The Two-Variable Fragment with Counting and Equivalence.

I. Pratt-Hartmann[1,*]

[1] School of Computer Science,
University of Manchester,
Manchester, UK

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

Key words Two-variable fragment, Satisfiability, Complexity
MSC (2010) 00-xx

We consider the two-variable fragment of first-order logic with counting, subject to the stipulation that a single distinguished binary predicate be interpreted as an equivalence. We show that the satisfiability and finite satisfiability problems for this logic are both NEXPTIME-complete. We further show that the corresponding problems for two-variable first-order logic with counting and *two* equivalences are both undecidable.

1 Introduction

The two-variable fragment of first-order logic, here denoted $\mathcal{L}^2$, is the set of function-free, first-order formulas (with equality) featuring at most two variables. The two-variable fragment with counting, here denoted $\mathcal{C}^2$, is the set of function-free, first-order formulas featuring at most two variables, but with the counting quantifiers $\exists_{[\leq M]}$, $\exists_{[\geq M]}$ and $\exists_{[=M]}$, (for every $M \geq 0$) allowed. It is impossible, in either logic, to express the fact that a given binary relation is an equivalence (i.e. is reflexive, symmetric and transitive). This suggests the possibility of enriching these logics by adding such a facility. We denote by $\mathcal{L}^2k$E the extension of $\mathcal{L}^2$ in which k distinguished binary predicates are required to be interpreted as equivalences, for any $k \geq 1$; and we denote by $\mathcal{C}^2k$E the analogous extension of $\mathcal{C}^2$.

The following facts are known. The logic $\mathcal{L}^2$ has the finite model property [1], and its satisfiability (= finite satisfiability) problem is NEXPTIME-complete [2]. The logic $\mathcal{C}^2$ is expressive enough for the finite model property to fail; nevertheless, its satisfiability and finite satisfiability problems remain NEXPTIME-complete [3–5]. The logic $\mathcal{L}^2$1E retains the finite model property, and its satisfiability problem remains NEXPTIME-complete [6]. The logic $\mathcal{L}^2$2E lacks the finite model property, and its satisfiability and finite satisfiability problems are both 2-NEXPTIME-complete [7]. The satisfiability and finite satisfiability problems for $\mathcal{L}^2k$E are both undecidable when $k \geq 3$ [8]. In this paper, we investigate $\mathcal{C}^2k$E—the two variable fragment with counting and k equivalences. We show that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$1E are both NEXPTIME-complete. We also show that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$2E are both undecidable. Note that the undecidability of the corresponding problems for $\mathcal{C}^2k$E where $k \geq 3$ follows anyway from the above-mentioned results on $\mathcal{L}^2k$E. Almost all of the paper is devoted to showing that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$1E are in NEXPTIME. The technique involves reduction to integer linear programming, along the lines of the treatment of $\mathcal{C}^2$ in [5]. A significant innovation of the present paper, however, is the use of Hilbert bases (of systems of linear Diophantine equations) to construct succinct certificates for finitely satisfiable $\mathcal{C}^2$1E-formulas.

A closely related family of logics is obtained by considering *transitive relations* in place of equivalences. We denote by $\mathcal{L}^2k$T the extension of $\mathcal{L}^2$ in which k distinguished binary predicates are required to be interpreted as transitive relations, for any $k \geq 1$; and we denote by $\mathcal{C}^2k$T the analogous extension of $\mathcal{C}^2$. It is easy to show that $\mathcal{L}^2$1T lacks the finite model property, and it is known (but by no means easy to show) that its satisfiability problem is in 2-NEXPTIME-time [9]. (No matching lower bound has yet been obtained, and the decidability of the finite satisfiability problem is still open.) The satisfiability and finite satisfiability problems for $\mathcal{L}^2k$T are

[*] Corresponding author E-mail: ipratt@cs.man.ac.uk, Phone: +00 44 275 6223

undecidable when $k \geq 2$ [6, 10]. In fact, the satisfiability and finite satisfiability problems for the weaker two-variable fragment with one equivalence and one transitive relation are also both undecidable [11]. In the context of logics with counting quantifiers, however, the satisfiability and finite satisfiability problems for $\mathcal{C}^2k$T are both undecidable for all $k \geq 1$ [12]. (Essentially the same result is obtainable from [13] by a simple adaptation.) It is possible to restore decidability—even in the presence of an arbitrary number of transitive relations—by restricting the underlying logical syntax. A great variety of such languages have been studied under the rubric of *description logics*, perhaps the best-known examples being $\mathcal{SHIQ}$ [14] or $\mathcal{SHOIQ}$ [15].

In the complexity-theoretic analysis of fragments of first-order logic, it is common to consider the restriction to guarded quantification [16], since doing so typically yields satisfiability problems (and finite satisfiability problems) with lower complexity. For example, the satisfiability problem for the guarded sub-fragment of $\mathcal{L}^2$ is ExpTime-complete [17]; likewise, although the guarded sub-fragment of $\mathcal{C}^2$ is still expressive enough for the finite model property to fail, its satisfiability and finite satisfiability problems are again both ExpTime-complete [18, 19]. On the other hand, it is easy to see that, in the presence of a single equivalence relation, restriction to guarded quantification produces no complexity-theoretic gains, since, within any equivalence class, all elements are accessible using guarded quantification (assuming that E may be used as a guard). In effect, the guarded fragment of $\mathcal{C}^2$1E enjoys the full power of $\mathcal{C}^2$.

Of some historical interest in this connection is the first-order theory of k equivalence relations. Here, we have full-first-order logic at our disposal (not just $\mathcal{C}^2$), but no non-logical predicates other than those denoting equivalences. It is reported in [20] that membership of a sentence in the first-order theory of one equivalence is decidable (even with equality); however, the first-order theory of two equivalences is undecidable (even without equality).

The structure of the paper is as follows. Section 2 establishes basic concepts and notation. Section 3 constructs the apparatus required to describe certain local configurations occuring in structures; this apparatus will allow us to apply techniques from linear integer programming to analyse of models of $\mathcal{C}^2$1E-formulas. Section 4 shows how, given a formula φ of $\mathcal{C}^2$1E, a data-structure can be constructed which, on the assumption that φ is satisfiable, is guaranteed to satisfy a collection of algorithmically checkable properties. We refer to such a data-structure as a *certificate*. If φ is in fact finitely satisfiable, this data-structure satisfies additional properties, and is referred to as a *finite certificate*. Section 5 establishes the converse: if a (finite) certificate for φ satisfies the requisite properties, then φ is (finitely) satisfiable. Section 6 establishes that if φ has a (finite) certificate, then it has one of exponentially bounded size, and that the properties it is required to satisfy can be checked in exponential time, as a function of the size of φ. This establishes that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$1E are both NExpTime-complete. Section 7 shows that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$2E are undecidable, by computable reduction to the halting problem for deterministic 2-counter machines.

A table listing the most common mathematical symbols that occur throughout the paper is given at the end.

2 Preliminaries

2.1 Logic

We employ standard terminology and notation from model-theory. Structures are denoted by (possibly decorated) fraktur letters $\mathfrak{A}$, $\mathfrak{B}$, and their domains by the corresponding Roman letters. Unless explicitly indicated to the contrary (as we do occasionally in the sequel), structures are assumed to have non-empty domains. If r is a symbol of the signature of $\mathfrak{A}$, we denote its interpretation by $r^{\mathfrak{A}}$. If φ is a formula, with free variables included in the list $x_1, \ldots, x_n$, $\mathfrak{A}$ is a structure and $\bar{a} = a_1, \ldots, a_n$ a tuple of elements of A, we write $\mathfrak{A} \models \varphi[a_1, \ldots, a_n]$ to indicate that $\bar{a}$ *satisfies* $\varphi(x_1, \ldots, x_n)$ in $\mathfrak{A}$; where φ is a sentence (i.e. has no free variables), we simply write $\mathfrak{A} \models \varphi$ to indicate that φ is true in $\mathfrak{A}$. We say φ is *satisfiable* if, for some $\mathfrak{A}$ and some $a_1, \ldots, a_n \in A$, $\mathfrak{A} \models \varphi[a_1, \ldots, a_n]$; if, in addition, A is finite, we say φ is *finitely satisfiable*. If $\neg\varphi$ is not satisfiable, we say φ is *valid*, and write $\models \varphi$. If $\mathcal{L}$ is a set of formulas, the (*finite*) *satisfiability problem* for $\mathcal{L}$ is the problem of determining whether a given $\varphi \in \mathcal{L}$ is (finitely) satisfiable. A set of formulas has the *finite model property* if its satisfiability and finite satisfiability problems coincide.

The two-variable fragment with counting, $\mathcal{C}^2$, is the set of function-free, first-order formulas (with equality), featuring only the variables x and y, but with the counting quantifiers $\exists_{[\leq M]}$, $\exists_{[\geq M]}$ and $\exists_{[=M]}$, for every $M \geq 0$, allowed. Formally, the subscripts M are bit-strings; however, we equivocate in the natural way between these bit-strings and the non-negative integers they encode. We read $\exists_{[\leq M]}x.\varphi$ as “There exist at most M x such that φ”, and similarly for the other counting quantifiers. The formal semantics are as expected. Evidently, $\exists_{[=0]}x.\varphi$ and $\exists_{[\leq 0]}x.\varphi$ are logically equivalent to $\forall x.\neg\varphi$, while $\exists_{[\geq 0]}x.\varphi$ is trivially true. We allow equality in $\mathcal{C}^2$-formulas; this represents no increase in expressive power, since the identity relation is anyway definable by the formula $\forall x.r(x,x) \wedge \forall x\exists_{[=1]}y.r(x,y)$. We do not allow individual constants in $\mathcal{C}^2$-formulas; this represents no effective decrease in expressive power, since we can always declare a unary predicate p to be uniquely instantiated by writing $\exists_{[=1]}x.p(x)$. Likewise, the use of predicates of arity greater than two adds no effective increase in expressive power, and we therefore assume, for simplicity, that all predicates are unary or binary. It is easy to see that $\mathcal{C}^2$ lacks the finite model property: the formula $\exists x\forall y\neg r(y,x) \wedge \forall x\exists y.r(x,y) \wedge \forall x\exists_{\leq 1}y.r(y,x)$ is satisfiable, but only over infinite domains.

The two-variable fragment with counting and one equivalence, $\mathcal{C}^2$1E, employs the same syntax and semantics as $\mathcal{C}^2$, but with the restriction that, in any structure $\mathfrak{A}$, the distinguished binary predicate E be interpreted as an equivalence. Where $\mathfrak{A}$ is clear from context, we refer to the cliques of $E^{\mathfrak{A}}$ simply as *equivalence classes*, and we say elements $a, b \in A$ are *equivalent* if $\mathfrak{A} \models E[a,b]$.

A formula ψ of $\mathcal{C}^2$1E is in *normal form* if it conforms to the pattern:

$$\forall x\forall y(x = y \vee \alpha) \wedge \bigwedge_{h=1}^{m} \forall x\exists_{[=M_h]}y(\beta_h \wedge x \neq y), \tag{1}$$

where α and the β_h are quantifier-free, equality-free $\mathcal{L}^2$-formulas, m is a positive integer, and the M_h are (bit-strings representing) positive integers. We refer to the integer $M = \max\{M_h \mid 1 \leq h \leq m\}$ as the *ceiling* of ψ. Observe that ψ is not satisfiable over any domain of cardinality less than or equal to M.

The following lemma uses a familiar technique originally employed in [21] to reduce the depth of quantification in $\mathcal{L}^2$-formulas.

Lemma 2.1 *Given a $\mathcal{C}^2$1E-formula φ, we can compute, in polynomial time, a normal-form $\mathcal{C}^2$1E-formula ψ, with ceiling M, such that, for any set A of cardinality greater than M, ψ is satisfiable over the domain A if and only if φ is.*

Proof. We may assume without loss of generality that φ is a universally quantified sentence. (Apply up to two existential quantifiers and a redundant universal quantifier if necessary.) We build ψ in three stages.

Stage 1: Write $\varphi^{(0)} = \varphi$, and suppose first that $\varphi^{(0)}$ contains a subformula $\pi(u) = \exists_{[\leq N]}v.\gamma$, where u, v are the variables x, y in some order, and γ is quantifier-free. Let p be a new unary predicate, let $\varphi^{(1)}$ be the result of replacing $\pi(u)$ in φ by the atomic formula $p(u)$, and define

$$\psi' := \varphi^{(1)} \wedge \forall u\exists_{[\leq N]}v(p(u) \wedge \gamma) \wedge \forall u\exists_{[\geq(N+1)]}v(p(u) \vee \gamma).$$

It is routine to check that $\forall u\exists_{[\leq N]}v(p(u) \wedge \gamma) \wedge \forall u\exists_{[\geq(N+1)]}v(p(u) \vee \gamma)$ entails $\forall u(p(u) \leftrightarrow \exists_{[\leq N]}v.\gamma)$; hence, ψ' entails φ. Moreover, if $\mathfrak{A} \models \varphi$ with $|A| > N$, then we may expand $\mathfrak{A}$ to a model $\mathfrak{A}'$ of ψ' by setting $p^{\mathfrak{A}'} = \{a \in A \mid \mathfrak{A} \models \pi[a]\}$. On the other hand, if $\varphi^{(0)}$ does not contain a subformula $\pi(u) = \exists_{[\leq N]}v.\gamma$, then it contains a subformula having one of the forms $\exists_{[\geq N]}v.\gamma$, $\exists_{[=N]}v.\gamma$, $\exists v.\gamma$ or $\forall v.\gamma$, and we may proceed similarly, subject to the obvious adjustments. Now apply the same process to the subformula $\varphi^{(1)}$ and continue until we obtain a universally quantified formula, say $\varphi^{(k)}$, in prenex form with quantifier depth at most two. Collecting $\varphi^{(k)}$ together with all the added conjuncts, and applying trivial logical manipulations, we obtain a formula

$$\psi^* := \forall x\forall y.\alpha' \wedge \bigwedge_{h=1}^{m'} \forall x\exists_{[\bowtie_h M'_h]}y.\beta'_h,$$

where: (*i*) α' and the β'_h are quantifier-free, (*ii*) the symbol $\bowtie_h$ stands for any of $\leq$, $\geq$ or $=$; (*iii*) $\psi^* \models \varphi$; and (*iv*) if $\mathfrak{A} \models \varphi$ with $|A| > \max_h M'_h$, then $\mathfrak{A}$ may be expanded to a model of ψ^*.

Stage 2: Replace any conjunct of the form $\forall x \exists_{\leq M'_h} y . \beta_h$ by $\forall x \exists_{=M'_h} y . q(x, y)$, where q is a new binary predicate, and add the conjunct $\forall x \forall y (\beta \rightarrow q(x, y))$; similarly, *mutatis mutandis*, for the case where $\bowtie_h$ is $\geq$. By rearranging conjuncts again, we may henceforth assume that each of the symbols $\bowtie_h$ in ψ^* is in fact $=$.

