{\#}$-class such a pair may be properly connected by an r_2-path going through some other pre-intersections of its $r_2^{\#}$-class). The notation $\mathfrak{D} \approx [\![f]\!]_2$ is defined symmetrically, with r_1 and r_2 exchanged. Observe that f does not fully determine $\mathfrak{D}$, since the connections (i.e. 2-types) between elements from different pre-intersections are not specified.

6.1. Approximating Classes. Fix a reduced normal-form EC_2^2-formula $\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \omega$ over signature τ. We take φ_1 to denote $\chi \wedge \psi_{00} \wedge \psi_{01}$, and φ_2 to denote $\chi \wedge \psi_{00} \wedge \psi_{10}$. Thus, φ_1 incorporates the universal requirements of φ, as well as its existential requirements in respect of the relation $r_1^{\#}$; similarly, *mutatis mutandis*, for φ_2. We employ the exponential function $K : \mathbb{N} \to \mathbb{N}$ of Lemma 4.2. In addition, we take $N : \mathbb{N} \to \mathbb{N}$ to be a doubly exponential function such that $N(|\tau|)$ bounds number of isomorphism types of τ-structures consisting of two pre-intersections of size at most $K(|\tau|)$. We define the function $L(n) = 45(N(n))^6$, corresponding to the size bound obtained in Lemma 4.3. We prove two simple facts regarding the $r_i^{\#}$-classes in a model of φ. The first allows us to add pre-intersections to an existing $r_1^{\#}$- or $r_2^{\#}$-class, provided that, for each pre-intersection being added, its type is realized in this class at least twice.

Lemma 6.1. *Let Δ be a finite set of isomorphism types of pre-intersections. Let f and f' be functions $\Delta \to \mathbb{N}^*$, such that, for all $\delta \in \Delta$, $f(\delta) \leq 1$ implies $f'(\delta) = f(\delta)$, and $f(\delta) \geq 2$ implies $f'(\delta) \geq f(\delta)$. For $i \in \{1, 2\}$, if $\mathfrak{D} \approx [\![f]\!]_i$ is such that $\mathfrak{D} \models \varphi_i$, then there exists $\mathfrak{D}' \approx [\![f']\!]_i$ such that $\mathfrak{D}' \models \varphi_i$.*

Proof. We prove the result for $i = 1$; the case $i = 2$ follows by symmetry. Consider first the case where, for some $\delta \in \Delta$, $f'(\delta) = f(\delta)+1$, with $f'(\delta') = f(\delta')$ for all $\delta' \neq \delta$. By assumption, $f(\delta) \geq 2$. We show how to add to $\mathfrak{D}$ a single pre-intersection of type δ to obtain a model $\mathfrak{D}' \models \varphi_1$. Let I_1, I_2 be pre-intersections in $\mathfrak{D}$ of type δ; and let $\mathfrak{D}'$ extend $\mathfrak{D}$ by a new pre-intersection I of type δ. For every pre-intersection I' of $\mathfrak{D}$, $I' \neq I_1$, set the 2-types between I and I', i.e. the 2-types realized by pairs of elements from, respectively, I and I', isomorphically to the connection between I_1 and I'. This ensures all the required witnesses for I inside $\mathfrak{D}'$, and, as I_1 has to be r_1-connected to the remaining part of $\mathfrak{D}$, this also makes $\mathfrak{D}'$ r_1-connected. Complete $\mathfrak{D}'$ by setting the connection between I and I_1 isomorphically to the connection between I_1 and I_2. Note that all 2-types in $\mathfrak{D}'$ are also realized in $\mathfrak{D}$, so $\mathfrak{D}' \models \chi$. Observe that, in this construction, $\mathfrak{D} \subseteq \mathfrak{D}'$.

Consider now the case where, for some $\delta \in \Delta$, $f'(\delta) > f(\delta) \geq 2$, with $f'(\delta') = f(\delta')$ for all $\delta' \neq \delta$. If $f'(\delta)$ is finite, iterating the above procedure $f'(\delta) - f(\delta)$ times yields the required $\mathfrak{D}'$. If $f'(\delta) = \aleph_0$, we define a sequence $\mathfrak{D}_1 \subseteq \mathfrak{D}_2 \subseteq \cdots$ of models of φ_1 with increasing numbers of copies of pre-intersections of type δ, and set $\mathfrak{D}' = \bigcup_i \mathfrak{D}_i$. The statement of the lemma is then obtained by applying the above construction successively for all $\delta \in \Delta$. □

In the next lemma we show that, from a local point of view, every class can be 'approximated' by a class in which the number of realizations of each pre-intersection type is bounded doubly exponentially in τ. (In fact, exponentially many realizations of each type suffice; however, a doubly exponential bound makes for a simpler proof.) This lemma is a counterpart of Lemma 16 from [21].

Lemma 6.2. *Let Δ be the set of all types of pre-intersections of size bounded by $K(|\tau|)$. Let f be a function $\Delta \to \mathbb{N}^*$, and let $f' = \lfloor f \rfloor_{L(|\tau|)}$. For $i \in \{1, 2\}$, if $\mathfrak{D} \approx [\![f]\!]_i$ is such that $\mathfrak{D} \models \varphi_i$, then there exists $\mathfrak{D}' \approx [\![f']\!]_i$ such that $\mathfrak{D}' \models \varphi_i$.*

Proof. Again, we prove the result for $i = 1$; the case $i = 2$ follows by symmetry. We translate $\mathfrak{D}$ into a structure $\mathfrak{F}$ whose domain is the set of all pre-intersections of $\mathfrak{D}$; atomic 1-types in $\mathfrak{D}$ represent isomorphism types of pre-intersections, and atomic 2-types represent connections among them. The signature σ of $\mathfrak{F}$ contains a binary symbol r_1', corresponding to r_1 from τ, a dummy binary symbol r_2' and some sets of unary and binary predicates bounded logarithmically in $N(|\tau|)$. We build $\mathfrak{F}$ in such a way that: (i) I_1, I_2 have the same 1-type in $\mathfrak{F}$ if and only if I_1 and I_2 are isomorphic in $\mathfrak{D}$; (ii) pairs of pre-intersections I_1, I_2 and I_1', I_2' have the same 2-types in $\mathfrak{F}$ if and only if $\mathfrak{D}\restriction(I_1 \cup I_2)$ is isomorphic to $\mathfrak{D}\restriction(I_1' \cup I_2')$; (iii) $\mathfrak{F} \models r_1'(I_1, I_2)$ if and only if there exist $a_1 \in I_1$, $a_2 \in I_2$ such that $\mathfrak{D} \models r_1(a_1, a_2)$; and (iv) r_2' is the universal relation: $\mathfrak{F} \models r_2'[I_1, I_2]$ for all $I_1, I_2 \in F$. Note that $\mathfrak{F}$ is r_1'-connected, and thus forms a single ${r_1'}^{\#}$-class, and, as ${r_2'}^{\#}$ is universal, $\mathfrak{F}$ is actually an intersection. Note also that $|\boldsymbol{\beta}[\mathfrak{F}]|$, i.e. the number of 2-types in $\mathfrak{F}$, is bounded by $N(|\tau|)$.

Let α be a 1-type realized in $\mathfrak{F}$. Let F_α be the set of realizations of α. If $|F_\alpha| > 45|\boldsymbol{\beta}[\mathfrak{F}]|^6$ then apply Lemma 4.3, taking $\mathfrak{A} := \mathfrak{F}$, $B := F_\alpha$, $D_1 := D_2 := F$. Repeat this step for all 1-types of $\mathfrak{F}$. Let $\mathfrak{F}'$ be the structure thus obtained.