Stage 3: Over domains of size at least 2, the formulas $\forall x \forall y . \alpha'(x, y)$ and $\forall x \forall y (x = y \vee (\alpha'(x, y) \wedge \alpha'(x, x)))$ are logically equivalent. Let α be the result of eliminating equalities from $(\alpha'(x, y) \wedge \alpha'(x, x))$ in the obvious way: i.e. replace any subformula $u = u$ by $\top$ and any subformula $u = v$, with u, v different, by $\perp$. Thus, over domains of size at least 2, $\forall x \forall y . \alpha'(x, y)$ is logically equivalent to $\forall x \forall y (x = y \vee \alpha)$. Similarly, replace each of the conjuncts $\forall x \exists_{=M'_h} y . \beta'_h$ (assuming $M'_h \geq 1$) with the corresponding conjunction

$$\forall x \exists_{[=(M'-1)]} y (q(x, y) \wedge x \neq y) \wedge \forall x \exists_{[=M']} y (q'(x, y) \wedge x \neq y) \wedge$$
$$\forall x \forall y \left(x = y \vee [(\beta'_h(x, x) \rightarrow (\beta'_h(x, y) \leftrightarrow q(x, y))) \wedge (\neg \beta'_h(x, x) \rightarrow (\beta'_h(x, y) \leftrightarrow q'(x, y)))]\right),$$

where q and q' are fresh binary predicates. Modulo trivial logical manipulations, the resulting formula ψ is of the form (1), and is satisfiable over the over a set A with $|A| > M$ if and only if φ is. □

2.2 Linear algebra

We write $\mathbb{N}$ for the set of *natural numbers*, $\{0, 1, 2, \dots\}$. If A is a matrix, we write $A[i, j]$ to denote the (i, j)th entry of A, and similarly for vectors. A matrix or vector (or indeed any collection of numbers) is *bounded by* a quantity M if each of its entries is, and *absolutely bounded by* M if the absolute value of each of its entries is. Matrices, vectors and scalars that it is helpful to think of as constants will frequently be indicated in bold type. If m, n are natural numbers with $m \leq n$, we write $[m, n]$ for the set $\{m, m + 1, \dots, n\}$.

A *linear Diophantine equation* is a linear equation $\underline{\mathbf{a}} \cdot \underline{w} + \mathbf{b} = \underline{\mathbf{f}} \cdot \underline{w} + \mathbf{g}$, where all coefficients are (possibly negative) integers. In this paper, we drop the modifier "linear" in the sequel, and simply speak of *Diophantine equations*. As long as variables take only finite values, any system $\mathcal{E}$ of Diophantine equations may without loss of generality be written in matrix form: $\mathbf{A}\underline{w} = \underline{\mathbf{b}}$. We occasionally write $\mathcal{E}(\underline{w})$ to make the tuple of variables in $\mathcal{E}$ explicit. We refer to the elements of $\mathbf{A}$ as the *variable coefficients* of $\mathcal{E}$ and the elements of $\mathbf{b}$ as the *constant coefficients* of $\mathcal{E}$. If $\underline{\mathbf{b}} = \underline{0}$, we say that $\mathcal{E}$ is *homogeneous*. If $\mathcal{E}$ is any system of Diophantine equations, we write $\|\mathcal{E}\|$ to denote the *size of* $\mathcal{E}$, i.e. the total number of bits required to write all its coefficients; and we write $|\mathcal{E}|$ to denote the *cardinality* of $\mathcal{E}$, i.e. the number or rows in $\mathbf{A}$.

We consider also *Diophantine inequalities* $\underline{\mathbf{a}} \cdot \underline{w} \leq \mathbf{b}$, where, coefficients are again (possibly negative) integers. Any such inequality can be converted into an equivalent (in the obvious sense) Diophantine equation $\underline{\mathbf{a}} \cdot \underline{w} + y = \mathbf{b}$, where y is a fresh variable, usually referred to as a *slack variable*. Thus, a mixed system of r Diophantine equations and inequalities can be converted into an equivalent system of r Diophantine equations by the addition of at most r slack variables. At various points in the sequel, we shall need to consider disjunctions of Diophantine equations and inequalities. We call such a disjunction a *Diophantine clause*.

We employ two basic results on systems of Diophantine equations, given in Propositions (2.2) and (2.4). Given two vectors over $\mathbb{N}$, say $\underline{a}$ and $\underline{b}$, we write $\underline{a} \trianglelefteq \underline{b}$ if, for every i $(1 \leq i \leq n)$, $\underline{a}[i] \leq \underline{b}[i]$. Thus, $\trianglelefteq$ is a partial order; we refer to it as the *pointwise order*. And if $\mathcal{E}$ is any homogeneous system of Diophantine equations, we define its *Hilbert basis*, denoted $H(\mathcal{E})$, to be the set of solutions of $\mathcal{E}$ over $\mathbb{N}$ that are minimal in the pointwise order. It is obvious that every solution of $\mathcal{E}$ over $\mathbb{N}$ is a non-negative integer linear combination of vectors in $H(\mathcal{E})$. The following bound on $H(\mathcal{E})$ can be established by a remarkably simple combinatorial argument.

Proposition 2.2 (Pottier [22], Theorem 1) *Let $\mathcal{E} : \mathbf{A}\underline{x} = \underline{0}$ be a homogeneous system of linear Diophantine equations, with $\mathbf{A}$ absolutely bounded by M, and having dimensions $r \times k$. Then every vector in $H(\mathcal{E})$ is bounded by $(kM + 1)^r$.*

The restriction to the homogeneous case is easily lifted.

Corollary 2.3 *Let $\mathcal{E} : \mathbf{A}\underline{x} = \underline{\mathbf{b}}$ be a system of Diophantine equations, with $\mathbf{A}$ and $\mathbf{b}$ absolutely bounded by M, and $\mathbf{A}$ of dimensions $r \times k$. Then the set of solutions of $\mathcal{E}$ over $\mathbb{N}$ can be written as*

$$\{\underline{w}_0 + \zeta_1 \underline{\mathbf{w}}_1 + \cdots + \zeta_L \underline{\mathbf{w}}_L \mid \underline{w}_0 \in \mathbf{W},\ \zeta_1, \dots, \zeta_L \in \mathbb{N}\},$$

where $\mathbf{W}$ *is a finite set of vectors over* $\mathbb{N}$*, and* $\underline{\mathbf{w}}_1, \ldots, \underline{\mathbf{w}}_L$ *a list of vectors over* $\mathbb{N}$*. Each vector in* $\mathbf{W}$ *and each of the vectors* $\underline{\mathbf{w}}_1, \ldots, \underline{\mathbf{w}}_L$ *is bounded by* $((k+1)M+1)^r$*; hence* $|\mathbf{W}|$ *and* L *are bounded by* $(1+((k+1)M+1)^r)^k$.

Proof. Let $\mathcal{E}'$ be the homogeneous system of Diophantine equations $(\mathbf{A} \mid -\mathbf{b}) \begin{pmatrix} x \\ y \end{pmatrix} = \underline{0}$, where y is a new variable. The solutions of $\mathcal{E}$ are evidently those of $\mathcal{E}'$ in which $y = 1$ (with the final 1 projected out). Now divide $H(\mathcal{E}')$ into sets of vectors $\mathbf{W}'$, $\mathbf{W}''$ and $\mathbf{W}'''$, according as the last element is 0, 1 or greater than 1, and enumerate $\mathbf{W}'$ as $\underline{\mathbf{w}}'_1, \ldots, \underline{\mathbf{w}}'_L$. Thus, any solution of $\mathcal{E}'$ in which $y = 1$ has the form $\underline{\mathbf{w}}''_0 + \zeta_1 \underline{\mathbf{w}}'_1 + \cdots + \zeta_L \underline{\mathbf{w}}'_L$, where $\underline{w}''_0 \in \mathbf{W}''$ and $\zeta_1, \ldots, \zeta_L$ are non-negative integers. Now let $\mathbf{W}$ and $\underline{\mathbf{w}}_1, \ldots, \underline{\mathbf{w}}_L$ be the result of projecting out the last components of all the vectors in $\mathbf{W}''$ and $\underline{\mathbf{w}}'_1, \ldots, \underline{\mathbf{w}}'_L$, respectively. The final statement then follows from Proposition 2.2. □

Note that the bounds in Corollary 2.3 imply the familiar fact that, if a system of Diophantine equations $\mathcal{E}$ has a solution over $\mathbb{N}$, then it has such a solution in which all values are bounded by $2^{p(\|\mathcal{E}\|)}$, where p is a fixed polynomial [23, Theorem 2], (see also [24]). This in turn implies that the problem of determining whether $\mathcal{E}$ has a solution is in NPTIME; indeed, the problem is easily seen to be NPTIME-complete.

In the sequel, we shall require a slightly sharper complexity bound when $\mathcal{E}$ has many more variables than equations. If $\underline{a}$ is a solution of $\mathcal{E}(\underline{w})$ over $\mathbb{N}$, and ω is a subset of the variables in $\underline{w}$, we say that $\underline{a}$ has *footprint* ω if all variables outside ω are assigned the value 0 in $\underline{a}$. The next result yields a bound on the cardinality of this set.

Proposition 2.4 ([25], Theorem 1(*ii*)) *Let* $\mathcal{E} : \mathbf{A}\underline{x} = \underline{\mathbf{b}}$ *be a system of Diophantine equations with* $\mathbf{A}$ *absolutely bounded by* M *and of dimensions* $r \times k$*. If* $\mathcal{E}$ *has a solution over* $\mathbb{N}$*, then it has a solution over* $\mathbb{N}$ *with footprint of size at most* $2r\log(4rM)$.

Strikingly, the bound obtained in Proposition 2.4 is independent both of the number of variables, k, and also of the constant coefficients, $\underline{\mathbf{b}}$. This feature will figure crucially in the argument of Sec. 4.4 and Sec. 6.

2.3 Extended linear algebra

We write $\mathbb{N}^*$ for the set of *extended natural numbers*, $\{0, 1, 2, \ldots, \aleph_0\}$. We interpret the arithmetic operations $+$ and $\cdot$ as well as the ordering $<$ over $\mathbb{N}^*$ as expected. Specifically: $\aleph_0 + n = n + \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0$ for all $n \in \mathbb{N}$; $\aleph_0 \cdot 0 = 0 \cdot \aleph_0 = 0$, and $\aleph_0 \cdot n = n \cdot \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$ for all $n \in \mathbb{N}$ with $n > 0$. The operations $+$ and $\cdot$ are associative and commutative, and satisfy the familiar distribution rule. We count $\aleph_0$ as *positive*. These definitions allow us to use extended natural numbers to reason about (possibly infinite) countable sets in the natural way. For brevity, we occasionally refer to tuples of values over $\mathbb{N}^*$ as $\mathbb{N}^*$-*vectors* (and similarly for $\mathbb{N}$-*vectors*). This section shows how results on linear Diophantine equations in Sec. 2.2 can be adapted to the extended natural numbers. Readers interested only in the *finite* satisfiability problem for $\mathcal{C}^2$1E may skip this material, ignoring all references to infinite values in the sequel.

When dealing with extended natural numbers, it is more convenient to present equations in a slightly different form. A *positive integer equation* is an equation of the form $\underline{\mathbf{a}} \cdot \underline{w} + \mathbf{b} = \underline{\mathbf{f}} \cdot \underline{w} + \mathbf{g}$, where the variable coefficients $\underline{\mathbf{a}}$, $\underline{\mathbf{f}}$ are natural numbers, and the constant coefficients $\mathbf{b}$, $\mathbf{g}$ are extended natural numbers. (We allow 0 as a coefficient.) Note that we again suppress the word "linear" here, as it is simply assumed. A system $\mathcal{E}$ of positive integer equations may thus be written in matrix form: $\mathbf{A}\underline{w} + \underline{\mathbf{b}} = \mathbf{F}\underline{w} + \underline{\mathbf{g}}$, where $\mathbf{A}$, $\mathbf{F}$ are matrices over $\mathbb{N}$ and $\underline{\mathbf{b}}$, $\underline{\mathbf{g}}$ are $\mathbb{N}^*$-vectors. Typically, we are interested in solutions over $\mathbb{N}^*$. Notice that, in this case, systems of positive integer equations cannot straightforwardly be re-written in the form $\mathbf{A}\underline{w} = \underline{\mathbf{b}}$. For instance, the single equation $w_1 + w_2 = w_1$ has solutions $(m, 0)$ and $(\aleph_0, n)$, for $m, n \in \mathbb{N}^*$; but the system $w_2 = 0$ has only the solutions $(m, 0)$. We use the terms *positive integer inequality* and *positive integer clause* to generalize Diophantine inequalities and Diophantine clauses in the obvious way.

We apply the notions of boundedness and footprints to the extended natural nubers as follows. A matrix or vector of extended natural numbers is said to be *finitely bounded by* M if each of its *finite* entries is bounded by M. (Thus, any infinite entries do not affect the bound.) Given a system $\mathcal{E}(\underline{w})$ of positive integer equations, having some solution $\underline{a}$ over $\mathbb{N}^*$, we say that $\underline{a}$ has *footprint* ω, where ω is a subset of the variables in $\underline{w}$, if all variables outside ω are assigned the value 0 in $\underline{a}$. (Thus, infinite entries in the solution do contribute to the footprint.)

Suppose $\mathcal{E}(\underline{w})$ is a system of positive integer equations having solutions $\underline{a}$, $\underline{b}$ over $\mathbb{N}^*$. By re-ordering variables if necessary, write $\underline{w} = \underline{w}', \underline{w}''$ where $\underline{w}'$ is the tuple of variables that take finite values in both $\underline{a}$ and $\underline{b}$, and $\underline{w}''$ is the tuple of variables that take infinite values in either $\underline{a}$ or $\underline{b}$. Now consider the system of equations $\mathcal{E}'(\underline{w}') \subseteq \mathcal{E}(\underline{w})$ featuring only the variables $\underline{w}'$ (i.e. all coefficients of variables in $\underline{w}''$ are 0) and no infinite constant coefficients. Thus, $\mathcal{E}'(\underline{w}')$ is a system of Diophantine equations with a solution, say $\underline{a}'$, over $\mathbb{N}$; and it is then obvious that setting all variables in $\underline{w}''$ to $\aleph_0$ yields a solution $\underline{a}', \underline{\aleph}_0$ of $\mathcal{E}$. (The argument relies at this point on the restriction of variable coefficients to $\mathbb{N}$.) Applying this argument to exhaustion, there is a unique smallest collection of variables $\underline{w}_0$ from $\underline{w}$, such that, re-ordering variables if necessary and writing $\underline{w} = \underline{w}_0, \underline{w}_1$, we see that $\mathcal{E}(\underline{w})$ has a solution $\underline{a}_0, \underline{\aleph}_0$, where $\underline{a}_0$ is an $\mathbb{N}$-vector and $\underline{\aleph}_0$ a tuple of $\aleph_0$s. We call such a solution *minimally finite*. Of course, the same argument applies when confining attention to solutions with a particular footprint. For any subset ω of the variables $\underline{w}$, if there is a solution of $\mathcal{E}$ over $\mathbb{N}^*$ with footprint ω, then we can speak of the *minimally finite* solutions of $\mathcal{E}$ *with footprint* ω. Notice that what is unique about minimally finite solutions (with a particular footprint) is the *set of variables taking finite values*: there may of course be more than one minimally finite solution.