Since, by Lemma 4.3 (ii) and (iii), no new 1-types or 2-types can appear in $\mathfrak{F}'$, it has a natural translation back into a structure $\mathfrak{D}''$, with elements of $\mathfrak{F}'$ corresponding to pre-intersections in $\mathfrak{D}''$. Thus, each isomorphism type δ is realized in $\mathfrak{D}''$ at most $45|\boldsymbol{\beta}[\mathfrak{F}]|^6 \leq L(|\tau|)$ times. If δ is realized fewer than $\min(f(\delta), L(|\tau|))$ times in $\mathfrak{D}''$, then we can use Lemma 6.1 to add an appropriate number of realizations of δ to $\mathfrak{D}''$ to obtain a model $\mathfrak{D}' \models \varphi_1$ with $\mathfrak{D}' \approx [\![f']\!]_1$. □

6.2. The (Finite) Satisfiability Problem for EC_2^2 and (Finite) BGESC. Let $\varphi, \varphi_1, \varphi_2, \tau$ and the function L be as in Sec. 6.1. (Recall: $\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{01} \wedge \omega$, $\varphi_1 = \chi \wedge \psi_{00} \wedge \psi_{01}$ and $\varphi_2 = \chi \wedge \psi_{00} \wedge \psi_{10}$.) We now explain how to transform φ non-deterministically into a BGESC-instance $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$. We show that φ is (finitely) satisfiable if and only if this transformation can be applied in such a way that the resulting tuple $\mathcal{P}$ is a positive instance of the problem (finite) BGESC.

We first define the components Δ, M, and X of $\mathcal{P}$. Let Δ be the set of isomorphism types of pre-intersections over the signature τ satisfying $\chi \wedge \psi_{00}$, and of size at most $K(|\tau|)$. Let $M = \max(L(|\tau|), 2)$, and let X be the set of pairs $(\delta, \delta') \in \Delta^2$ for which there exists a model $\mathfrak{D} \models \chi$ consisting of exactly one pre-intersection of type δ and another of type δ', each forming its own $r_1^\#$-class and its own $r_2^\#$-class. Thus, $|\Delta|$, M and $|X|$ are all bounded by a doubly exponential function of $|\tau|$.

The remaining components of $\mathcal{P}$, namely, Δ_0, F and G, will be guessed. The following terminology and notation will prove useful. Say that a set of pre-intersection types $\Delta' \subseteq \Delta$ *certifies* ω if, for every conjunct $\omega_i = \exists x.p_i(x)$ of ω we can find δ in Δ' such that in any structure $\mathfrak{I}$ consisting of a single pre-intersection of type δ there is a such that $\mathfrak{I} \models p_i[a]$. Now let F^* be the set of functions $f : \Delta \to [0, M]$ for which there exists a structure $\mathfrak{D} \approx [\![f]\!]_1$ such that $\mathfrak{D} \models \varphi_1$. Similarly, let G^* be the set of functions $g : \Delta \to [0, M]$ for which there exists a structure $\mathfrak{D} \approx [\![g]\!]_2$ such that $\mathfrak{D} \models \varphi_2$. (Note that $|F^*|$ and $|G^*|$ are bounded by a *triply* exponential function of $|\varphi|$.)

LEMMA 6.3. *Let φ, Δ, F^*, G^*, X be as defined above, and let h_0 be the polynomial guaranteed by Corollary 5.17. Then φ is (finitely) satisfiable if and only if there exist $\Delta_0 \subseteq \Delta$ certifying ω, and collections of functions $F \subseteq F^*$, $G \subseteq G^*$, both of cardinality bounded by $h_0(|\Delta|, M)$, such that $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$ is a positive instance of the problem (finite) BGESC.*

Proof. $\Rightarrow$ By Lemma 4.2, let $\mathfrak{A} \models \varphi$ be a model with intersections bounded by $K(|\tau|)$. Let E be the set of intersections in $\mathfrak{A}$. For each conjunct ω_i of ω choose one element of E satisfying ω_i. Let Δ_0 be the set of isomorphism types of the chosen intersections. Clearly Δ_0 certifies ω. We show that the BGESC-instance $\mathcal{P}^* = (\Delta, \Delta_0, M, F^*, G^*, X)$ is positive. (Of course: F^* and G^* do not satisfy the cardinality bounds of the lemma.) Let U be the set of $r_1^\#$-classes in $\mathfrak{A}$, and V the set of $r_2^\#$-classes. (As before, any 'loner'—i.e., an intersection which is both an $r_1^\#$-class and an $r_2^\#$-class—contributes one element of U and a distinct element of V.) Since each intersection is contained in exactly one $r_1^\#$-class and exactly one $r_2^\#$-class, and indeed is determined by those classes, we may regard the intersections in E as *edges* in a bipartite graph (U, V, E). Denoting by E_δ the set of intersections in E having any type $\delta \in \Delta$, we obtain a Δ-graph $H = (U, V, \{E_\delta\}_\Delta)$. We show that H is a solution of $\mathcal{P}^*$ by checking properties (G1), (G2′), (G3′), (G4) from Section 5.3. Property (G1) is obvious. For (G2′), we show that, for each $\mathfrak{D} \in U$, $\lfloor \mathrm{ord}_\mathfrak{D}^H \rfloor_M \in F^*$. Since $\mathfrak{A} \models \varphi$, and $\mathfrak{D}$ is an $r_1^\#$-class in $\mathfrak{A}$, $\mathfrak{D} \models \varphi_1$; moreover, by definition, $\mathfrak{D} \approx [\![\mathrm{ord}_\mathfrak{D}^H]\!]_1$. Setting $f = \mathrm{ord}_\mathfrak{D}^H$ and $f' = \lfloor f \rfloor_M$, Lemma 6.2 then states that there exists a model $\mathfrak{D}' \models \varphi_1$ such that $\mathfrak{D}' \approx [\![f']\!]_1$. Thus by the definition of F^*, $\lfloor \mathrm{ord}_\mathfrak{D}^H \rfloor_M \in F^*$ as required.

Property (G3′) follows symmetrically. For property (G4), consider any pair (I, I') of skew edges in H, $I \in E_\delta$, $I' \in E_{\delta'}$. Observe that the structure $\mathfrak{A}\restriction(I \cup I')$ consists of two pre-intersections of types δ, δ', each forming its own $r_1^\#$- and $r_2^\#$-class. Thus (δ, δ') is a member of X. Applying Corollary 5.17, we may find $F \subseteq F^*$ and $G \subseteq G^*$, of size bounded by $h_0(|\Delta|, M)$, such that $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$ is a positive instance.

$\Leftarrow$ Assume now that there exist Δ_0 certifying ω, $F \subseteq F^*$ and $G \subseteq G^*$, such that $\mathcal{P} = (\Delta, \Delta_0, M, F, G, X)$ is positive. Let $H = (U, V, \{E_\delta\}_\Delta)$ be an edge-coloured bipartite graph which is a solution of $\mathcal{P}$. Thus, H satisfies (G1), (G2′), (G3′), (G4). We show how to construct a model $\mathfrak{A} \models \varphi$ from the graph H. Intersections of $\mathfrak{A}$ correspond to the edges of H: for each $\delta \in \Delta$ and each $e \in E_\delta$, we put into $\mathfrak{A}$ a pre-intersection I_e of type δ. Property (G1) ensures that $\mathfrak{A} \models \omega$; and the fact that all intersections have types from Δ ensures that $\mathfrak{A} \models \psi_{00}$.

Consider now any vertex $u \in U$. Let $\mathcal{J}$ be the set of all pre-intersections corresponding to the edges incident to u. Our task is to compose from them an $r_1^\#$-class $\mathfrak{D}_u$ satisfying φ_1. First, writing f for ord_u^H and f' for $\lfloor f \rfloor_M$, we form from some subset of $\mathcal{J}$ a class $\mathfrak{D} \approx [\![f']\!]_1$ such that $\mathfrak{D} \models \varphi_1$. This is possible by (G2′) and the construction of F^*. For each of the remaining intersections from $\mathcal{J}$ of type δ, note that the number of intersections of type δ realized in $\mathfrak{D}$ is bigger than $M \geq 2$ and thus the preconditions of Lemma 6.1 are fulfilled. Thus all the remaining intersections of $\mathcal{J}$ can be joined to $\mathfrak{D}$ using Lemma 6.1, forming a desired $\mathfrak{D}_u$. We repeat this construction for all vertices in U. This ensures that $\mathfrak{A} \models \psi_{01}$. It also makes every pre-intersection r_1-connected.

Similarly, from any vertex $v \in V$, we form a $r_2^\#$-class consisting of all pre-intersections corresponding to edges incident on v, using (G3′) and the construction of G. This step ensures that $\mathfrak{A} \models \psi_{10}$ and makes every pre-intersection r_2-connected. Thus, all pre-intersections become both r_1- and r_2-connected; moreover, no two pre-intersections can be connected to each other by both r_1 and r_2 (because no two edges of H can have common vertices in both U and V); hence, every pre-intersection becomes an intersection of $\mathfrak{A}$, as required.

At this point, we have specified the 2-type in $\mathfrak{A}$ of any pair of elements not in free position. To complete the definition of $\mathfrak{A}$, consider a pair of intersections $I_e, I_{e'}$ which are in free position, i.e. are not members of the same $r_1^\#$-class or $r_2^\#$-class. But then the edges e and e' are skew in H. Assume that $e \in E_\delta$ and $e' \in E_{\delta'}$, so that I_e and $I_{e'}$ have respective isomorphism types δ and δ'. By (G4), $(\delta, \delta') \in X$. By the definition of X, there is a structure $\mathfrak{D} \models \chi$ consisting of exactly one intersection of type δ and another of type δ', each forming its own $r_1^\#$-class and its own $r_2^\#$-class. We make $\mathfrak{A}\restriction I_e \cup I_{e'}$ isomorphic to $\mathfrak{D}$. Finally, we point out that each pair of intersections in $\mathfrak{A}$ has been connected by copying the connections between a pair of intersections from a structure which satisfied χ. This ensures that $\mathfrak{A} \models \chi$. □

6.3. Main Theorem. THEOREM 6.4. *The satisfiability and finite satisfiability problems for* EC_2^2 *are in 2-*NEXPTIME.

Proof. Let $\varphi \in \mathrm{EC}_2^2$ be given. By Lemma 3.2, we may assume that $\varphi = \chi \wedge \psi_{00} \wedge \psi_{01} \wedge \psi_{10} \wedge \omega$ is in reduced normal form, since satisfiability of φ over models of at most exponential size can be tested in doubly exponential time. We continue to write φ_1 for $\chi \wedge \psi_{00} \wedge \psi_{01}$, and φ_2 for $\chi \wedge \psi_{00} \wedge \psi_{10}$. Let M, Δ, F^*, G^* and X be as in Sec. 6.2. To determine the (finite) satisfiability of φ', execute the following procedure. Non-deterministically guess a subset $\Delta_0 \subseteq \Delta$, and sets of functions F and G of type $\Delta \to [0, M]$, such that $|F|$ and $|G|$ are bounded by $h_0(|\Delta|, M)$, where h_0 is the polynomial guaranteed by Corollary 5.17. Check, in deterministic doubly exponential time, that Δ_0 certifies ω, and fail if not. For each $f \in F$, guess a structure $\mathfrak{D} \approx [\![f]\!]_1$, and check that $\mathfrak{D} \models \varphi_1$, failing if not; and similarly, for each $g \in G$, guess a structure $\mathfrak{D} \approx [\![g]\!]_2$, and check that $\mathfrak{D} \models \varphi_2$, failing if not. This non-deterministic process runs in doubly exponential time, and has a successful run just in case $F \subseteq F^*$ and $G \subseteq G^*$. Let $\mathcal{P}$ be the BGESC-instance $(\Delta, \Delta_0, M, F, G, X)$; thus the size of $\mathcal{P}$ is bounded doubly exponentially in $|\tau|$. Check the existence of a (finite) solution of $\mathcal{P}$ using the NPTime-algorithm guaranteed by Theorem 5.10, and report the result. This non-deterministic procedure runs in time bounded by a doubly exponential function of $|\varphi|$. By Lemma 6.3, it has a successful run if and only if φ is (finitely) satisfiable. □

The following corollary is an improvement of Theorem 13 of [22].

Corollary 6.5. *Any finitely satisfiable* EC_2^2*-formula* φ *has a model of cardinality at most* $2^{2^{2^{p(\|\varphi\|)}}}$*, for some fixed polynomial* p.

Proof. The proof of Theorem 6.4 constructs a finite model $\mathfrak{A}$ of φ from the solution G of some BGESC-instance $\mathcal{P}$, where $\mathcal{P}$ is of size doubly exponential in $\|\varphi\|$. More specifically, $\mathfrak{A}$ consists of a collection of intersections, each with size bounded by a singly-exponential function of $\|\varphi\|$, and each corresponding to a specific edge of G. We showed in the proof of Theorem 5.10 that $\mathcal{P}$ translates into a system $\mathcal{E}$ of linear equations and inequalities, with the size of G given by the integer solutions of $\mathcal{E}$. From Proposition 5.3, these numbers are all at most triply exponential in $\|\varphi\|$. Hence the number of edges in G is triply exponential in $\|\varphi\|$. □

Of course, the size bound in Corollary 6.5 is insufficient to secure the complexity bound of Theorem 6.4. On the other hand, we know from [21] that it cannot be improved upon: there exists a series φ_n of finitely satisfiable EC_2^2-formulas such that $\|\varphi_n\|$ grows polynomially with n, but the smallest satisfying model of φ_n has at least $2^{2^{2^n}}$ elements.

7. Lower Bound for FO^2 with Two Equivalences. In this section we show that the satisfiability and finite satisfiability problems for EQ_2^2 are both 2-NExpTime-hard. It follows that the satisfiability and finite satisfiability problems for both EQ_2^2 and EC_2^2 are 2-NExpTime-complete. Adapting notation and terminology used above in the natural way, we henceforth assume that the binary predicates r_1 and r_2 are interpreted as equivalences; and when a structure $\mathfrak{A}$ is clear from context, we refer to equivalence classes of $r_1^{\mathfrak{A}} \cap r_2^{\mathfrak{A}}$ as *intersections*. The lower bounds are obtained by a reduction from a variant of the tiling problem. Let $\mathfrak{G}_m$ denote the standard grid on a finite $m \times m$ torus: $\mathfrak{G}_m = ([0, m-1]^2, h, v)$, $h = \{((p, q), (p', q)) : p' - p \equiv 1 \mod m\}$, $v = \{((p, q), (p, q')) : q' - q \equiv 1 \mod m\}$. A *tiling system* is a quadruple $\mathcal{T} = \langle C, c_0, H, V\rangle$, where C is a non-empty, finite set of *colours*, c_0 is an element of C, and H, V are binary relations on C called the *horizontal* and *vertical* constraints, respectively. A *tiling* for $\mathcal{T}$ of a grid $\mathfrak{G}_m$ is a function $f : [0, m]^2 \to C$ such that $f(0, 0) = c_0$ and, for all $d \in [0, m]^2$, the pair $\langle f(d), f(h(d))\rangle$ is in H and the pair $\langle f(d), f(v(d))\rangle$ is in V. The *doubly exponential tiling problem* is defined as follows.