In Secs. 4.5 and 5.2 we shall have occasion to deal with equations, inequalities and clauses having infinite variable coefficients. An *extended integer equation* is an equation of the form $\underline{\mathbf{a}} \cdot \underline{w} + \mathbf{b} = \underline{\mathbf{f}} \cdot \underline{w} + \mathbf{g}$, where all coefficients are extended natural numbers. A system of such equations will again typically be written in matrix form: $\mathbf{A}\underline{w} + \underline{\mathbf{b}} = \mathbf{F}\underline{w} + \underline{\mathbf{g}}$ where $\mathbf{A}$, $\mathbf{F}$ are matrices over $\mathbb{N}^*$, and $\underline{\mathbf{b}}$, $\underline{\mathbf{g}}$ are $\mathbb{N}^*$-vectors. We use the terms *extended integer inequality* and *extended integer clause* in the expected way. Corollary 2.3 has the following consequence.

Corollary 2.5 *Let $\mathcal{E}$ be a system of r extended integer equations in k variables, with all coefficients finitely bounded by M. If $\mathcal{E}$ has a solution over $\mathbb{N}^*$, then $\mathcal{E}$ has a solution over $\mathbb{N}^*$ finitely bounded by $((2k+1)M+1)^{(r+k)}$.*

Proof. Suppose $\mathcal{E}(\underline{w}, \underline{w}_0, \underline{w}^*)$ has a solution $\underline{a}, \underline{0}, \underline{\aleph}_0$, with all values in $\underline{a}$ finite and positive. Let $\mathcal{E}' \subseteq \mathcal{E}$ be the set of equations featuring no infinite terms in this solution (i.e. both sides of the equation are finite), and let $\mathcal{E}^+(\underline{w})$ be the result of removing all terms featuring the variables $\underline{w}_0$ from $\mathcal{E}'$. Let $\mathcal{E}''(\underline{w}, \underline{z})$ be the set of equations $w = z + 1$, where w is a variable in $\underline{w}$, and z is a fresh variable. Thus, $\mathcal{E}''$ asserts that the vector $\underline{w}$ has only positive entries, and $(\mathcal{E}^+ \cup \mathcal{E}'')(\underline{w}, \underline{z})$ is a system of at most $r + k$ Diophantine equations featuring at most $2k$ variables with solution, say, $\underline{a}, \underline{b}$ over $\mathbb{N}$. Corollary 2.3 then ensures that $(\mathcal{E}^+ \cup \mathcal{E}'')(\underline{w}, \underline{z})$ has a solution $\underline{a}', \underline{b}'$ bounded by $((2k+1)M+1)^{(r+k)}$. But in that case, $\underline{a}', \underline{0}, \underline{\aleph}_0$ is a solution of $\mathcal{E}$, since both sides of every equation in $\mathcal{E} \setminus \mathcal{E}'$ will evaluate to $\aleph_0$. □

Extending the notion of *footprint* to solutions of systems of extended integer equations in the obvious way, Proposition 2.4 has the following generalization.

Corollary 2.6 *Let $\mathcal{E}$ be a system of r extended integer equations, with all variable coefficients finitely bounded by M. If $\mathcal{E}$ has a solution over $\mathbb{N}^*$, then it has a solution over $\mathbb{N}^*$ with footprint of size at most $2r\log(4rM)$.*

Proof. Take any solution $\underline{a}$ of $\mathcal{E}$ over $\mathbb{N}^*$. For each equation in $\mathcal{E}$, if both sides of that equation evaluate to $\aleph_0$, mark that equation, pick one infinite term on each side, and, if that term involves a variable, mark that variable. Let $\mathcal{E}'$ be the set of marked equations and $\mathcal{E}''$ the set of unmarked equations. Let $\underline{w}'$ be the tuple of marked variables and $\underline{w}''$ the tuple of unmarked variables (in some order). Now write $\underline{w}'' = \underline{w}_0, \underline{w}^+, \underline{w}^*$, where all variables in $\underline{w}_0$ take value 0 in $\underline{a}$, all variables in $\underline{w}^+$ take a positive, finite value in $\underline{a}$, and all variables in $\underline{w}^*$ take value $\aleph_0$ in $\underline{a}$. Write $\underline{a} = \underline{a}', \underline{0}, \underline{a}^+, \underline{\aleph}_0$ corresponding to the decomposition $\underline{w} = \underline{w}', \underline{w}_0, \underline{w}^+, \underline{w}^*$. Let $r' = |\mathcal{E}'|$. Thus, $|\mathcal{E}''| = (r - r')$, and $\underline{w}'$ is of length at most $2r'$. If $r' = r$, then $\underline{a}', \underline{0}, \underline{0}, \underline{0}$ is evidently a solution of $\mathcal{E}$ with footprint of size at most $2r$, and we are done. So we may assume $r' < r$.

Notice that it is perfectly possible for $\mathcal{E}''$ to contain marked variables. (This might occur if, for example, some equation in $\mathcal{E}'$ involves a term $\aleph_0 w_1$, where w_1 takes a finite, non-zero value in $\underline{a}$.) Let us freeze these values. Formally, let $\mathcal{E}'''(\underline{w}_0, \underline{w}^+, \underline{w}^*)$ be the result of fixing all values of the variables $\underline{w}'$ in $\mathcal{E}''$ to the corresponding values in $\underline{a}'$. In fact, by construction of $\mathcal{E}''$, we know that no variables $\underline{w}^*$ occur in $\mathcal{E}'''$ (i.e. they have zero-coefficients). Moreover, the variables $\underline{w}_0$ do not contribute to the solution; so let $\mathcal{E}^+(\underline{w}^+)$ be the result of removing from $\mathcal{E}'''$ all terms involving the variables in $\underline{w}_0$. Thus, $\mathcal{E}^+(\underline{w}^+)$ is a system of Diophantine equations with solution $\underline{a}^+$ over $\mathbb{N}$. By Proposition 2.4, $\mathcal{E}^+(\underline{w}^+)$ has a solution $\underline{b}^+$ over $\mathbb{N}$ with footprint of size at most $2(r - r')\log(4(r - r')M)$. It follows that $\mathcal{E}'''(\underline{w}_0, \underline{w}^+, \underline{w}^*)$ has a solution $\underline{0}, \underline{b}^+, \underline{0}$ for tuples $\underline{0}$ of the appropriate length. But then $\underline{a}', \underline{0}, \underline{b}^+, \underline{0}$ is a solution of $\mathcal{E}$, since all equations in $\mathcal{E}'$ are trivially secured by the fixed values $\underline{a}'$. Clearly, this solution has footprint of size at most $2r' + 2(r - r') \log(4(r - r')M)$. □

We finish this section with an easy observation:

Lemma 2.7 *Let $\mathcal{E}$ be a system of extended integer equations, and $\underline{\mathbf{w}}_0, \ldots, \underline{\mathbf{w}}_L$, $\mathbb{N}^*$-vectors. The following are equivalent:*

(i) *$\underline{\mathbf{w}}_0$ is a solution of $\mathcal{E}$ and, for all ℓ ($1 \leq \ell \leq L$), $\underline{\mathbf{w}}_0 + \underline{\mathbf{w}}_\ell$ is a solution of $\mathcal{E}$;*

(ii) *for all $\zeta_1, \ldots, \zeta_L \in \mathbb{N}^*$, $\underline{\mathbf{w}}_0 + \zeta_1 \underline{\mathbf{w}}_1, \ldots, \zeta_L \underline{\mathbf{w}}_L$ is a solution of $\mathcal{E}$.*

Proof. The implication (ii) ⇒ (i) is trivial. For the converse, suppose (i) holds, write $\mathcal{E}$ as $\mathbf{A}\underline{x} + \underline{\mathbf{b}} = \mathbf{F}\underline{x} + \underline{\mathbf{g}}$, and let $\zeta_1, \ldots, \zeta_L$ be extended natural numbers. We wish to establish the vector equation:

$$\mathbf{A}\underline{\mathbf{w}} + \underline{\mathbf{b}} + \zeta_1 \mathbf{A}\underline{\mathbf{w}}_1 + \cdots + \zeta_L \mathbf{A}\underline{\mathbf{w}}_L = \mathbf{F}\underline{\mathbf{w}} + \underline{\mathbf{g}} + \zeta_1 \mathbf{F}\underline{\mathbf{w}}_1 + \cdots + \zeta_L \mathbf{F}\underline{\mathbf{w}}_L. \qquad (2)$$

Consider the jth entry of this vector equation. If $(\mathbf{A}\underline{\mathbf{w}}_0 + \underline{\mathbf{b}})[j] = (\mathbf{F}\underline{\mathbf{w}}_0 + \underline{\mathbf{g}})[j]$ is infinite, then the jth entry of (2) is trivial. On the other hand, if $(\mathbf{A}\underline{\mathbf{w}}_0 + \underline{\mathbf{b}})[j] = (\mathbf{F}\underline{\mathbf{w}}_0 + \underline{\mathbf{g}})[j]$ is finite, then (i) implies $(\mathbf{A}\underline{\mathbf{w}}_\ell)[j]$= $(\mathbf{F}\underline{\mathbf{w}}_\ell)[j]$ for all ℓ ($1 \leq \ell \leq L$), from which the jth entry of (2) again follows. □

2.4 Combinatorics

The following fact will be used in Sec. 5.

Lemma 2.8 *Let $\{V_x\}_{x \in X}$ be a finite family of disjoint (possibly infinite) sets, and let $V = \bigcup_{x \in X} V_x$. Suppose that:*

(i) *$|V|$ is either even or infinite;*

(ii) *for all $y \in X$, $|V_y| \leq \sum_{x \in X \setminus \{y\}} |V_x|$.*

Then the elements of V can be arranged in pairs so that no pair has both its elements from the same V_x.

We might express condition (ii) of the lemma by saying that no set V_y contains an absolute majority of elements in the family.

Proof. If any of the sets V_x is infinite, then, by (ii), at least two are, and the existence of the required pairing is again immediate. Therefore, we may assume that V is finite, and proceed by induction on $|V|$. If $|V| = 0$, there is nothing to show. Otherwise, let $y \in X$ be such that $|V_y|$ is largest, and choose $u \in V_y$. By (ii), there exists $z \neq y$ such that $V_z \neq \emptyset$, so choose z such that $|V_z|$ is largest (for $z \in X \setminus \{y\}$), and choose $v \in V_z$. Now define the family $\{V'_x\}_{x \in X}$ by setting $V'_y = V_y \setminus \{u\}$; $V'_z = V_z \setminus \{v\}$; and $V'_x = V_x$ for any other $x \in X$. Then $\{V'_x\}_{x \in X}$ satisfies (i). To see that it also satisfies (ii), we observe that, if V'_y is strictly largest among the sets $\{V'_x\}$, then the result is immediate since $|V'_y| = |V_y| - 1 \leq \sum_{x \in X \setminus \{y\}} |V_x| - 1 = \sum_{x \in X \setminus \{y\}} |V'_x|$. So we may assume that there exists some $w \in X \setminus \{y, z\}$ such that V'_w is largest among the sets V'_x: $|V'_w| = |V_w| = |V_y| = |V_z| = k$, say. Now if $k = 1$, by (i), there exists $x \in X \setminus \{y, z, w\}$ with $V'_x = V_x \neq \emptyset$ and it is immediate that V'_w does not have an absolute majority. And if $k > 1$, $|V'_w| = |V_w| = |V_y| = |V'_y| + 1 \leq |V'_y| + |V'_z|$, so that V_w again does not have an absolute majority. By inductive hypothesis, the family $\{V'_x\}_{x \in X}$ has a pairing in which no pair has both its elements from the same set V'_x. Now add (u, v) to this pairing. □

The following lemma will be used in Sec. 4.5.

Lemma 2.9 *Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be $\mathbb{N}$-vectors such that $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$. The following are equivalent:*

(a) *for all k ($1 \leq k \leq n$), $u_k \leq \sum_{i \neq k} v_i$;*

(b) *there exists k ($1 \leq k \leq n$) such that: (i) $\sum_{i=1}^{k-1} u_i \leq \sum_{i=k}^n v_i$, ($ii$) $\sum_{i=k+1}^n u_i \leq \sum_{i=1}^k v_i$ and (iii) $u_k \leq \sum_{i=1}^{k-1} v_i + \sum_{i=k+1}^n v_i$.*

We might express condition (a) of the lemma by saying that no integer u_k constitutes an absolute *skew-majority*.

Proof. Suppose (a) holds, and write $S = \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} v_i$. If $n \leq 2$, putting $k = 1$ evidently secures (b); thus, we may assume that $n > 2$. From (a), $u_1 \leq \sum_{i=2}^{n} v_i$; so let k be the largest number ($2 \leq k \leq n-1$) such that $\sum_{i=1}^{k-1} u_i \leq \sum_{i=k}^{n} v_i$. This secures (*i*). If $k < n-1$, then we have $\sum_{i=1}^{k} u_i > \sum_{i=k+1}^{n} v_i$ by the maximality of k; and, using the fact that $\sum_{i=1}^{k} u_i = S - \sum_{i=k+1}^{n} u_i$ and $\sum_{i=k+1}^{n} v_i = S - \sum_{i=1}^{k} v_i$, we certainly obtain (*ii*). On the other hand, if $k = n-1$, then $\sum_{i=k+1}^{n} u_i = u_n$, so that (a) immediately implies (*ii*). Finally, the inequality (*iii*) is likewise an immediate consequence of (a). Conversely, suppose (b) holds, and choose any i ($1 \leq i \leq n$). Then i is in one of the intervals $[1, k-1]$, $[k+1, n]$ or $[k, k]$, so that one of the inequalities (*i*)–(*iii*) certainly implies $u_i \leq \sum_{i' \neq i} v_{i'}$, as required by (a). □

Lemma 2.9 guarantees that the n inequalities in (a), if true, are always 'witnessed' by some value of k ($1 \leq k \leq n$) satisfying the three inequalities in (b).

3 Local configurations in structures

In the sequel, we fix a normal-form $\mathcal{C}^2$1E-formula φ. We use the symbols α, m, M_h, β_h throughout to refer to the parts of φ as indicated in (1), and we aditionally define M to be the ceiling of φ. We denote the number of symbols occurring in φ by $\|\varphi\|$, it being understood that a counting subscript M_h contributes $\lceil \log M_h \rceil$ symbols. In this section, we construct the basic apparatus required to describe certain local configurations in structures interpreting the signature of φ. This apparatus, which is essentially the same as that presented in [5], will allow us to characterize collections of elements in these structures using non-negative integer vectors, and thence to apply techniques from integer linear programming to analyse them.

For technical reasons, we shall work with a signature featuring a number of unary predicates not occurring in φ. Henceforth, let

$$Z = \max(3mM + 1, (mM + 1)^2 + 1). \tag{3}$$

and fix Σ to be the signature of φ together with $(\|\varphi\| + 5\lceil \log Z \rceil)$ additional unary predicates, which we shall refer to as *spare predicates*. Since φ is fixed in the sequel, we refer to any quantity bounded by $p(\|\varphi\|)$, where p is a fixed polynomial, as *polynomially bounded*, or simply *polynomial*. Likewise, quantities bounded by $2^{p(\|\varphi\|)}$ are said to be (*singly*) *exponentially bounded* or (*singly*) *exponential*; and quantities bounded by $2^{2^{p(\|\varphi\|)}}$ are said to be *doubly exponentially bounded* or *doubly exponential*. Thus, $|\Sigma|$ is polynomial, while M and Z are exponential. The quantity Z will feature at various points in the sequel as a 'moderately large' number.