Given: a number $n \in \mathbb{N}$ written in unary, and a tiling system $\mathcal{T}$.
Output: Yes, if $\mathcal{T}$ has a tiling of the grid $\mathfrak{G}_m$, where $m = 2^{2^n}$; No otherwise.

It is well known that the doubly exponential tiling problem is 2-NExpTime-complete (see, e.g. [28], p. 501).

Theorem 7.1. *The satisfiability and finite satisfiability problems for* EQ_2^2 *are 2-NExpTime-hard.*

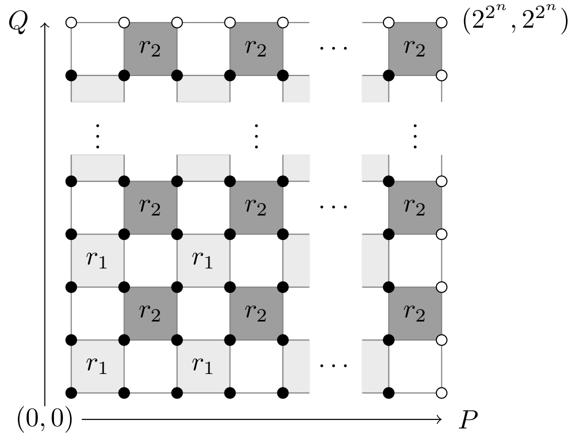

Figure 7.1: A doubly-exponential toroidal grid of intersections: the top and bottom rows are identified, as are the left- and right-most columns; r_1-classes are indicated by light grey squares, and r_2-classes by dark grey squares.

Proof. We proceed to reduce the doubly-exponential tiling problem to the satisfiability and finite satisfiability problems for EQ_2^2. The crux of the proof is a succinct axiomatization of a toroidal grid structure of doubly exponential size by means of an EQ_2^2-formula. In this axiomatization, the nodes of the grid are *intersections* (in our technical sense) containing at least 2^n elements. By regarding these elements as indices of binary digits, we can endow each intersection with a pair of (x, y)-coordinates in the range $[0, 2^{2^n} - 1]$. Our axiomatization forces each intersection to have a vertical and a horizontal successor with appropriate coordinates. This ensures that, for each pair of numbers (i, j) in the range $[0, 2^{2^n} - 1]$, there is at least one intersection having coordinates (i, j). In addition, our axioms ensure that horizontally successive intersections having respective coordinates (i, j) and $(i + 1, j)$ are related by r_1 if i is even, and by r_2 if i is odd; a similar condition holds for vertical successors. To guarantee that there is at most one intersection having coordinates (i, j), it is sufficient to assert: (i) there is at most one intersection having coordinates $(2^{2^n} - 1, 2^{2^n} - 1)$; and (ii) no two intersections possess a common horizontal or a common vertical successor. To enforce the latter condition, we use the pattern of r_1- and r_2-relations between successive intersections: we simply say that, if two elements are joined by one of the equivalence relations and if the parities of their (x, y)-coordinates agree, then they are also joined by the other equivalence relation, and hence are members of the same intersection. Thus, any model of our axioms has intersections arranged in the pattern shown in Fig. 7.1. Having established our grid, encoding an instance of the tiling problem can be done in a standard fashion. Below we describe the construction in detail.

Given an instance $(\mathcal{T}, n)$ of the doubly exponential tiling problem, where $\mathcal{T} = (C, c_0, H, V)$, we construct an EQ_2^2-formula Ω of length polynomial in n and $\mathcal{T}$, such that the following are equivalent: (i) Ω is satisfiable; (ii) Ω is finitely satisfiable; (iii) $(\mathcal{T}, n)$ is positive. As usual, we take r_1, r_2 to be distinguished binary predicates interpreted as equivalence relations. For ease of reading, we abbreviate $r_1(x, y) \wedge r_2(x, y)$ by $r_{12}(x, y)$, and we introduce the conjuncts of Ω in groups.

Let $o_1, \ldots, o_n$ be unary predicates. By taking the o_i to indicate the values of binary digits, we may take each element in any structure interpreting these predicates to have a 'local coordinate' in the form of a (single) number in the range $[0, 2^n - 1]$. For our purposes, it helps to think of an element's local coordinate as fixing its position within its intersection. We employ the abbreviation $\varepsilon(x, y)$ to state that x and y (which may be from different intersections) have the same local coordinates, $\lambda(x, y)$ to state that the local coordinate of y is one greater than the local coordinate of x (addition modulo 2^n), and $\zeta(x)$ to state that the local coordinate of x is 0. All these formulas can be defined in a straightforward way. The conjunct

$$\forall x \exists y \big(r_{12}(x, y) \wedge \lambda(x, y)\big) \tag{7.1}$$

then ensures that each intersection contains a collection of 2^n elements, distinguished by local coordinates in the range $[0, 2^n - 1]$.

We now endow each intersection with a pair of 'global coordinates' corresponding to the grid coordinates, in the range $[0, 2^{2^n} - 1]$, though the process here is more involved than with local coordinates. Let p and q be unary predicates. The conjunct

$$\forall x, y \big(r_{12}(x, y) \wedge \varepsilon(x, y) \rightarrow ((p(x) \leftrightarrow p(y)) \wedge (q(x) \leftrightarrow q(y)))\big) \tag{7.2}$$

ensures that elements of the same intersection with the same local coordinates agree on the satisfaction of p and q. To avoid cumbersome circumlocutions in the sequel, we allow ourselves to speak of *the* element of some intersection with a given local coordinate, since all such elements will turn out to have identical properties. If I is an intersection, we take the global P-coordinate of I to be the number in the range $[0, 2^{2^n} - 1]$ whose jth bit ($0 \leq j \leq 2^n - 1$) is 1 just in case the element of I whose local coordinate is j satisfies the predicate p. Likewise, we take the global Q-coordinate of I to be the number in the range $[0, 2^{2^n} - 1]$ whose jth bit ($0 \leq j \leq 2^n - 1$) is 1 just in case the element of I whose local coordinate is j satisfies the predicate q.

Recalling that $\zeta(y)$ states that the local coordinate of y is 0, we abbreviate the formula $\exists y (r_{12}(x, y) \wedge \zeta(y) \wedge \neg p(y))$ by $p^\circ(x)$. Thus, we may read $p^\circ(x)$ as "x belongs to an intersection whose global P-coordinate is an even number". Similarly, we may write a formula $q^\circ(x)$ to mean "x belongs to an intersection whose global Q-coordinate is an even number". Of course, all elements in an intersection agree on the satisfaction of these predicates; hence, we may speak of the satisfaction of $p^\circ(x)$ or $q^\circ(x)$ *by an intersection*.

We employ the abbreviations

$$\begin{aligned}
\eta(x, y) \equiv & \big(r_1(x, y) \wedge \neg r_2(x, y) \wedge \neg p^\circ(x) \wedge p^\circ(y)\big) \vee \\
& \big(\neg r_1(x, y) \wedge r_2(x, y) \wedge p^\circ(x) \wedge \neg p^\circ(y)\big) \\
\nu(x, y) \equiv & \big(r_1(x, y) \wedge \neg r_2(x, y) \wedge \neg q^\circ(x) \wedge q^\circ(y)\big) \vee \\
& \big(\neg r_1(x, y) \wedge r_2(x, y) \wedge q^\circ(x) \wedge \neg q^\circ(y)\big).
\end{aligned}$$

Evidently, if a pair of elements satisfies $\eta(x, y)$, then so does any other pair of elements from the same respective intersections. We wish to read $\eta(x, y)$ as "the intersection of y is a horizontal successor of the intersection of x", and $\nu(x, y)$ as "the intersection of y is a vertical successor of the intersection of x": we proceed to add conjuncts to Ω justifying these readings.

Suppose I and J are intersections. We shall write conjuncts ensuring that if J is a horizontal successor of I (in the sense of the previous paragraph), then I and J have successive P-coordinates and identical Q-coordinates. Let $\hat{p}$ be a unary predicate. Observing that the elements of an intersection are naturally ordered by their local coordinates, and recalling that $\lambda(x, y)$ states that the local coordinate of y is one greater than the local coordinate of x, the conjuncts

$$\forall x(\zeta(x) \to \hat{p}(x)) \tag{7.3}$$

$$\forall x \forall y(\lambda(x, y) \to (\hat{p}(y) \leftrightarrow p(x) \wedge \hat{p}(x))) \tag{7.4}$$

allow us to read $\hat{p}(x)$ as stating that all the bits in the global P-coordinate of the intersection containing x up to (but not necessarily including) the bit x are 1. Thus, the formula $\hat{p}(x) \wedge \neg p(x)$ says "x is the least significant zero-bit in the global P-coordinate of its intersection". Recalling that $\varepsilon(x, y)$ states that x and y have the same local (but not necessarily global) coordinates, we can enforce the required global coordinate constraints on horizontal successors using the conjuncts

$$\forall x \forall y(\eta(x, y) \wedge \varepsilon(x, y) \to (\hat{p}(x) \to (p(x) \leftrightarrow \neg p(y)))) \tag{7.5}$$

$$\forall x \forall y(\eta(x, y) \wedge \varepsilon(x, y) \to (\neg\hat{p}(x) \to (p(x) \leftrightarrow p(y)))) \tag{7.6}$$

$$\forall x \forall y(\eta(x, y) \wedge \varepsilon(x, y) \to (q(x) \leftrightarrow q(y))). \tag{7.7}$$

That is: two equivalence classes whose elements are related by η have global coordinates (P, Q) and $(P + 1, Q)$, for some P, Q in the range $[0, 2^{2^n} - 1]$ (addition modulo 2^{2^n}).

Let (7.8)–(7.12) be a counterparts of (7.3)–(7.7) for ν. Thus, by arranging the intersections in any model of Ω according to their global coordinates, we see that these intersections are related by r_1 and r_2 according to the pattern of Fig. 7.1, forming a doubly-exponential toroidal grid of interlocking r_1-classes and r_2-classes. Notice incidentally that intersections in even numbered columns satisfy p°, while those in odd-numbered columns do not. Likewise, the intersections in even numbered rows satisfy q°; those in odd-numbered rows do not.

Now we can enforce the existence of at least one intersection with any given pair of global coordinates in the range $[0, 2^{2^n} - 1]$, by writing conjuncts requiring each element to have at least one horizontal successor and at least one vertical successor:

$$\forall x \exists y.\eta(x, y) \wedge \forall x \exists y.\nu(x, y). \tag{7.13}$$

The main idea of the proof is that we can also enforce the existence of at *most* one intersection with any given pair of global coordinates in this range. Let $e(x)$ abbreviate $\forall y(r_{12}(x, y) \to (p(y) \wedge q(y)))$, stating that "$x$ belongs to an intersection whose global coordinates are $(2^{2^n} - 1, 2^{2^n} - 1)$". Hence, the conjunct

$$\forall x \forall y(e(x) \wedge e(y) \to r_{12}(x, y)). \tag{7.14}$$

ensures that there is exactly one such intersection.

We now write conjuncts preventing two intersections from having a common horizontal successor or a common vertical successor. To this end, observe from the definitions of $\eta(x, y)$ and $\nu(x, y)$ that, if x and y belong to intersections with a common horizontal or vertical successor, then they are related by either r_1 or r_2, and agree on $p^\circ(x)$ and $q^\circ(x)$. Thus, it suffices to add the conjunct

$$\forall x \forall y\big((r_1(x, y) \vee r_2(x, y)) \wedge (p^\circ(x) \leftrightarrow p^\circ(y)) \wedge (q^\circ(x) \leftrightarrow q^\circ(y)) \to r_{12}(x, y)\big). \tag{7.15}$$

(A glance at the arrangement of Fig. 7.1 shows that (7.15) is satisfied in this case.) Thus, in any model of Ω: (i) there is at most one intersection with global coordinates $(2^{2^n}-1, 2^{2^n}-1)$; (ii) every intersection possesses at least one horizontal successor and at least one vertical successor, with the global coordinates of these intersections related in the expected ways; (iii) no two intersections have a common horizontal successor or a common vertical successor. A straightforward double (backwards) induction, starting from the coordinates $(2^{2^n}-1, 2^{2^n}-1)$, then establishes that there is at most one intersection with any given pair of global coordinates, as required. That is: any model of Ω has precisely the pattern of intersections depicted in Fig. 7.1.

Having established a grid of doubly exponential size, the encoding of any instance of the doubly-exponential tiling problem on some tiling system (C, c_0, H, V) is routine. We simply add to Ω the conjuncts

$$\forall x \Big(\bigvee_{c \in C} c(x) \wedge \bigwedge_{\substack{c,d \in C \\ c \neq d}} \neg(c(x) \wedge d(x)) \Big) \tag{7.16}$$

$$\forall x \forall y \big(r_{12}(x,y) \wedge c(x) \rightarrow c(y) \big) \tag{7.17}$$

$$\forall x \forall y \big(\eta(x,y) \wedge c(x) \rightarrow \neg d(y) \big) (\langle c, d \rangle \notin H) \tag{7.18}$$

$$\forall x \forall y \big(\nu(x,y) \wedge c(x) \rightarrow \neg d(y) \big) (\langle c, d \rangle \notin V) \tag{7.19}$$

$$\exists x \left(\forall y (r_{12}(x,y) \rightarrow (\neg p(y) \wedge \neg q(y))) \wedge c_0(x) \right). \tag{7.20}$$

Notice that (7.20) states that the grid square with coordinates (0,0) is coloured with c_0.