3.1 Basic definitions

A *1-type* is a maximal consistent set of literals over Σ involving only the variable x. Likewise, a *2-type* is a maximal consistent set of literals over Σ involving only the variables x and y and containing the literal $x \neq y$. Here, consistency is understood to take account of the requirement that E is interpreted as an equivalence: every 1-type contains the literal $E(x, x)$; every 2-type contains $E(x, x)$ and $E(y, y)$; and every 2-type contains $E(x, y)$ if and only if it contains $E(y, x)$. We denote by τ^{-1} the 2-type obtained by exchanging the variables x and y in τ, and call τ^{-1} the *inverse* of τ. We denote by $\text{tp}_1(\tau)$ the 1-type obtained by removing from τ any literals containing y; and we denote by $\text{tp}_2(\tau)$ the 1-type obtained by first removing from τ any literals containing x, and then replacing all occurrences of y by x. Evidently, $\text{tp}_2(\tau) = \text{tp}_1(\tau^{-1})$. We equivocate freely between finite sets of formulas and their conjunctions.

Let $\mathfrak{A}$ be any structure interpreting Σ. If $a \in A$, then there exists a unique 1-type $\pi(x)$ such that $\mathfrak{A} \models \pi[a]$; we denote π by $\text{tp}^{\mathfrak{A}}[a]$. If, in addition, $b \in A \setminus \{a\}$, then there exists a unique 2-type $\tau(x, y)$ such that $\mathfrak{A} \models \tau[a, b]$; we denote τ by $\text{tp}^{\mathfrak{A}}[a, b]$. Evidently, $\tau^{-1} = \text{tp}^{\mathfrak{A}}[b, a]$; $\text{tp}_1(\tau) = \text{tp}^{\mathfrak{A}}[a]$; and $\text{tp}_2(\tau) = \text{tp}^{\mathfrak{A}}[b]$. If π is a 1-type, we say that π is *realized* in $\mathfrak{A}$ if there exists $a \in A$ with $\text{tp}^{\mathfrak{A}}[a] = \pi$. If τ is a 2-type, we say that τ is *realized* in $\mathfrak{A}$ if there exist distinct $a, b \in A$ with $\text{tp}^{\mathfrak{A}}[a, b] = \tau$.

We call a 2-type τ *galactic* if it contains the atom $E(x, y)$, and *cosmic* otherwise, i.e. if it contains the atom $\neg E(x, y)$. For any 2-type τ, τ is galactic (cosmic) if and only if τ^{-1} is. Recalling the form (1) of φ, we call the 2-type τ a *ray-type* if $\models \tau \rightarrow \beta_h$ for some h ($1 \leq h \leq m$). If ρ is a ray-type such that ρ^{-1} is also a ray-type, we say that ρ is *invertible*. A ray-type ρ is *polarized* if it is either non-invertible or if $\text{tp}_1(\rho) \neq \text{tp}_2(\rho)$. It will be convenient, in the sequel, to pair polarized, invertible, cosmic ray-types with their inverses: if ρ is a polarized, invertible cosmic ray-type, we refer to the unordered pair (ρ, ρ^{-1}) as a *symmetrized cosmic ray-type*. (We do not require a corresponding notion for other sorts of ray-types.) If τ is a 2-type such that neither τ nor τ^{-1} is a ray-type, we say that τ is *dark*.

Informally, if $\text{tp}^{\mathfrak{A}}[a, b] = \rho$ is a ray-type, we speak of the ordered pair $\langle a, b\rangle$ as a *ray of type* ρ, and we are invited to imagine that this ray is emitted by a and absorbed by b. If ρ is invertible, then b reciprocates with a ray of type ρ^{-1}. Accordingly, we refer to the 1-types $\text{tp}_1(\rho)$ and $\text{tp}_2(\rho)$ as the *emission-type* and *absorption-type* of ρ, respectively. Polarized ray-types (galactic or cosmic) are thus ray-types which are either non-invertible or whose emission and absorption types are distinct. If $\text{tp}^{\mathfrak{A}}[a, b]$ is dark, then neither element emits a ray that is absorbed by the other. By inspection of (1), we see that, if $\mathfrak{A} \models \varphi$, and $a \in A$, then a emits at least one ray. On the other hand, a cannot emit more than M rays of any given type, and indeed cannot emit more than Mm rays in total. The diagram in Fig. 1a illustrates a pair of elements a and b in such a structure $\mathfrak{A}$, with the first sending a ray to the second.

Enumerate the 1-types as $\pi_1, \ldots, \pi_I$. Notice that the number of 1-types, $I = 2^{|\Sigma|-1}$, is singly exponentially bounded. We fix this enumeration for the remainder of this paper. Enumerate the polarized ray-types as $\rho_1, \ldots, \rho_{8J}$. We may choose the enumeration so that the ray-types $\rho_1, \ldots, \rho_{2J}$ are all galactic and invertible, the ray-types $\rho_{2J+1}, \ldots, \rho_{4J}$ are all galactic and non-invertible, the ray-types $\rho_{4J+1}, \ldots, \rho_{6J}$ are all cosmic and invertible, and the ray-types $\rho_{6J+1}, \ldots, \rho_{8J}$ are all cosmic and non-invertible, thus:

$$\underbrace{\rho_1, \ldots, \rho_J, \rho_{J+1}, \ldots, \rho_{2J}}_{\text{galactic, invertible}}, \underbrace{\rho_{2J+1}, \ldots, \rho_{4J}}_{\substack{\text{galactic}\\ \text{non-invertible}}}, \underbrace{\rho_{4J+1}, \ldots, \rho_{5J}, \rho_{5J+1}, \ldots, \rho_{6J}}_{\text{cosmic, invertible}}, \underbrace{\rho_{6J+1}, \ldots, \rho_{8J}}_{\substack{\text{cosmic}\\ \text{non-invertible}}}.$$

We need not worry that there are more invertible than non-invertible ray-types: we can simply 'pad out' the latter with unrealized dummy types. Notice that the quantity J (approximately one eighth the number of polarized ray-types) is singly exponentially bounded. Since the rays in question are polarized, we certainly have $\rho_j \neq \rho_j^{-1}$ for all invertible ray-types ρ_j (galactic or cosmic); hence we may unproblematically arrange the enumeration so that $\rho_j^{-1} = \rho_{J+j}$ for all j in the ranges $1 \leq j \leq J$ and $4J + 1 \leq j \leq 5J$. It follows that the symmetrized cosmic ray-types are exactly the pairs $(\rho_{4J+j}, \rho_{5J+j})$, where $1 \leq j \leq J$. We fix the enumeration $\rho_1, \ldots, \rho_{8J}$ for the remainder of this paper. The following notation will be used later in the paper for j in the range $1 \leq j \leq 2J$. We write:

$$j^* = \begin{cases} j + J & \text{if } 1 \leq j \leq J \\ j - J & \text{if } J < j \leq 2J. \end{cases}$$

Thus, for j in the range in question, ρ_{4J+j} and ρ_{4J+j^*} are mutually inverse cosmic ray-types, and $j^{**} = j$.

3.2 Differentiation and polarization

Recall that Σ features $(\|\varphi\| + 5\lceil\log Z\rceil)$ spare unary predicates, not occurring in φ. We proceed to introduce a class of Σ-structures in which the 1-type of each element encodes useful information about that element's locality. We show that we may without loss of generality restrict attention to such structures.

Say that a Σ-structure $\mathfrak{A}$ is *polarized* if no element sends an invertible ray to any other element with the same 1-type as itself—equivalently, if every ray in $\mathfrak{A}$ is polarized. Likewise, say that $\mathfrak{A}$ is *2-polarized* if it is polarized, and no element sends invertible rays to any two elements with the same 1-type as *each other*. A moment's thought shows that $\mathfrak{A}$ is 2-polarized just in case no two elements with the same 1-type are joined by a chain of at most two invertible rays—i.e. just in case, for all distinct a, b such that $\text{tp}^{\mathfrak{A}}[a] = \text{tp}^{\mathfrak{A}}[b]$, the ray-type $\text{tp}^{\mathfrak{A}}[a, b]$ is not invertible and, moreover, there exists no c, distinct from a and b, such that $\text{tp}^{\mathfrak{A}}[a, c]$ and $\text{tp}^{\mathfrak{A}}[c, b]$ are invertible ray-types.

Recalling the quantity Z defined in (3), say that a Σ-structure $\mathfrak{A}$ is *differentiated* if, for every 1-type π:

(*i*) for every equivalence class B of $\mathfrak{A}$, π is realized either by at most one or by at least Z elements of B;

(*ii*) π is realized either in at most one or in at least Z equivalence classes of $\mathfrak{A}$.

Lemma 3.1 *If φ has a model over some domain A, then φ has a 2-polarized, differentiated model interpreting Σ over A.*

Proof. Let $\mathfrak{A}$ be a model of φ interpreting Σ. Without loss of generality, we can assume that all spare predicates have empty extension in $\mathfrak{A}$. We first construct a 2-polarized model $\mathfrak{A}'$. Consider the graph (A, E), where E is the set of pairs of distinct elements of A joined by a chain of one or two invertible rays:

$$\begin{aligned} E = & \{(a,b) \in A \times A \mid a \neq b \text{ and } \mathrm{tp}^{\mathfrak{A}}[a,b] \text{ is an invertible ray-type}\} \cup \\ & \quad \{(a,b) \in A \times A \mid a \neq b \text{ and, for some } c \in A \setminus \{a,b\}, \\ & \qquad\qquad\qquad\qquad \mathrm{tp}^{\mathfrak{A}}[a,c] \text{ and } \mathrm{tp}^{\mathfrak{A}}[c,b] \text{ are invertible ray-types}\}. \end{aligned}$$

This graph has degree at most $(mM)^2$. Hence its vertices may be coloured using $(mM)^2 + 1$ colours in such a way that no two vertices joined by an edge have the same colour. Now take $\lceil \log((mM)^2 + 1) \rceil \leq \lceil \log Z \rceil$ spare predicates, and re-interpret them so as to encode these colours in the obvious way. The resulting structure, $\mathfrak{A}'$, is evidently 2-polarized.

We next construct a 2-polarized model $\mathfrak{A}''$, in which every 1-type is realized in every equivalence class either at most once or at least Z times. Let B be an equivalence class of $\mathfrak{A}'$, and suppose π is any 1-type realized in $\mathfrak{A}'$ by at least two, and fewer than Z elements of B. Colour these elements with at most $Z - 1$ colours in such a way that each has a different colour. Re-using the same colours, do the same for every 1-type in $\mathfrak{A}'$, and every equivalence class. Colour any remaining elements uniformly with any colour. Take $\lceil \log(Z - 1) \rceil$ additional spare predicates and re-interpret them so as to encode these colours, denoting the resulting structure by $\mathfrak{A}''$. Evidently, if B is an equivalence class, then every 1-type in $\mathfrak{A}''$ is realized either by at most one or by at least Z elements of B. In addition, $\mathfrak{A}''$ is 2-polarized, since it was obtained from $\mathfrak{A}'$ by further differentiating the realized 1-types.

We next construct a 2-polarized differentiated model $\mathfrak{A}'''$, by modifying $\mathfrak{A}''$ so as to ensure that every 1-type is realized in either at most one or in at least than Z equivalence classes. Since we assumed all spare predicates to have empty extensions in $\mathfrak{A}$, and re-interpreted at most $2\lceil \log Z \rceil$ of these to form $\mathfrak{A}''$, we know that $\mathfrak{A}''$ realizes at most $2^{(\|\varphi\| + 2\lceil \log Z \rceil)}$ different 1-types. Let $\mathbb{B}$ be the set of equivalence classes in $\mathfrak{A}$ (equivalently, in $\mathfrak{A}''$), and, for any 1-type π realized in $\mathfrak{A}''$, let $\mathbb{B}_\pi$ be the set of those equivalence classes in which π is realized. We first claim that there exists a subset $\mathbb{B}^*$ of $\mathbb{B}$ with $|\mathbb{B}^*| \leq 2^{(\|\varphi\| + 2\lceil \log Z \rceil))} \cdot (Z - 1)$ such that, for every 1-type π, either $\mathbb{B}_\pi \subseteq \mathbb{B}^*$ or $|\mathbb{B}_\pi \setminus \mathbb{B}^*| \geq Z$. To see this, start with $\mathbb{B}^* = \emptyset$; if any π fails to satisfy the required condition, add all the elements of $\mathbb{B}_\pi$ to $\mathbb{B}^*$, proceeding until there are no more 1-types to consider. This process must terminate in at most $2^{(\|\varphi\|) + 2\lceil \log Z \rceil)}$ rounds, in each of which at most $(Z - 1)$ equivalence classes are added to $\mathbb{B}^*$. Having obtained $\mathbb{B}^*$, take $|\mathbb{B}^*| \leq 2^{(\|\varphi\|) + 2\lceil \log Z \rceil)} \cdot (Z - 1)$ colours, and modify the structure $\mathfrak{A}''$ as follows. For each equivalence class $B \in \mathbb{B}^*$, pick a fresh colour, and colour all elements of B uniformly with that colour. Then pick any colour (not necessarily fresh) and use it to colour all other elements of A uniformly. Encode these colours using at most $\lceil \log(2^{(\|\varphi\| + 2\lceil \log Z \rceil))} \cdot Z) \rceil = (\|\varphi\| + 3\lceil \log Z \rceil)$ additional spare predicates. Let the resulting structure be $\mathfrak{A}'''$. Since each equivalence class is coloured uniformly, the previous step in the construction is not undone, whence $\mathfrak{A}'''$ is differentiated. In addition, $\mathfrak{A}'''$ is 2-polarized, since it was obtained from $\mathfrak{A}''$ by further differentiating the realized 1-types. The number of spare predicates required in the entire construction is at most $(\lceil \log Z \rceil) + (\lceil \log Z \rceil) + (\|\varphi\| + 3\lceil \log Z \rceil) = \|\varphi\| + 5\lceil \log Z \rceil$, so we do not run out. □

3.3 Coupling: galactic and cosmic

Fix 1-types π and π', not necessarily distinct. Recalling the form (1) of φ, we see that the conjunct $\forall x \forall y (x = y \vee \alpha)$ might force a pair of distinct elements realizing these 1-types to be joined by a (galactic or cosmic) ray-type. This situation will careful treatment in the sequel, and we need a mechanism to describe it. Define γ to be the formula

$$(\pi(x) \wedge \pi'(y) \wedge \alpha(x,y) \wedge \alpha(y,x)) \rightarrow \bigvee_{h=1}^{m} (\beta_h(x,y) \vee \beta_h(y,x)) \, .$$

Informally, for pairs of distinct elements, $\gamma(x,y)$ expresses the condition that, if x and y have respective 1-types π and π', and are related in both directions by α, then one of them emits a ray absorbed by the other. We say that π and π' are *galactically coupled*, and write $\pi \overset{g}{\sim} \pi'$, if $\models \forall x \forall y(E(x,y) \wedge x \neq y \rightarrow \gamma)$; and we say that π and π' are *cosmically coupled*, and write $\pi \overset{c}{\sim} \pi'$, if $\models \forall x \forall y(\neg E(x,y) \rightarrow \gamma)$. Galactic and cosmic coupling are important for the following reason. Suppose $\mathfrak{A} \models \varphi$, and a, b are equivalent but distinct elements of A such that $\text{tp}^{\mathfrak{A}}[a] = \pi$ and $\text{tp}^{\mathfrak{A}}[b] = \pi'$. If $\pi \overset{g}{\sim} \pi'$, then either $\text{tp}^{\mathfrak{A}}[a,b]$ or $\text{tp}^{\mathfrak{A}}[b,a]$ (possibly both) is a galactic ray-type. Similarly, suppose a, b are non-equivalent elements of A such that $\text{tp}^{\mathfrak{A}}[a] = \pi$ and $\text{tp}^{\mathfrak{A}}[b] = \pi'$. If $\pi \overset{c}{\sim} \pi'$, then either $\text{tp}^{\mathfrak{A}}[a,b]$ or $\text{tp}^{\mathfrak{A}}[b,a]$ (possibly both) is a cosmic ray-type.