Let Ω be the conjunction of constraints (7.1)–(7.20). From any model of Ω, we can read off a $\mathcal{T}$-tiling of size 2^{2^n}—for example, by looking at the colours assigned to the elements with local coordinate 0 in each of the $2^{2 \cdot 2^n}$ intersections. On the other hand, given any tiling for $\mathcal{T}$, we can construct a finite model of Ω in the obvious way using the arrangement of Fig. 7.1. Thus we see that: (i) if Ω is satisfiable, then $(\mathcal{T}, n)$ is positive; (ii) if $(\mathcal{T}, n)$ is positive, then Ω is finitely satisfiable. This proves the theorem. □

We remark that, in the above proof, (7.14) is the only conjunct of Ω that is not—modulo trivial logical manipulations—a guarded formula. The function of this formula is to ensure that there is only one intersection with global coordinates $(2^{2^n} - 1, 2^{2^n} - 1)$—an effect which could be achieved using a constant. Recalling that the satisfiability problem for the two-variable guarded fragment with two equivalence relations is 2-ExpTime-complete [16], we see that adding a single individual constant to this fragment results in the same complexity as the full (unguarded) fragment. That is:

Corollary 7.2. *The satisfiability problem for the guarded fragment of* FO^2 *with two equivalence relations and a single individual constant is 2*-NExpTime-*complete.*

8. Undecidability of FO^2 with one equivalence and one transitive relation. In this section we show that the (finite) satisfiability problem for two-variable first-order logic in which one distinguished predicate, r, is required to denote an equivalence and another, t, a transitive relation, is undecidable. This logic contains EQ^2_2: we may write FO^2 conjuncts requiring t to be reflexive and symmetric, and thus to be an equivalence. The result may also be a seen as a strengthening of an earlier theorem that FO^2 with two transitive relations is undecidable [13, 16]. Actually, our proof will show rather more: the logic in question is undecidable even under the stronger assumption that t is a strict partial order, rather than an arbitrary transitive relation.

The following proof closely follows the approach taken in [22], but additionally avoids the use of the equality predicate. We begin by recalling some definitions and lemmas from [26].

Let $\mathfrak{G}_m$ be the standard grid on a finite $m \times m$ torus as defined in Section 7, and let $\mathfrak{G}_{\mathbb{N}}$ be the standard grid structure on $\mathbb{N}^2$: $\mathfrak{G}_{\mathbb{N}} = (\mathbb{N}^2, h, v)$, $h = \{((p,q),(p+1,q)) : p,q \in \mathbb{N}\}$, $v = \{((p,q),(p,q+1)) : p,q \in \mathbb{N}\}$. An infinite structure $\mathfrak{G} = (G, h, v)$ is called *grid-like* if $\mathfrak{G}_{\mathbb{N}}$ is homomorphically embeddable into $\mathfrak{G}$; a finite $\mathfrak{G}$ is grid-like if some $\mathfrak{G}_m$ is homomorphically embeddable into $\mathfrak{G}$. Grid-likeness is implied by a simple local criterion. We say that h is *complete over* V in $\mathfrak{G} = (G, h, v)$ if $\mathfrak{G} \models \forall x \forall y \forall x' \forall y' ((h(x,y) \wedge v(x,x') \wedge v(y,y')) \to h(x',y'))$.

LEMMA 8.1. *Assume that $\mathfrak{G} = (G,h,v)$ satisfies the* FO^2*-axiom* $\forall x(\exists y\ h(x,y) \wedge \exists y\ v(x,y))$. *If h is complete over v, then $\mathfrak{G}$ is grid-like.*

LEMMA 8.2. *Let $\mathcal{C}$ be a class of structures, and suppose that there exists an* FO^2 *sentence Ω such that:*

(*a*) *$\mathfrak{G}_{\mathbb{N}}$ can be expanded to a structure in $\mathcal{C}$ satisfying Ω;*
(*b*) *for every $n \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that the grid $\mathfrak{G}_m$ with $m = kn$ can be expanded to a structure in $\mathcal{C}$ satisfying Ω;*
(*c*) *every model of Ω from $\mathcal{C}$ is grid-like.*

Then both satisfiability and finite satisfiability of FO^2 *over $\mathcal{C}$ are undecidable. In fact,* FO^2 *forms a conservative reduction class over $\mathcal{C}$.*

Now we are ready to prove the main result for this section.

THEOREM 8.3. *The satisfiability and finite satisfiability problems for* FO^2 *with one equivalence and one transitive relation* (*but without equality*) *are both undecidable.*

Proof. We construct a sentence Ω satisfying conditions (a)–(c) of Lemma 8.2. We add to the formula Ω suitable conjuncts to ensure that both the infinite grid, $\mathfrak{G}_{\mathbb{N}}$, and every finite toroidal grid, $\mathfrak{G}_{8n}$, can be expanded to a model of Ω.

The formula Ω employs unary predicates c_{ij} with $0 \le i \le 3$ and $0 \le j \le 7$, together with binary predicates h v, r and t. We refer to the c_{ij} as *colours*, and to h and v as the *horizontal* and *vertical grid relations*, respectively. We assume that r is interpreted as an equivalence, and t as a transitive relation. The colour $c_{i,j}$ describes elements whose column number, modulo 8, is i, and whose row number, modulo 4, is j, as shown in Figure 8.1. When we use addition in subscripts of the $c_{i,j}$ s, it is always understood modulo 4 in the first position, and modulo 8 in the second position, i.e. $c_{i+a,j+b}$ denotes $c_{(i+a) \bmod 4, (j+b) \bmod 8}$. We start by writing the initial formula

$$\exists x c_{00}(x) \wedge \forall x(\exists y\ h(x,y) \wedge \exists y\ v(x,y)). \tag{8.1}$$

Both grid relations, h and v, interact with t in two possible ways. To define these, we employ the abbreviations

$$\begin{aligned}
\theta_{i,j} &\equiv \forall x \forall y (c_{i,j}(x) \wedge h(x,y) \to c_{i+1,j}(y) \wedge t(x,y)) \\
\bar{\theta}_{i,j} &\equiv \forall x \forall y (c_{i,j}(x) \wedge h(x,y) \to c_{i+1,j}(y) \wedge t(y,x)) \\
\xi_{i,j} &\equiv \forall x \forall y (c_{i,j}(x) \wedge v(x,y) \to c_{i,j+1}(y) \wedge t(x,y)) \\
\bar{\xi}_{i,j} &\equiv \forall x \forall y (c_{i,j}(x) \wedge v(x,y) \to c_{i,j+1}(y) \wedge t(y,x)),
\end{aligned}$$

and add to Ω the conjuncts

$$\bigwedge_{i=0,2} \bigwedge_{j=1,2,5,6} \theta_{i,j} \wedge \bigwedge_{i=1,3} \bigwedge_{j=0,3,4,7} \theta_{i,j} \wedge \bigwedge_{i=0,2} \bigwedge_{j=0,3,4,7} \bar{\theta}_{i,j} \wedge \bigwedge_{i=1,3} \bigwedge_{j=1,2,5,6} \bar{\theta}_{i,j} \tag{8.2}$$

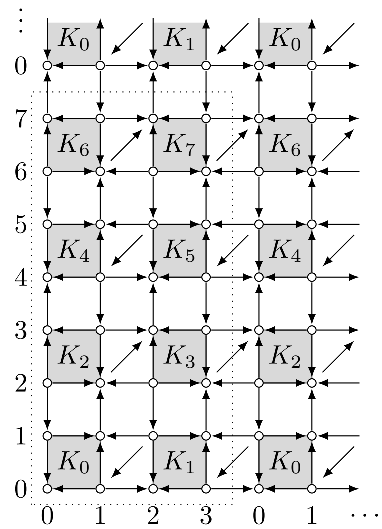

Figure 8.1: Expansion of $\mathfrak{G}_{\mathbb{N}}$ to a structure interpreting r, t and the colours $c_{i,j}$: the grid element $(n,m) \in \mathbb{N} \times \mathbb{N}$ is coloured with $c_{i,j}$, where $i = n \mod 4$ and $j = m \mod 8$; r-classes are indicated by grey shading; arrows depict t-connections.