A pair of 1-types that are both numerous in some equivalence class cannot be galactically coupled, with corresponding remarks applying to cosmic coupling. This is shown in the next two lemmas.

Lemma 3.2 *Suppose $\mathfrak{A} \models \varphi$, and let B be any equivalence class of $\mathfrak{A}$. If π and π' are 1-types both realized at least Z times in B, then π and π' are not galactically coupled.*

Proof. Certainly, from (3), $Z \geq mM(mM+1)+1$. Suppose the conditions of the lemma hold, and let $D \subseteq B$ be a set of exactly $mM+1$ elements with 1-type π'. Mark every element of B which absorbs a (galactic) ray emitted by any element of D. There can be at most $mM(mM+1)$ marked elements, so choose an unmarked element $a \in B$ having 1-type π. Since a emits only mM galactic rays, there exists $b \in D$ such that neither $\text{tp}^{\mathfrak{A}}[a,b]$ nor $\text{tp}^{\mathfrak{A}}[b,a]$ is a galactic ray-type. □

Lemma 3.3 *Suppose $\mathfrak{A} \models \varphi$, and π, π' are 1-types for which either of the following holds:*

(*i*) *π is realized in at least Z different equivalence classes of $\mathfrak{A}$, and π' is realized at least Z times in $\mathfrak{A}$;*

(*ii*) *π is realized in some equivalence class B at least Z times, and π' is realized in some equivalence class B' at least Z times, where $B \neq B'$.*

Then π and π' are not cosmically coupled.

Proof. From (3), $Z \geq (mM+1)^2+1$. For condition (*i*), take a collection D of exactly $(mM+1)$ elements realizing π', and for each element b in this collection, mark the equivalence class of b, and mark every equivalence class containing some element a such that $\text{tp}^{\mathfrak{A}}[b,a]$ is a cosmic ray-type. The total number of marked equivalence classes is thus at most $(mM+1)^2$. Now choose an element a realizing π, and lying in an umarked equivalence class. Since a emits at most mM cosmic rays, there exists $b \in D$ such that $\text{tp}^{\mathfrak{A}}[a,b]$ is a dark cosmic 2-type.

For condition (*ii*), take a collection $D \subseteq B'$ of exactly $(mM+1)$ elements realizing π'. For each $b \in D$, mark any element a of B such that $\text{tp}^{\mathfrak{A}}[b,a]$ is a cosmic ray-type. The number of marked elements is at most $mM(mM+1)$, so choose some unmarked $a \in B$ with 1-type π. Since a emits at most mM cosmic rays, there exists $b \in D$ such that $\text{tp}^{\mathfrak{A}}[a,b]$ is a dark cosmic 2-type. □

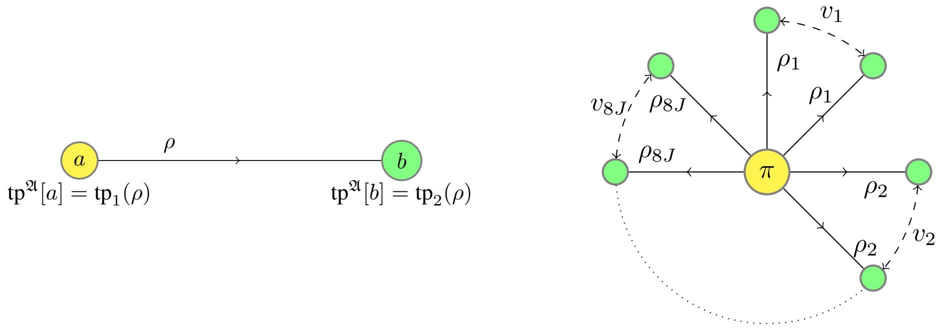
3.4 Local finiteness and star-types

We now construct apparatus for describing the 'local environment' of elements in polarized structure interpreting Σ. A *star-type* is a pair $\sigma = \langle \pi, (v_1, \ldots, v_{8J}) \rangle$ where π is a 1-type and the v_j are cardinal numbers such that $v_j \neq 0$ implies $\text{tp}_1(\rho_j) = \pi$ for all j $(1 \leq j \leq 8J)$. We write $\text{tp}(\sigma) = \pi$ and, abusing our vector notation slightly, $\sigma[j] = v_j$. To motivate this terminology, suppose $\mathfrak{A}$ is a polarized structure interpreting Σ. For any $a \in A$, we define

$$\text{st}^{\mathfrak{A}}[a] = \langle \text{tp}^{\mathfrak{A}}[a], (v_1, \ldots, v_{8J}) \rangle, \tag{4}$$

where $v_j = |\{b \in A : b \neq a \text{ and } \text{tp}^{\mathfrak{A}}[a,b] = \rho_j\}|$ for all j $(1 \leq j \leq 8J)$. Evidently, $\text{st}^{\mathfrak{A}}[a]$ is a star-type. We call $\text{st}^{\mathfrak{A}}[a]$ the *star-type of* a in $\mathfrak{A}$. If $\sigma = \text{st}^{\mathfrak{A}}[a]$ for some $a \in A$, we say σ is realized in $\mathfrak{A}$. We say that $\mathfrak{A}$ is *locally finite* if for all $a \in A$ and all j $(1 \leq j \leq 8J)$, the number of rays of type ρ_j emitted by a—i.e., the quantity $|\{b \in A : b \neq a \text{ and } \text{tp}^{\mathfrak{A}}[a,b] = \rho_j\}|$—is finite. By inspection of (1), if $\mathfrak{A} \models \varphi$, then $\mathfrak{A}$ is locally finite: indeed, the quantities in question are all uniformly bounded by M. Thus, we will only ever need to deal with locally finite structures in the sequel, and we will only ever encounter star-types with finite entries. Notice that this applies even when $\mathfrak{A}$ is infinite.

It is often useful to regard a star-type σ as a finite multiset over the list of polarized ray-types $\rho_1, \ldots, \rho_{8J}$, with multiplicities indicated by $v_1, \ldots, v_{8J}$. Accordingly, we speak informally of σ *emitting* $\sigma[j]$ *rays of type* ρ_j, for



a
ρ
b
$\mathrm{tp}^{\mathfrak{A}}[a] = \mathrm{tp}_1(\rho)$
$\mathrm{tp}^{\mathfrak{A}}[b] = \mathrm{tp}_2(\rho)$
v_1
ρ_1
ρ_1
v_{8J}
ρ_{8J}
ρ_{8J}
π
ρ_2
ρ_2
v_2

(a) (b)

Fig. 1 Depiction of: (a) an element a sending a ray of type ρ to an element b in a structure $\mathfrak{A}$; and (b) a star type $\langle \pi, (v_1, v_2, \ldots, v_{8J})\rangle$, emitting v_j rays of type ρ_j for all j $(1 \leq j \leq 8J)$.

all j $(1 \leq j \leq 8J)$, as depicted in Fig. 1b. We remark in passing that, even in a locally finite structure, there is no bound on the number of non-invertible rays *absorbed* by an element: indeed, it may be that an element absorbs a ray emitted by every other element.

Later in the paper, it will be useful to consider a truncated form of star-types featuring only galactic rays. Recalling that $\rho_1, \ldots, \rho_{4J}$ are the galactic polarized ray-types, we say that a *galactic star-type* is a pair $\langle \pi, (v_1, \ldots, v_{4J})\rangle$ where π is a 1-type and the v_j are cardinal numbers such that $v_j > 0$ implies $\mathrm{tp}_1(\rho_j) = \pi$ for all j $(1 \leq j \leq 4J)$. If $\mathfrak{A}$ is a structure interpreting Σ, and $a \in A$, define

$$\mathrm{st}^{\mathfrak{A}}_{\star}[a] = \langle \mathrm{tp}^{\mathfrak{A}}[a], (v_1, \ldots, v_{4J})\rangle, \tag{5}$$

where the v_j are as in (4). We call $\mathrm{st}^{\mathfrak{A}}_{\star}[a]$ the *galactic star-type of* a in $\mathfrak{A}$. Evidently, $\mathrm{st}^{\mathfrak{A}}_{\star}[a]$ is a galactic star-type. Galactic star-types are simply projections of star-types in which cosmic rays are ignored. They will not be used until Sec. 5. Again, if $\mathfrak{A}$ is locally finite, all values occurring in the realized galactic star-types in $\mathfrak{A}$ will be finite.

A star-type $\sigma = \langle \mathrm{tp}^{\mathfrak{A}}[a], (v_1, \ldots, v_{8J})\rangle$ is *2-polarized* if it does not emit two invertible rays (galactic or cosmic) with the same absorption-type as each other—that is to say, if, for every 1-type π,

$$\sum\{v_j : 1 \leq j \leq 2J,\ \mathrm{tp}_2(\rho_j) = \pi\} + \sum\{v_j : 4J < j \leq 6J,\ \mathrm{tp}_2(\rho_j) = \pi\} \leq 1.$$

A polarized structure interpreting Σ is thus 2-polarized if and only if every star-type it realizes is 2-polarized.

Recalling the form (1) of φ, we say that a 2-type τ is *compatible with* φ if $\models \tau \rightarrow (\alpha(x,y) \wedge \alpha(y,x))$. (Remember that τ by definition contains the literal $x \neq y$.) Evidently, in any model of φ, all realized 2-types are compatible with φ. Similarly, we say that a star-type σ is *compatible with* φ if:

(i) for all j $(1 \leq j \leq 8J)$, if $\sigma[j] > 0$, then ρ_j is compatible with φ; and

(ii) for all h $(1 \leq h \leq m)$, $\sum\{\sigma[j] \mid 1 \leq j \leq 8J,\ \models \rho_j \rightarrow \beta_h\} = M_h$.

Informally, σ is compatible with φ if it emits no rays forbidden by the conjunct $\forall x \forall y (x = y \vee \alpha)$, and emits the right numbers of rays required by the conjuncts $\forall x \exists_{[=M_h]} y (\beta_h \wedge x \neq y)$. Evidently, if $\mathfrak{A}$ is a polarized structure interpreting Σ, then $\mathfrak{A} \models \varphi$ if and only if every star-type and every dark 2-type realized in $\mathfrak{A}$ is compatible with φ. Finally, we observe that, if σ is compatible with φ, then $\sigma[j] \leq M$ for all j $(1 \leq j \leq 8J)$. It follows that the number of star-types compatible with φ is at most $(M+1)^J$—i.e., is doubly exponentially bounded.

Henceforth, we silently assume all star-types to have finite entries.

4 From models to certificates

In this section, we suppose that the $\mathcal{C}^2$1E-formula φ given in (1) has a 2-polarized, differentiated model $\mathfrak{A}$ interpreting the signature Σ, featuring $(\|\varphi\| + 5 \log Z)$ spare predicates. By the Löwenheim-Skolem-Tarski theorem, we may assume without loss of generality that $\mathfrak{A}$ is finite or countably infinite. Moreover, since $\mathfrak{A} \models \varphi$, $\mathfrak{A}$ is locally finite. Our aim is to construct a *certificate* for φ, namely, a data structure satisfying certain conditions that depend only on φ (and not on $\mathfrak{A}$). In Sec. 5, we show that the existence of such a certificate constitutes a guarantee that φ is satisfiable, and in Sec. 6 we show that this certificate may be assumed to be of size bounded by an exponential function of $\|\varphi\|$. This suffices to prove that the satisfiability problem for $\mathcal{C}^2$1E is in NExpTime. In the course of the argument, we further show that, if $\mathfrak{A}$ is finite, the certificate obtained satisfies some simple additional conditions, and that, conversely, the existence of a certificate satisfying these additional conditions constitutes a guarantee that φ is finitely satisfiable. This suffices to prove that the finite satisfiability problem for $\mathcal{C}^2$1E is in NExpTime.

4.1 Numerical characterizations

Let us first enumerate the star-types realized in $\mathfrak{A}$ as $\sigma_1, \ldots, \sigma_K$; we fix this enumeration for the remainder of Sec. 4. Unlike the enumerations $\{\pi_i\}_1^I$ and $\{\rho_j\}_1^{8J}$, the enumeration $\{\sigma_k\}_1^K$ depends on $\mathfrak{A}$. However, the number of 2-polarized star-types compatible with φ is doubly exponentially bounded as a function of $\|(\varphi)\|$; and so, therefore, is K.

The following notion provides the fundamental numerical characterization of subsets of $\mathfrak{A}$ used in this paper. Let $A' \subseteq A$; define the *profile* of A' to be the $\mathbb{N}^*$-vector

$$\text{pr}^{\mathfrak{A}}[A'] =(w_1, \ldots, w_K),$$

where $w_k = |\{a \in A' : \text{st}^{\mathfrak{A}}[a] = \sigma_k\}|$ for all k $(1 \leq k \leq K)$. The profile of A' thus lists the numbers of each star-type realized in A'. While it does not tell us how these elements are connected to each other (or to elements outside A'), it nevertheless gives us relatively detailed information about A' as a whole. It is easy to see that A' is finite if and only if $\text{pr}^{\mathfrak{A}}[A']$ is—i.e. is an $\mathbb{N}$-vector. Of primary interest in the sequel will be the case where A' is an equivalence class or a union of equivalence classes.

In a similar vein, we define the *cosmic spectrum* (or *c-spectrum*) of A' to be the $\mathbb{N}^*$-vector

$$\text{cs}^{\mathfrak{A}}[A'] = (u_1, \ldots, u_{2J}),$$

where $u_j = |\{\langle a, b\rangle \in A' \times A : b \neq a \text{ and } \text{tp}^{\mathfrak{A}}[a, b] = \rho_{4J+j}\}|$ for all j $(1 \leq j \leq 2J)$. The c-spectrum of A' thus lists the total number of rays of each invertible cosmic type emitted by elements of A'. If A' is finite, then $\text{cs}^{\mathfrak{A}}[A']$ is finite; however, the converse implication does not hold, as elements may emit no cosmic rays at all. The definition of $\text{cs}^{\mathfrak{A}}[A']$ takes no account of whether the rays involved are absorbed by other elements of A' or by elements outside A': all that matters is that the rays should be emitted by elements of A'. We may think of the c-spectrum of A' as a more laconic version of its profile. In particular, while the length of the vector $\text{pr}^{\mathfrak{A}}[A']$ is doubly exponentially bounded (as a function of $\|\varphi\|$), the length of $\text{cs}^{\mathfrak{A}}[A']$ is singly exponentially bounded.

The following notation will be useful in the sequel. Define the integer array $\mathbf{U}$, of dimensions $(2J \times K)$, as follows:

$$\mathbf{U}[j, k] = \sigma_k[4J + j] \qquad (1 \leq j \leq 2J, 1 \leq k \leq K). \tag{6}$$

for any $A' \subseteq A$ the following equations relate the c-spectrum of A' to its profile:

$$\text{cs}^{\mathfrak{A}}[A'] = \mathbf{U} \cdot \text{pr}^{\mathfrak{A}}[A'].$$

Note that this equation holds even when A' is infinite.

4.2 Special and ordinary elements

Recall that the model $\mathfrak{A}$ is, by hypothesis, differentiated. For all i $(1 \leq i \leq I)$, execute the following procedure. If π_i is realized in at least Z equivalence classes of $\mathfrak{A}$, select Z of those equivalence classes. If, on the other hand, π_i is realized in just one equivalence class B, select B, and if, in addition, π_i is realized by exactly one element a of B (and hence by exactly one element in the whole of $\mathfrak{A}$), also select every equivalence class B' containing any b such that $\text{tp}^{\mathfrak{A}}[a, b]$ is a cosmic ray-type. Call an equivalence class *special* if it is selected in this process.

Thus, a special equivalence class is one which realizes a 1-type not realized outside that equivalence class, or which absorbs a cosmic ray emitted by an element whose 1-type is realized uniquely, or which has simply been chosen as one of Z equivalence classes in which some 1-type is realized. An equivalence-class that is not special is *ordinary*, and an element is *special* (*ordinary*) if its equivalence class is. Let $A^\dagger$ be the set of special elements, and A^* the set of ordinary elements. Thus, $A = A^\dagger \cup A^*$, and $A^\dagger \neq \emptyset$. Both $A^\dagger$ and A^* are unions of equivalence classes.

Enumerate the special equivalence classes as $B^1, \ldots, B^G$. Thus, $1 \leq G < IZ{+}I(1{+}mM)$, i.e. G is (positive and) singly exponentially bounded. Define $\mathcal{I} = \{i \mid \pi_i \text{ is realized exactly once in } \mathfrak{A}\}$. For all i ($1 \leq i \leq I$) define $\mathcal{G}_i = \{g \in [1, G] \mid \pi_i \text{ is realized at least once in } B^g\}$. Thus, $\mathcal{I}$ and the $\mathcal{G}_i$ are all exponentially bounded. Now consider the following sets of statements:

$$\{(|\mathcal{G}_i| \leq 1) \text{ or } (|\mathcal{G}_i| \geq Z) \mid 1 \leq i \leq I\} \tag{$\mathcal{B}_1$}$$

$$\{\mathcal{G}_i \text{ is a singleton} \mid i \in \mathcal{I}\} \tag{$\mathcal{B}_2$}$$

$$\{(\mathcal{G}_i = \emptyset) \text{ or } (\mathcal{G}_{i'} = \emptyset) \text{ or } (\mathcal{G}_i = \mathcal{G}_{i'} \text{ and } |\mathcal{G}_i| = 1) \mid i, i' \in [1, I] \setminus \mathcal{I},\ \pi_i \overset{c}{\sim} \pi_{i'}\}. \tag{$\mathcal{B}_3$}$$

We write $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$.

Lemma 4.1 *All the statements in $\mathcal{B}$ are true.*

Proof. The statements in $\mathcal{B}_1$ and $\mathcal{B}_2$ are immediate by the construction of the sets $\mathcal{G}_i$ and $\mathcal{I}$ and the fact that $\mathfrak{A}$ is differentiated. For $\mathcal{B}_3$, fix i and i', and suppose that π_i and $\pi_{i'}$ are c-coupled 1-types with $i, i' \notin \mathcal{I}$. Thus, if the first two disjuncts are false, then π_i and $\pi_{i'}$ are each realized in $\mathfrak{A}$ more than once, and hence, since $\mathfrak{A}$ is differentiated, at least Z times. In that case, suppose first that π_i is realized in more than one equivalence class of $\mathfrak{A}$. Since $\mathfrak{A}$ is differentiated, π_i is realized in at least Z equivalence classes of $\mathfrak{A}$. By Lemma 3.3(*i*), then, π_i and π'_i are not c-coupled—a contradiction. If $\pi_{i'}$ is realized in more than one equivalence class of $\mathfrak{A}$, the same argument applies exchanging i and i'. So we may assume that π_i is realized (at least Z times) in a single special equivalence class B^g and $\pi_{i'}$ is realized (at least Z times) in a single special equivalence class $B^{g'}$. Since π_i and $\pi_{i'}$ are c-coupled, Lemma 3.3(*ii*) implies that $g = g'$, and the third disjunct holds as required. □

The set $\mathcal{I}$ and the collection of sets $\mathcal{G}_i$ will form part of the certificate for φ, which we assemble in Sec. 4.6. The statements $\mathcal{B}$ will feature among the conditions to which certificates are subject. Note that the bounds given above on G, $\mathcal{I}$ and the $\mathcal{G}_i$ apply to any 2-polarized, differentiated model $\mathfrak{A}$, finite or infinite.

4.3 Equivalence classes

Let us remind ourselves that our goal in this section is to construct a certificate guaranteeing the (finite) satisfiability of $\mathfrak{A}$. Our strategy will be to tackle the equivalence classes one at a time. Within each equivalence class, say B, the equivalence relation referred to in φ simply becomes the total relation. Thus, we can think of B as a model of a modified version of φ in which E does not feature, so that we are, in effect, dealing with the logic $\mathcal{C}^2$. This observation allows us to characterize single equivalence classes by adapting the linear-programming-based approach used in [5] to show that the (finite) satisfiability problem for $\mathcal{C}^2$ is in NExpTime. Two main technical hurdles need to be overcome in order to extend this analysis to an algorithm for deciding (finite) satisfiability in $\mathcal{C}^2$1E. First, there is no bound on the number of equivalence classes in $\mathfrak{A}$, so that we cannot, for example, simply assemble a collection of separate linear integer programming problems corresponding to the different equivalence classes. Second, we need to take into account the constraints which φ imposes on the cosmic 2-types. The first of these hurdles will be overcome in Sec. 4.4; the second, in Sec. 4.5. For the present, we confine ourselves to the analysis of individual equivalence classes.

Recall the enumeration of the 1-types $\pi_1, \ldots, \pi_I$, polarized ray-types $\rho_1, \ldots, \rho_{8J}$ and star-types $\sigma_1, \ldots, \sigma_K$. We define the following integer constants for all i, j and k ($1 \leq i \leq I$, $1 \leq j \leq 2J$, $1 \leq k \leq K$):

$$\mathbf{p}_{i,k} = \begin{cases} 1 & \text{if } \mathrm{tp}(\sigma_k) = \pi_i; \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{t}_{j,k} = \sigma_k[j].$$

Thus, the equation $\mathbf{p}_{i,k} = 1$ states that any element with star-type σ_k has 1-type π_i, while $\mathbf{t}_{j,k}$ gives the number of rays of (invertible galactic) type ρ_j emitted by any element having star-type σ_k.

In a similar vein, for all i, j, k in the above ranges, all i' ($1 \leq i' \leq I$), and all c, d ($1 \leq c, d \leq mM + 1$), we define the following integer constants:

$$\mathbf{o}^c_{i,i',k} = \begin{cases} 1 & \text{if } \mathrm{tp}(\sigma_k) = \pi_i \text{ and } \sum\{\sigma_k[j] \mid 1 \leq j \leq 4J \text{ and } \mathrm{tp}_2(\rho_j) = \pi_{i'}\} \geq c; \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{o}^*_{i,i',k} = \begin{cases} 1 & \text{if } \mathrm{tp}(\sigma_k) = \pi_i \text{ and } \sum\{\sigma_k[j] \mid 2J < j \leq 4J \text{ and } \mathrm{tp}_2(\rho_j) = \pi_{i'}\} = 0; \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{q}^d_{i,i',k} = \begin{cases} 1 & \text{if } \mathrm{tp}(\sigma_k) = \pi_i \text{ and } \sum\{\sigma_k[j] \mid 4J < j \leq 8J \text{ and } \mathrm{tp}_2(\rho_j) = \pi_{i'}\} \geq d; \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{q}^*_{i,i',k} = \begin{cases} 1 & \text{if } \mathrm{tp}(\sigma_k) = \pi_i \text{ and } \sum\{\sigma_k[j] \mid 6J < j \leq 8J \text{ and } \mathrm{tp}_2(\rho_j) = \pi_{i'}\} = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\mathbf{o}^c_{i,i',k} = 1$ states that any element with star-type σ_k has 1-type π_i and emits at least c galactic rays (invertible or non-invertible) that are absorbed by elements with 1-type $\pi_{i'}$; and $\mathbf{o}^*_{i,i',k} = 1$ states that any element with star-type σ_k has 1-type π_i and emits no *non-invertible* galactic rays that are absorbed by elements with 1-type $\pi_{i'}$. The corresponding equations involving the constants $\mathbf{q}^d_{i,i',k}$ and $\mathbf{q}^*_{i,i',k}$ are interpreted analogously, but with "galactic ray" replaced by "cosmic ray".

For convenience, we collect those constants with indices differing only in the value of k ($1 \leq k \leq K$) into vectors of length K, thus:

$$\begin{aligned} \underline{\mathbf{p}}_i &= (\mathbf{p}_{i,1}, \ldots, \mathbf{p}_{i,K}) & \underline{\mathbf{t}}_j &= (\mathbf{t}_{j,1}, \ldots, \mathbf{t}_{j,K}) \\ \underline{\mathbf{o}}^c_{i,i'} &= (\mathbf{o}^c_{i,i',1}, \ldots, \mathbf{o}^c_{i,i',K}) & \underline{\mathbf{q}}^d_{i,i'} &= (\mathbf{q}^d_{i,i',1}, \ldots, \mathbf{q}^d_{i,i',K}) \\ \underline{\mathbf{o}}^*_{i,i'} &= (\mathbf{o}^*_{i,i',1}, \ldots, \mathbf{o}^*_{i,i',K}) & \underline{\mathbf{q}}^*_{i,i'} &= (\mathbf{q}^*_{i,i',1}, \ldots, \mathbf{q}^*_{i,i',K}). \end{aligned}$$

Let $\underline{w} = (w_1, \ldots, w_K)$ be an $\mathbb{N}^*$-vector. Informally, we may think of the values of these variables as the profile of some equivalence class in $\mathfrak{A}$. Consider the following sets of positive integer clauses:

$$\{\underline{\mathbf{t}}_j \cdot \underline{w} = \underline{\mathbf{t}}_{J+j} \cdot \underline{w} \mid 1 \leq j \leq J\} \tag{$\mathcal{C}^0_1$}$$

$$\{(\underline{\mathbf{p}}_i \cdot \underline{w} \leq 1) \vee (\underline{\mathbf{p}}_i \cdot \underline{w} \geq Z) \mid 1 \leq i \leq I\} \tag{C^0_2}$$

$$\{(\underline{\mathbf{p}}_i \cdot \underline{w} \geq c) \vee (\underline{\mathbf{o}}^c_{i',i} \cdot \underline{w} = 0) \mid 1 \leq i, i' \leq I,\ c = 1, 2\} \tag{$\mathcal{C}^0_3$}$$

$$\{(\underline{\mathbf{p}}_i \cdot \underline{w} > 1) \vee (\underline{\mathbf{o}}^c_{i,i'} \cdot \underline{w} = 0) \vee (\underline{\mathbf{o}}^*_{i',i} \cdot \underline{w} \geq c) \mid 1 \leq i, i' \leq I,\ 1 \leq c \leq mM\} \tag{$\mathcal{C}^0_4$}$$

$$\{(\underline{\mathbf{p}}_i \cdot \underline{w} = 0) \vee (\underline{\mathbf{o}}^*_{i',i} \cdot \underline{w} < c) \vee (\underline{\mathbf{o}}^c_{i,i'} \cdot \underline{w} \geq 1) \mid 1 \leq i, i' \leq I,\ \pi_i \overset{g}{\sim} \pi_{i'},\ c \leq mM + 1\} \tag{$\mathcal{C}^0_5$}$$

$$\{\underline{\mathbf{p}}_i \cdot \underline{w} \leq 1) \mid i \in \mathcal{I}\} \tag{$\mathcal{C}^0_6$}$$

$$\{\underline{\mathbf{q}}^2_{i',i} \cdot \underline{w} = 0) \mid i \in \mathcal{I},\ 1 \leq i' \leq I\} \tag{$\mathcal{C}^0_7$}$$

$$\{\underline{\mathbf{q}}^1_{i',i} \cdot \underline{w} = 0) \mid 1 \leq i, i' \leq I, \mathcal{G}_i = \emptyset\} \tag{$\mathcal{C}^0_8$}$$

$$\{(\underline{\mathbf{p}}_i \cdot \underline{w} \leq 1) \vee (\underline{\mathbf{p}}_{i'} \cdot \underline{w} \leq 1) \mid 1 \leq i, i' \leq I,\ \pi_i \overset{g}{\sim} \pi_{i'}\}. \tag{$\mathcal{C}^0_9$}$$

We write $\mathcal{C}^0 = \mathcal{C}^0_1 \cup \cdots \cup \mathcal{C}^0_9$. Notice that $|\mathcal{C}^0|$—that is, the *number* of clauses in $\mathcal{C}^0$—is singly exponentially bounded, whereas the number K of variables in $\mathcal{C}^0$ is doubly exponentially bounded.

Lemma 4.2 *Suppose B is an equivalence class of $\mathfrak{A}$. Then* $\mathrm{pr}^{\mathfrak{A}}[B]$ *satisfies* $\mathcal{C}^0(\underline{w})$.

We remark that, although all coefficients (variable or constant) mentioned in $\mathcal{C}^0$ are finite, the solution $\mathrm{pr}^{\mathfrak{A}}[B]$ need not be. Similar remarks apply to Lemmas 4.3 and 4.4.

Proof. We write $\underline{w}$ for $\mathrm{pr}^{\mathfrak{A}}[B]$, and consider the sets of clauses $\mathcal{C}^0_1, \ldots, \mathcal{C}^0_9$ in turn.

$\mathcal{C}^0_1$**:** The expression $\underline{\mathbf{t}}_j \cdot \underline{w}$ gives the total number of invertible galactic rays of type ρ_j emitted by elements of B. The expression $\underline{\mathbf{t}}_{J+j} \cdot \underline{w}$ gives the total number of invertible galactic rays of type $\rho_{J+j} = \rho_j^{-1}$ emitted by elements of B. Since B is an equivalence class, these must be equal.

$\mathcal{C}^0_2$**:** The expression $\underline{\mathbf{p}}_i \cdot \underline{w}$ gives the total number of elements of B having 1-type π_i. The statement then follows from the fact that $\mathfrak{A}$ is differentiated.

$\mathcal{C}^0_3$**:** If there are fewer than c elements in B of 1-type π_i, then no element of B (having any 1-type $\pi_{i'}$) emits c or more galactic rays that are absorbed by elements whose 1-type is π_i.

$\mathcal{C}^0_4$**:** If $a \in B$ emits a galactic ray that is absorbed by $b \in B$, then b emits no non-invertible, galactic ray that is absorbed by a. Hence, if a is in addition the unique element of B with 1-type π_i, then b emits no non-invertible, galactic ray absorbed by any element whose 1-type is π_i. Therefore, if a in fact emits at least c rays that are absorbed by elements having 1-type $\pi_{i'}$, there must exist at least c elements of B having 1-type $\pi_{i'}$ and emitting no non-invertible, galactic ray absorbed by any element whose 1-type is π_i.