$$\bigwedge_{i=0,2} \bigwedge_{j=2,3,6,7} \xi_{i,j} \wedge \bigwedge_{i=1,3} \bigwedge_{j=0,1,4,5} \xi_{i,j} \wedge \bigwedge_{i=0,2} \bigwedge_{j=0,1,4,5} \bar{\xi}_{i,j} \wedge \bigwedge_{i=1,3} \bigwedge_{j=2,3,6,7} \bar{\xi}_{i,j}. \tag{8.3}$$

The equivalence relation r partitions the required model into many equivalence classes. We define a partition with eight classes, denoted $K_0, \ldots, K_7$, with each K_i equal to the union of those r-classes whose elements realize a particular combination of colours $c_{i,j}$. (We note that not all combinations of these colours are possible in models of Ω.) The intended arrangement of these *colour classes* is depicted in Figure 8.1. To enforce this partition, we employ the following abbreviations

$$\text{for every } l = 0,2,4,6\text{: } K_l(x) \equiv c_{0,l}(x) \vee c_{0,l+1}(x) \vee c_{1,l}(x) \vee c_{1,l+1}(x)$$
$$\text{for every } l = 1,3,5,7\text{: } K_l(x) \equiv c_{2,l-1}(x) \vee c_{2,l}(x) \vee c_{3,l-1}(x) \vee c_{3,l}(x),$$

and we add to Ω the conjunct

$$\forall x \forall y \Big(r(x,y) \rightarrow \bigwedge_{k \neq l} \neg(K_k(x) \wedge K_l(y)) \wedge \bigwedge_{i,j} \big(c_{i,j}(x) \wedge c_{i,j}(y) \rightarrow t(x,y) \wedge t(y,x)\big)\Big) \tag{8.4}$$

which expresses that elements belonging to the same equivalence class and having the same colour form a t-clique. This means that the structure of our possible models is similar to Fig. 8.1, where white circles represent t-cliques. If we allowed equality, we could write formulas identifying elements of the same colour within a t-clique; but this is not needed for undecidability.

We also induce the diagonal t-edges drawn in Figure 8.1 by adding to Ω the conjuncts

$$\begin{aligned} &\bigwedge_{i=1,3}\bigwedge_{j=0,4}\Big(\forall x\big(c_{i,j}(x)\to\exists y(t(y,x)\wedge c_{i+1,j+1}(y))\big)\Big)\\ &\bigwedge_{i=1,3}\bigwedge_{j=2,6}\Big(\forall x\big(c_{i,j}(x)\to\exists y(t(x,y)\wedge c_{i+1,j+1}(y))\big)\Big)\end{aligned} \tag{8.5}$$

and we add to Ω a formula saying that certain elements connected by t are in the same r-class

$$\bigwedge_{l=0}^{7}\forall x\forall y\big(t(x,y)\wedge K_l(x)\wedge K_l(y)\to r(x,y)\big). \tag{8.6}$$

To ensure that every model of Ω is grid-like, we need additional conjuncts saying that certain elements connected by t are also connected by the horizontal grid relation

$$\begin{aligned} &\bigwedge_{i=0,2}\bigwedge_{j=0,3,4,7}\forall x\forall y\big(t(y,x)\wedge c_{i,j}(x)\wedge c_{i+1,j}(y)\to h(x,y)\big),\\ &\bigwedge_{i=1,3}\bigwedge_{j=1,2,5,6}\forall x\forall y\big(t(y,x)\wedge c_{i,j}(x)\wedge c_{i+1,j}(y)\to h(x,y)\big),\\ &\bigwedge_{i=0,2}\bigwedge_{j=1,2,5,6}\forall x\forall y\big(t(x,y)\wedge c_{i,j}(x)\wedge c_{i+1,j}(y)\to h(x,y)\big),\\ &\bigwedge_{i=1,3}\bigwedge_{j=0,3,4,7}\forall x\forall y\big(t(x,y)\wedge c_{i,j}(x)\wedge c_{i+1,j}(y)\to h(x,y)\big)\end{aligned} \tag{8.7}$$

and a similar formula for elements connected by r

$$\bigwedge_{i=0,2}\bigwedge_{j=1,3,5,7}\forall x\forall y\big(r(x,y)\wedge c_{i,j}(x)\wedge c_{i,j+1}(y)\to h(x,y)\big). \tag{8.8}$$

We show that the expansion of $\mathfrak{G}_{\mathbb{N}}$ illustrated in Figure 8.1 is a model of the formula Ω. It is clear that in the model all conjuncts of the form (8.1)–(8.6) hold. To see that also conjuncts of the form (8.7)–(8.8) are satisfied, observe that every t-path in the structure is finite and of length at most 6. Moreover, any t-path connects at most three adjacent columns and at most five adjacent rows. So, the distribution of the colours $c_{i,j}$ ensures that formulas (8.7)–(8.8) cannot force new pairs of elements, apart from those already connected in the standard grid, to become connected by h or v.

By considering two copies of the arrangement in the dotted rectangle of Fig. 8.1 placed side by side, an identical argument shows that every grid $\mathfrak{G}_{8m}$ can be expanded to a model of Ω.

To show that every model of Ω is grid-like, i.e. that condition (c) of Lemma 8.2 holds, we use Lemma 8.1 and prove the following claim.

Claim. In every model $\mathfrak{A}$ of Ω, h is complete over v, i.e.:

$$\mathfrak{A}\models\forall x\forall y\forall x'\forall y'\big(h(x,y)\wedge v(x,x')\wedge v(y,y')\to h(x',y')\big).$$

Assume that $\mathfrak{A}\models h[a,b]\wedge v[a,a']\wedge v[b,b']$. We show that $\mathfrak{A}\models h[a',b']$. Several cases need to be considered, depending on the colour of the element a. We discuss three typical ones.

Case 1: $\mathfrak{A} \models c_{00}[a]$. By $\bar{\theta}_{00}$ from (8.2) we have $\mathfrak{A} \models t[b,a] \wedge c_{10}[b]$. Formula $\bar{\xi}_{00}$ from (8.3) implies $\mathfrak{A} \models t[a',a] \wedge c_{01}[a']$. Formula ξ_{10} from (8.3) implies $\mathfrak{A} \models t[b,b'] \wedge c_{11}[b']$. By (8.6), $\mathfrak{A} \models r[a,b]$, $\mathfrak{A} \models r[a,a']$ and $\mathfrak{A} \models r[b,b']$. Since r is an equivalence, we have $\mathfrak{A} \models r[a',b']$. And by (8.8), we get $\mathfrak{A} \models h[a',b']$. A similar argument works for a coloured by c_{20}, c_{02}, c_{22}, c_{04}, c_{24}, c_{06} or c_{26}.