$\mathcal{C}^0_5$**:** If $a \in B$ has 1-type π_i, and $b \in B$ has 1-type $\pi_{i'}$, where π_i and $\pi_{i'}$ are galactically coupled, then *either* a emits a ray absorbed by b, or b emits a non-invertible ray absorbed by a. Hence, if there is at least one element $a \in B$ having 1-type π_i, and at least c elements of B having 1-type $\pi_{i'}$ that emit no non-invertible, galactic ray absorbed by any element of 1-type π_i (and hence by a), then a emits at least c galactic rays absorbed by elements of type $\pi_{i'}$.

$\mathcal{C}^0_6$**:** If the 1-type π_i is realized exactly once in $\mathfrak{A}$, then it is realized at most once in any equivalence class.

$\mathcal{C}^0_7$**:** If the 1-type π_i is realized exactly once in $\mathfrak{A}$, then no element of B emits two or more cosmic rays absorbed by elements of type π_i.

$\mathcal{C}^0_8$**:** If the 1-type π_i is not realized in $\mathfrak{A}$, then no element of B emits any cosmic rays absorbed by elements of type π_i.

$\mathcal{C}^0_9$**:** If π_i and $\pi_{i'}$ are galactically coupled, then, by Lemma 3.2, these 1-types cannot both be realized at least Z times in B. Since $\mathfrak{A}$ is differentiated, one of them is realized at most once.

□

Focussing specifically on the *ordinary* equivalence classes of $\mathfrak{A}$, consider the following sets of positive integer equations:

$$\{\underline{\mathbf{p}}_i \cdot \underline{w} = 0 \mid 1 \leq i \leq I,\ |\mathcal{G}_i| \leq 1\} \qquad (\mathcal{C}^*_1)$$

$$\{\underline{\mathbf{q}}^*_{i',i} \cdot \underline{w} = 0 \mid i \in \mathcal{I},\ 1 \leq i' \leq I,\ \pi_i \overset{c}{\sim} \pi_{i'}\}. \qquad (\mathcal{C}^*_2)$$

We write $\mathcal{C}^* = \mathcal{C}^0 \cup \mathcal{C}^*_1 \cup \mathcal{C}^*_2$. Again, we see that $|\mathcal{C}^*|$ is singly exponentially s bounded.

Lemma 4.3 *Suppose B is an ordinary equivalence class of $\mathfrak{A}$. Then* $\mathrm{pr}^{\mathfrak{A}}[B]$ *satisfies* $\mathcal{C}^*(\underline{w})$.

Proof. We write $\underline{w}$ for $\mathrm{pr}^{\mathfrak{A}}[B]$. The constraints in $\mathcal{C}^0$ are dealt with by Lemma 4.2. We consider the sets of clauses $\mathcal{C}^*_1$ and $\mathcal{C}^*_2$ in turn.

$\mathcal{C}^*_1$**:** If π_i is realized in at most one special equivalence class, then it is not realized in any ordinary equivalence class, since $Z > 1$.

$\mathcal{C}^*_2$**:** Suppose π_i is realized exactly once in $\mathfrak{A}$, say by the element a. By definition, no ordinary element absorbs a ray emitted by a. Hence, if b is an ordinary element, of 1-type $\pi_{i'}$, such that π_i and $\pi_{i'}$ are c-coupled, b must emit a non-invertible ray absorbed by a. Hence B cannot contain an element of 1-type $\pi_{i'}$ that emits no cosmic ray absorbed by any element of type π_i.

□

Turning now to the *special* equivalence classes of $\mathfrak{A}$, namely, $B^1, \ldots, B^G$, let $\underline{w}^g$ be a K-tuple of fresh variables for all g $(1 \leq g \leq G)$. We may think of the values of these variables as the profile of B^g in $\mathfrak{A}$. Write $\underline{1}$ for the vector $(1, \ldots, 1)$ of length K. For any g, consider the following sets of positive integer clauses:

$$\{\underline{\mathbf{p}}_i \cdot \underline{w}^g = 0 \mid 1 \leq i \leq I \text{ and } g \notin \mathcal{G}_i\} \tag{$\mathcal{C}^g_1$}$$

$$\{\underline{\mathbf{p}}_i \cdot \underline{w}^g \geq 1 \mid 1 \leq i \leq I \text{ and } g \in \mathcal{G}_i\} \tag{$\mathcal{C}^g_2$}$$

$$\{\underline{\mathbf{p}}_i \cdot \underline{w}^g \geq 2 \mid 1 \leq i \leq I,\ \mathcal{G}_i = \{g\} \text{ and } i \notin \mathcal{I}\} \tag{$\mathcal{C}^g_3$}$$

$$\{\underline{1} \cdot \underline{w}^g \geq 1\} \tag{$\mathcal{C}^g_4$}$$

$$\{\underline{\mathbf{q}}^1_{i',i} \cdot \underline{w}^g = 0 \mid 1 \leq i, i' \leq I,\ \mathcal{G}_i = \{g\}\} \tag{$\mathcal{C}^g_5$}$$

$$\left\{ (\underline{\mathbf{q}}^d_{i,i'} \cdot \underline{w}^g = 0) \vee \left(\sum_{1 \leq h \leq G}^{h \neq g} \underline{\mathbf{q}}^*_{i',i} \cdot \underline{w}^h \geq d \right) \middle| \, i \in \mathcal{I},\ \mathcal{G}_i = \{g\},\ 1 \leq i' \leq I, d \leq mM \right\} \tag{$\mathcal{C}^g_6$}$$

$$\left\{ (\underline{\mathbf{q}}^d_{i,i'} \cdot \underline{w}^g \geq 1) \vee \left(\sum_{1 \leq h \leq G}^{h \neq g} \underline{\mathbf{q}}^*_{i',i} \cdot \underline{w}^h < d \right) \middle| \, 1 \leq i, i' \leq I,\ g \in \mathcal{G}_i, \pi_i \overset{c}{\sim} \pi_{i'}, d \leq mM + 1 \right\}. \tag{$\mathcal{C}^g_7$}$$

We remark that $\mathcal{C}^g_1$–$\mathcal{C}^g_5$ involve only the variables $\underline{w}^g$; by contrast, $\mathcal{C}^g_6$ and $\mathcal{C}^g_7$ involve all the variables $\underline{w}^1, \ldots, \underline{w}^G$. For all g $(1 \leq g \leq G)$, we write $\mathcal{C}^g = \mathcal{C}^g_1 \cup \cdots \cup \mathcal{C}^g_7$; again, $|\mathcal{C}^g|$ is singly exponentially bounded.

Lemma 4.4 *For all g $(1 \leq g \leq G)$, the (KG)-tuple* $\mathrm{pr}^{\mathfrak{A}}[B^1], \ldots, \mathrm{pr}^{\mathfrak{A}}[B^G]$ *satisfies* $\mathcal{C}^g(\underline{w}^1, \ldots, \underline{w}^G)$.

Proof. For all g $(1 \leq g \leq G)$, we write $\underline{w}^g$ for $\mathrm{pr}^{\mathfrak{A}}[B^g]$. As we observed in the proof of Lemma 4.2, $\underline{\mathbf{p}}_i \cdot \underline{w}^g$ gives the number of elements in B^g realizing the 1-type π_i. Similarly, the expression $\sum\{\underline{\mathbf{q}}^*_{i',i} \cdot \underline{w}^h \mid 1 \leq h \leq G,\ h \neq g\}$ gives the number of elements in special equivalence classes other than B^g having 1-type $\pi_{i'}$, and emitting no non-invertible, cosmic ray absorbed by any element whose 1-type is π_i. Fixing g, we consider the sets of clauses $\mathcal{C}^g_1, \ldots, \mathcal{C}^g_7$ in turn.

$\mathcal{C}^g_1$: If $g \notin \mathcal{G}_i$, then π_i is not realized in B^g.

$\mathcal{C}^g_2$: If $g \in \mathcal{G}_i$, then π_i is realized in B^g.

$\mathcal{C}^g_3$: If $\mathcal{G}_i = \{g\}$, but $i \notin \mathcal{I}$, then π_i is realized in B^g, but not uniquely.

$\mathcal{C}^g_4$: B^g is non-empty.

$\mathcal{C}^g_5$: If $\mathcal{G}_i = \{g\}$, then π_i is realized only in B^g, and so no element of B^g (of any 1-type $\pi_{i'}$) emits a cosmic ray absorbed by an element of 1-type π_i.

$\mathcal{C}^g_6$: If $a \in B^g$ emits a cosmic ray that is absorbed by an element b of a special equivalence class B^h (where $h \neq g$), then b emits no *non-invertible* cosmic ray that is absorbed by a. Hence, if a is in addition the unique element realizing the 1-type π_i, then b emits no non-invertible, cosmic ray absorbed by any element whose 1-type is π_i. Therefore, if a in fact emits at least d cosmic rays to be absorbed by elements having 1-type $\pi_{i'}$ (with these elements being special, by definition), there must exist at least d elements in the various special equivalence classes other than B^g having 1-type π' and emitting no non-invertible, cosmic ray absorbed by any element whose 1-type is π_i.

$\mathcal{C}^g_7$: Let π_i and $\pi_{i'}$ be cosmically coupled, and suppose $g \in \mathcal{G}_i$. Pick $a \in B^g$ with 1-type π_i. If b is an element of a special equivalence class B^h $(h \neq g)$, having 1-type $\pi_{i'}$, then *either* a emits a cosmic ray absorbed by b, *or* b emits a non-invertible cosmic ray absorbed by a. Hence, if a emits fewer than d cosmic rays absorbed by elements of type $\pi_{i'}$, then there must be fewer than d elements in special equivalence classes other than B^g having 1-type $\pi_{i'}$ that emit no non-invertible, cosmic ray absorbed by a, and hence fewer than d elements in special equivalence classes other than B^g having 1-type $\pi_{i'}$ that emit no non-invertible, cosmic ray absorbed by any element of 1-type π_i.

□

Writing $\underline{w}^\dagger$ to denote the (KG)-tuple $\underline{w}^1, \ldots, \underline{w}^G$, define the set of positive integer clauses $\mathcal{C}^\dagger$ to be

$$\mathcal{C}^\dagger(\underline{w}^\dagger) = \bigcup_{g=1}^{G} \left(\mathcal{C}^0(\underline{w}^g) \cup \mathcal{C}^g(\underline{w}^\dagger)\right).$$

The tuple (KG)-tuple $\mathrm{pr}^{\mathfrak{A}}[B^1], \ldots, \mathrm{pr}^{\mathfrak{A}}[B^G]$ will form part of the certificate for φ, which we assemble in Sec. 4.6. The positive integer clauses $\mathcal{C}^\dagger(\underline{w}^\dagger)$ will feature among the conditions to which certificates are subject. The positive integer clauses $\mathcal{C}^*(\underline{w})$, by contrast, will feature only implicitly in these conditions, as part of additional machinery constructed in Secs. 4.4–4.5.

4.4 Clusters

We now group the equivalence classes of $\mathfrak{A}$ into larger units, called clusters. The number of clusters will be bounded as a (doubly exponential) function of $\|\varphi\|$; the number of equivalence classes within each cluster, by contrast, will not be bounded *a priori*. However, within each cluster, any equivalence class will be characterized by a linear combination of a fixed set of constant $\mathbb{N}^*$-vectors ($\mathbb{N}$-vectors in case $\mathfrak{A}$ is finite). These vectors will form a part of the certificate for φ, which we assemble in Sec. 4.6. We remind the reader of the notions of *profile* and *c-spectrum*, established in Sec. 4.1, as well as the matrix $\mathbf{U}$ relating them, and defined in (6).

We first consider the special elements $A^\dagger = B^1 \cup \cdots \cup B^G$. Here, clustering is degenerate: we define a *special cluster* to be a special equivalence class, and we list the special clusters as $C^1, \ldots, C^G$, where $C^g = B^g$ for all g ($1 \leq g \leq G$). This completes the definition of the special clusters of $\mathfrak{A}$.

Now let us turn to the ordinary elements of $\mathfrak{A}$, namely the set $A^* = A \setminus A^\dagger$. The argument here is rather intricate, and so we break it up into three stages.

Hyper-clusters

Lemma 4.3 states that the profile of any ordinary equivalence class satisfies the clauses $\mathcal{C}^*(\underline{w})$. Now replace any clause in $\mathcal{C}^*$ by one of its disjuncts, so that a mixed system of positive integer equations and inequalities results. Enumerate the systems obtained in this way as $\mathcal{Q}_1, \ldots, \mathcal{Q}_{\mathfrak{z}}$. Thus $\mathfrak{z}$ is doubly exponentially bounded. For each z ($1 \leq z \leq \mathfrak{z}$), let $\mathbb{E}_z$ denote the set of those ordinary equivalence classes B such that z is the smallest integer for which $\mathrm{pr}^{\mathfrak{A}}[B]$ satisfies $\mathcal{Q}_z$; and let $E_z = \bigcup \mathbb{E}_z$. Discarding any empty E_z and re-numbering if necessary, we ensure that $E_1, \ldots, E_{\mathfrak{z}}$ is a partition of A^*. We call any set E_z a *hyper-cluster*. Thus, the hyper-cluster E_z is a union of equivalence classes all of which have profiles satisfying $\mathcal{Q}_z$. Hyper-clusters are by construction non-empty: if $A^* = \emptyset$, we have $\mathfrak{z} = 0$, i.e. there are no hyper-clusters at all. By adding slack variables to $\underline{w}$, each $\mathcal{Q}_z$ can be written as a system of positive integer equations

$$\mathcal{E}_z : \mathbf{A}_z \underline{w} + \underline{\mathbf{b}}_z = \mathbf{F}_z \underline{w} + \underline{\mathbf{g}}_z. \tag{7}$$

We may take the dimensions of all the matrices $\mathbf{A}_z$ and $\mathbf{F}_z$ to be $R \times K^*$, where $R = |\mathcal{C}^*|$ and $K \leq K^* \leq K + R$. Thus, R is singly exponential in $\|\varphi\|$, and K^* doubly exponential. We remark that $\underline{w}$ features at most R slack variables.

Super-clusters

Fixing some value z ($1 \leq z \leq \mathfrak{z}$), consider any equivalence class $B \subseteq E_z$. Thus, $\mathrm{pr}^{\mathfrak{A}}[B]$, together with appropriate values for slack variables, yields an $\mathbb{N}^*$-vector satisfying (7). Recall now the matrix $\mathbf{U}$ defined in (6), so that we have $\mathbf{U} \cdot \mathrm{pr}^{\mathfrak{A}}[B] = \mathrm{cs}^{\mathfrak{A}}[B]$. By adding extra columns of zeros to $\mathbf{U}$ to accommodate the slack variables, we see that $\mathrm{pr}^{\mathfrak{A}}[B]$ (together with the chosen values for slack variables) satisfies not only (7), but also:

$$\mathbf{U}\underline{w} = \mathrm{cs}^{\mathfrak{A}}[B]. \tag{8}$$

Observe that the total number of equations in the combined system (7)–(8) is $R + 2J$. This system has a solution $\mathrm{pr}^{\mathfrak{A}}[B]$ over $\mathbb{N}^*$ (over $\mathbb{N}$ if $\mathfrak{A}$ is finite), and, by inspection of $\mathcal{C}^*$ and $\mathbf{U}$, its *variable* coefficients are bounded by $Mm + 1$. Writing

$$K_0 = 2(R + 2J)\log(4(R + 2J)(Mm + 1)),$$

Corollary 2.6 guarantees that (7) and (8) have a solution over $\mathbb{N}^*$ in which at most K_0 values are non-zero—i.e. with footprint of cardinality at most K_0. Indeed, if $\mathfrak{A}$ is finite, Proposition 2.4 guarantees that (7) and (8) have a solution over $\mathbb{N}$ with footprint of cardinality at most K_0. Observe that K_0 is singly exponentially bounded. Notice that it is crucial here that the bound returned by Corollary 2.6 (Proposition 2.4) is independent of the constant coefficients in the system of equations to which it applies; indeed, equations (8) involve constant coefficients in $\mathrm{cs}^{\mathfrak{A}}[B]$ on which there is no *a priori* bound.