Case 2: $\mathfrak{A} \models c_{10}[a]$. As before, by θ_{10} from (8.2), we have $\mathfrak{A} \models t[a,b] \wedge c_{20}[b]$. Formula ξ_{10} from (8.3) implies $\mathfrak{A} \models t[a,a'] \wedge c_{11}[a']$, and $\bar{\xi}_{20}$ implies $\mathfrak{A} \models t[b',b] \wedge c_{21}[b']$. Now, by (8.5), for some $c \in A$, $\mathfrak{A} \models t[c,a] \wedge c_{21}[c]$. By transitivity of t, $\mathfrak{A} \models t[c,a']$ and $\mathfrak{A} \models t[c,b]$. As $\mathfrak{A} \models K_1[b] \wedge K_1[b'] \wedge K_1[c]$, by (8.6), we have $\mathfrak{A} \models r[c,b]$ and $\mathfrak{A} \models r[b',b]$. Since r is an equivalence, we have $\mathfrak{A} \models r[b',c]$ and so, using (8.4), $\mathfrak{A} \models t[b',c] \wedge t[c,b']$. So, by transitivity of t, $\mathfrak{A} \models t[b',a']$. Now, as $\mathfrak{A} \models c_{11}[a'] \wedge c_{21}[b']$, by (8.7), we get $\mathfrak{A} \models h[a',b']$. A similar argument works for a coloured by c_{30}, c_{12}, c_{32}, c_{14}, c_{34}, c_{16} or c_{36}.

Case 3: $\mathfrak{A} \models c_{11}[a]$. By $\bar{\theta}_{11}$ from (8.2) we have $\mathfrak{A} \models t[b,a] \wedge c_{21}[b]$. Formula ξ_{11} from (8.3) implies $\mathfrak{A} \models t[a,a'] \wedge c_{12}[a']$, and $\bar{\xi}_{21}$ implies $\mathfrak{A} \models t[b',b] \wedge c_{22}[b']$. Now, by transitivity of t, $\mathfrak{A} \models t[b',a']$. Using (8.7) we get $\mathfrak{A} \models h[a',b']$. The remaining cases are similar to Case 3. □

We conclude by noting that the grid relations h and v can be replaced with appropriate combinations of r, t and the unary predicates $c_{i,j}$. Furthermore, all the resulting formulas are—modulo trivial logical manipulations—guarded. Moreover, the transitive relation t is not required to contain non-trivial cliques, and thus we may assume that it is a partial order. Therefore:

Corollary 8.4. *The (finite) satisfiability problem for the guarded fragment of* FO^2 *with one equivalence and one transitive relation (or with one equivalence and one partial order) is undecidable even if no other binary relation symbols are allowed (including equality).*

As mentioned in Section 1, the satisfiability problem for FO^2 in the presence of one transitive relation is in 2-NExpTime [33]. The satisfiability of FO^2 in the presence of a single transitive closure operation, however, is not currently known to be decidable. The decidability of finite satisfiability for both of these logics is likewise open.

Acknowledgements. We would like the thank the anonymous referees for their insightful comments and numerous suggestions.

REFERENCES

[1] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274, 1998.

[2] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[3] I. Borosh, M. Flahive, and B. Treybig. Small solutions of linear Diophantine equations. *Discrete Mathematics*, 58(3):215–220, 1986.

[4] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 3rd edition, 1990.

[5] F. Eisenbrand and G. Shmonin. Carathéodory bounds for integer cones. *Operations Research Letters*, 34(5):564–568, 2006.

[6] H. Ganzinger, Ch. Meyer, and M. Veanes. The two-variable guarded fragment with transitive relations. In *LICS*, pages 24–34. IEEE Computer Society, 1999.

[7] E. Grädel. On the restraining power of guards. *J. Symbolic Logic*, 64:1719–1742, 1999.

[8] E. Grädel, P. Kolaitis, and M. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.

[9] E. Grädel and M. Otto. On Logics with Two Variables. *Theoretical Computer Science*, 224:73–113, 1999.
[10] E. Grädel, M. Otto, and E. Rosen. Two-variable logic with counting is decidable. In *Logic in Computer Science*, pages 306–317. IEEE, 1997.
[11] E. Grädel, M. Otto, and E. Rosen. Undecidability results on two-variable logics. *Archiv für Mathematische Logik und Grundlagenforschung*, 38(4-5):313–354, 1999.
[12] N. Immerman, A. Rabinovich, T. Reps, S. Sagiv, and G. Yorsh. The boundary between decidability and undecidability for transitive-closure logics. In *Computer Science Logic*, volume 3210 of *LNCS*, pages 160–174. Springer, 2004.
[13] Y. Kazakov. *Saturation-based decision procedures for extensions of the guarded fragment.* PhD thesis, Universität des Saarlandes, Saarbrücken, Germany, 2006.
[14] L. G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20:191–194, 1979.
[15] E. Kieroński. The two-variable guarded fragment with transitive guards is 2EXPTIME-Hard. In *FOSSACS*, volume 2620 of *LNCS*, pages 299–312. Springer, 2003.
[16] E. Kieroński. Results on the guarded fragment with equivalence or transitive relations. In *Computer Science Logic*, volume 3634 of *LNCS*, pages 309–324. Springer, 2005.
[17] E. Kieroński. Decidability issues for two-variable logics with several linear orders. In *Computer Science Logic*, volume 12 of *LIPIcs*, pages 337–351. Schloß Dagsuhl - Leibniz-Zentrum für Informatik, 2011.
[18] E. Kieronski and J. Michaliszyn. Two-variable universal logic with transitive closure. In *Computer Science Logic*, volume 16 of *LIPIcs*, pages 396–410. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.
[19] E. Kieroński, J. Michaliszyn, I. Pratt-Hartmann, and L. Tendera. Two-variable first-order logic with equivalence closure. In *Logic in Computer Science*, pages 431–440. IEEE, 2012.
[20] E. Kieroński and M. Otto. Small substructures and decidability issues for first-order logic with two variables. In *Logic in Computer Science*, pages 448–457. IEEE, 2005.
[21] E. Kieroński and M. Otto. Small substructures and decidability issues for first-order logic with two variables. *Journal of Symbolic Logic*, 77:729–765, 2012.
[22] E. Kieroński and L. Tendera. On finite satisfiability of two-variable first-order logic with equivalence relations. In *Logic in Computer Science*, pages 123–132. IEEE, 2009.
[23] L. Libkin. *Elements Of Finite Model Theory.* Springer, 2004.
[24] J. Michaliszyn. Decidability of the guarded fragment with the transitive closure. In *ICALP (2)*, volume 5556 of *LNCS*, pages 261–272. Springer, 2009.
[25] M. Mortimer. On languages with two variables. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 21:135–140, 1975.
[26] M. Otto. Two-variable first-order logic over ordered domains. *Journal of Symbolic Logic*, 66:685–702, 2001.
[27] L. Pacholski, W. Szwast, and L. Tendera. Complexity of two-variable logic with counting. In *Logic in Computer Science*, pages 318–327. IEEE, 1997.
[28] C.H. Papadimitriou. *Computational Complexity.* Addison Wesley Longman, Reading, MA, 1994.
[29] I. Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *Journal of Logic, Language and Information*, 14(3):369–395, 2005.
[30] T. Schwentick and T. Zeume. Two-variable logic with two order relations - (extended abstract). In *Computer Science Logic*, volume 6247 of *LNCS*, pages 499–513. Springer, 2010.
[31] D. Scott. A decision method for validity of sentences in two variables. *Journal Symbolic Logic*, 27:477, 1962.
[32] W. Szwast and L. Tendera. On the decision problem for the guarded fragment with transitivity. In *Logic in Computer Science*, pages 147–156. IEEE, 2001.
[33] W. Szwast and L. Tendera. FO^2 with one transitive relation is decidable. In *STACS*, volume 20 of *LIPIcs*, pages 317–328. Schloß Dagsuhl - Leibniz-Zentrum für Informatik, 2013.