Continuing to fix z, list all subsets of the set of variables $\{w_1, \ldots, w_K\}$ of size at most K_0 as $\omega_1, \ldots, \omega_{\mathfrak{y}}$. Thus, $\mathfrak{y}$ is doubly exponentially bounded. For all y ($1 \leq y \leq \mathfrak{y}$), let $\mathbb{D}_{y,z}$ be the set of equivalence classes $B \subseteq E_z$ for which y is the smallest number such that (7) and (8) have a solution with footprint ω_y; and let $D_{y,z} = \bigcup \mathbb{D}_{y,z}$. Discarding any empty $D_{y,z}$ (notice that this depends on z) and re-numbering if necessary, we may assume that each $D_{y,z}$ is non-empty, whence the sets $D_{1,z}, \ldots, D_{\mathfrak{y}(z),z}$ partition E_z, where $\mathfrak{y}(z) \leq \mathfrak{y}$. We call any such set $D_{y,z}$ a *super-cluster*. Thus, if B is an equivalence class included in $D_{y,z}$, then the system of equations (7) and (8) have a solution with footprint, say, $\omega_{y,z}$ of cardinality at most K_0. We remark that, while $\mathrm{pr}^{\mathfrak{A}}[B]$ (together with appropriate values for slack variables) is a solution of (7) and (8), it might not have footprint $\omega_{y,z}$.

Fixing also the value y ($1 \leq y \leq \mathfrak{y}(z)$), let $\mathcal{E}_{y,z}$ be the result of ignoring all terms in $\mathcal{E}_z$ involving variables outside the footprint $\omega_{y,z}$. Thus, $\mathcal{E}_{y,z}$ is a system of positive integer equations involving at most K_0 variables, with all coefficients bounded by $Mm + 1$. (Remember: $\mathcal{E}_z$ does not include the equations (8).) We now establish the following: there exist a set of $\mathbb{N}^*$-vectors, $\mathbf{W}$ and a list of $\mathbb{N}$-vectors $\underline{\mathbf{w}}_1, \ldots \underline{\mathbf{w}}_L$ such that, for any equivalence class $B \subseteq D_{y,z}$, there exists a vector in the set

$$\{\underline{\mathbf{w}}_0 + \zeta_1 \underline{\mathbf{w}}_1 + \cdots + \zeta_L \underline{\mathbf{w}}_L \mid \underline{\mathbf{w}}_0 \in \mathbf{W} \text{ and } \zeta_1, \ldots, \zeta_L \in \mathbb{N}\}.$$

satisfying (7) and (8). Moreover, for all $\underline{\mathbf{w}}_0 \in \mathbf{W}$ and all ℓ ($1 \leq \ell \leq L$), both $\underline{\mathbf{w}}_0$ and $\underline{\mathbf{w}}_0 + \underline{\mathbf{w}}_\ell$ are solutions (over $\mathbb{N}^*$) of the system of equations $\mathcal{E}_z$ given in (7). In addition, the $\mathbb{N}^*$-vectors $\mathbf{W}^{y,z}$ are doubly exponentially finitely bounded and the $\mathbb{N}$-vectors $\underline{\mathbf{w}}_L$ doubly exponentially bounded. Finally, $\mathfrak{A}$ is finite, no infinite values occur: i.e., $\mathbf{W}^{y,z}$ is a set of doubly exponentially bounded $\mathbb{N}$-vectors.

We deal first with the case where $\mathfrak{A}$ is finite, since it involves less clutter. Since $\mathcal{E}_{y,z}$ has a solution over $\mathbb{N}$, Corollary 2.3 guarantees the existence of a set of $\mathbb{N}$-vectors, $\mathbf{W}^{y,z}$ and a list of $\mathbb{N}$-vectors $\underline{\mathbf{w}}_1^{y,z}, \ldots \underline{\mathbf{w}}_L^{y,z}$ such that the set of solutions of (7) over $\mathbb{N}$ with footprint $\omega_{y,z}$ is exactly

$$\{\underline{\mathbf{w}}_0 + \zeta_1 \underline{\mathbf{w}}_1^{y,z} + \cdots + \zeta_L \underline{\mathbf{w}}_L^{y,z} \mid \underline{\mathbf{w}}_0 \in \mathbf{W}^{y,z} \text{ and } \zeta_1, \ldots, \zeta_L \in \mathbb{N}\}.$$

In particular, for any equivalence class $B \subseteq D_{y,z}$, there exists a vector in this set satisfying (7) and (8). Note that we take vectors $\underline{\mathbf{w}}_0, \underline{\mathbf{w}}_1^{y,z}, \ldots, \underline{\mathbf{w}}_L^{y,z}$ here to have the same length as $\underline{w}$, it being understood that all entries corresponding to variables lying outside the footprint $\omega_{y,z}$ are zero. All these vectors are non-negative and bounded by $((K_0 + 1)(Mm + 1) + 1)^{(R+2J)}$—i.e., are doubly exponentially bounded. It follows that $|\mathbf{W}^{y,z}|$ and L are bounded by $(1 + ((K_0 + 1)(Mm + 1) + 1)^{(R+2J)})^{K_0}$—again, doubly exponentially. It immediately follows from the above observations that, for all $\underline{\mathbf{w}}_0 \in \mathbf{W}$ and all ℓ ($1 \leq \ell \leq L$), both $\underline{\mathbf{w}}_0$ and $\underline{\mathbf{w}}_0 + \underline{\mathbf{w}}_\ell$ are solutions (over $\mathbb{N}$) of the system of Diophantine equations $\mathcal{E}_z$ given in (7).

Now consider the case where $\mathfrak{A}$ is infinite, so that the combined system of positive integer equations (7)–(8) has a solution over $\mathbb{N}^*$ with footprint $\omega_{y,z}$. Recalling the discussion of *minimally finite solutions* from Sec. 2.3, let us identify the collection of variables $\omega_{y,z}^-$ taking finite values in any minimally finite solution with footprint $\omega_{y,z}$, and let us take $\mathcal{E}_{y,z}^-$ to be the result of ignoring all those equations in $\mathcal{E}_{y,z}$ which involve no infinite terms in any minimally finite solutions (i.e., have no infinite constant coefficients and no non-zero coefficients of any variables outside $\omega_{y,z}^-$). Let us write $\mathcal{E}_{y,z}^-$ as $\mathbf{A}_{y,z}^- \underline{w}^- + \underline{\mathbf{b}}_{y,z}^- = \mathbf{F}_{y,z}^- \underline{w}^- + \underline{\mathbf{g}}_{y,z}^-$. Thus, $\mathcal{E}_{y,z}^-$ is a system of Diophantine equations with a solution over $\mathbb{N}$, and by exactly the same reasoning as in the finite case, there exist a set of $\mathbb{N}$-vectors, $\mathbf{W}^-$ and a list of $\mathbb{N}$-vectors $\underline{\mathbf{w}}_1^-, \ldots \underline{\mathbf{w}}_L^-$, all doubly exponentially bounded, such that the set of solutions of $\mathcal{E}_{y,z}^-$ over $\mathbb{N}$ is exactly

$$\{\underline{\mathbf{w}}_0^- + \zeta_1 \underline{\mathbf{w}}_1^- + \cdots + \zeta_L \underline{\mathbf{w}}_L^- \mid \underline{\mathbf{w}}_0^- \in \mathbf{W}^- \text{ and } \zeta_1, \ldots, \zeta_L \in \mathbb{N}\}.$$

Notice that, in this case we must have, for all ℓ ($1 \leq \ell \leq L$), $\mathbf{A}_{y,z}^- \underline{\mathbf{w}}_\ell^- = \mathbf{F}_{y,z}^- \underline{\mathbf{w}}_\ell^-$. We now provide values to the remaining variables as expected: we give all entries in the vectors $\mathbf{W}$ corresponding to variables of $\omega_{y,z} \setminus \omega_{y,z}^-$

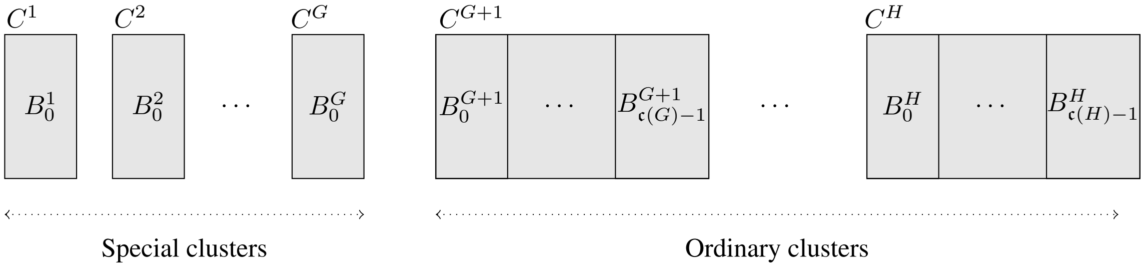


Fig. 2 The organization of equivalence classes into clusters (finite case).

the value $\aleph_0$, and all entries corresponding to variables outside $\omega_{y,z}$ the value 0; and we give all entries in the $\underline{\mathbf{w}}_\ell$ corresponding to variables outside $\omega^-_{y,z}$ the value 0. That is, we let $\mathbf{W} = \{(\underline{\mathbf{w}}^-_0, \underline{\aleph}_0, \underline{0}) \mid \underline{\mathbf{w}}^-_0 \in \mathbf{W}'\}$, and we let $\underline{\mathbf{w}}_\ell = (\underline{\mathbf{w}}^-_\ell, \underline{0}, \underline{0})$ for all ℓ $(1 \leq \ell \leq L)$. It is immediate that $\mathbf{W}$ and the $\underline{\mathbf{w}}_\ell$ have the properties claimed above.

Ordinary clusters

Continuing to fix z and y, let us enumerate $\mathbf{W}^{y,z}$ as $\underline{\mathbf{w}}^{1,y,z}_0, \ldots, \underline{\mathbf{w}}^{\mathfrak{x}(y,z),y,z}_0$, where $\mathfrak{x}(y,z) = |\mathbf{W}^{y,z}|$. For all x $(1 \leq x \leq \mathfrak{x}(y,z))$, let $\mathbb{C}_{x,y,z}$ be the set of equivalence classes $B \subseteq D_{z,y}$ such that x is the smallest value for which there exists a vector of the form $\underline{\mathbf{w}}^{x,y,z}_0 + \zeta_1 \underline{\mathbf{w}}^{y,z}_1 + \cdots + \zeta_L \underline{\mathbf{w}}^{y,z}_L$ satisfying (7) and (8); and let $C_{x,y,z} = \bigcup \mathbb{C}_{x,y,z}$. Again, by discarding any empty $C_{x,y,z}$ and re-numbering if necessary, we may assume that the $C_{1,y,z}, \ldots, C_{\mathfrak{x}(z,y),y,z}$ partition $D_{y,z}$. We call any set $C_{x,y,z}$ an *ordinary cluster*. For notational convenience, we define, for each x $(1 \leq x \leq \mathfrak{x}(y,z))$ and each ℓ $(1 \leq \ell \leq L)$, $\underline{\mathbf{w}}^{x,y,z}_\ell = \underline{\mathbf{w}}^{y,z}_\ell$. (That is: when $\ell \geq 1$, we allow ourselves to add redundant x-superscripts to $\underline{\mathbf{w}}^{y,z}_\ell$.) It follows that, if B is an equivalence class included in $C_{x,z,y}$, then there exists a vector of the form

$$\underline{\mathbf{w}}^{x,y,z}_0 + \zeta_1 \underline{\mathbf{w}}^{x,y,z}_1 + \cdots + \zeta_L \underline{\mathbf{w}}^{x,y,z}_L,$$

where $\zeta_1, \ldots, \zeta_L \in \mathbb{N}$, satisfying (7) and (8). We remind ourselves at this point that the only infinite values occurring in this expression are in the vectors $\underline{\mathbf{w}}^{x,y,z}_0$; and if $\mathfrak{A}$ is finite, there are no infinite values at all.

Almost there. Let us say that a *cluster* is a special cluster or ordinary cluster; let us enumerate the ordinary clusters $C_{x,y,z}$ $(1 \leq z \leq \mathfrak{z}, 1 \leq y \leq \mathfrak{y}(z), 1 \leq x \leq \mathfrak{x}(y,z))$ as $C^{G+1}, \ldots, C^H$; and let us re-index the vectors $\underline{\mathbf{w}}^{x,y,z}_\ell$ as $\underline{\mathbf{w}}^h_\ell$ in a corresponding fashion. In addition, if h $(G < h \leq H)$ is the new index corresponding to the triple (x,y,z), we write $\mathcal{Q}^h = \mathcal{Q}_z$ and $\mathcal{E}^h = \mathcal{E}_z$. In this way, the sequence $C^1, \ldots, C^G, C^{G+1}, \ldots, C^H$ enumerates all the clusters of $\mathfrak{A}$, with the special clusters first. For all h $(G < h \leq H)$ and all ℓ $(1 \leq \ell \leq L)$, $\underline{\mathbf{w}}^h_0$ and $\underline{\mathbf{w}}^h_0 + \underline{\mathbf{w}}^h_\ell$ are solutions of $\mathcal{E}^h$. Furthermore, if B is an equivalence class included in the ordinary cluster C^h $(G < h \leq H)$, then there exists an $\mathbb{N}^*$-vector

$$\underline{w} = \underline{\mathbf{w}}^h_0 + \zeta_1 \underline{\mathbf{w}}^h_1 + \cdots + \zeta_L \underline{\mathbf{w}}^h_L, \tag{9}$$

satisfying $\mathcal{E}^h$ together with the equations $\mathbf{U}\underline{w} = \text{cs}^{\mathfrak{A}}[B]$. If $\mathfrak{A}$ is finite, $\underline{w}$ is an $\mathbb{N}$-vector. This completes the definition of the clusters of $\mathfrak{A}$. Observe that the total number of clusters, H, is doubly exponentially bounded.

For all h $(1 \leq h \leq H)$, let $\mathfrak{c}(h)$ be the number of equivalence classes included in C^h, and list these equivalence classes as B^h_s $(0 \leq s < \mathfrak{c}(h))$. Note that we allow the possibility that $\mathfrak{c}(h)$ may be $\aleph_0$; of course, if $\mathfrak{A}$ is finite, then $\mathfrak{c}(h)$ is finite. Remember also that each special cluster consists of a single (special) equivalence class; thus, for $1 \leq g \leq G$, we have $\mathfrak{c}(g) = 1$ and $B^g_0 = B^g$. The resulting arrangement of clusters is shown, for the case where $\mathfrak{A}$ is finite, in Fig. 2; if $\mathfrak{A}$ is infinite, an ordinary cluster C^h may contain $\aleph_0$ equivalence classes $B^h_0, B^h_1, \ldots$, but the picture is essentially the same.

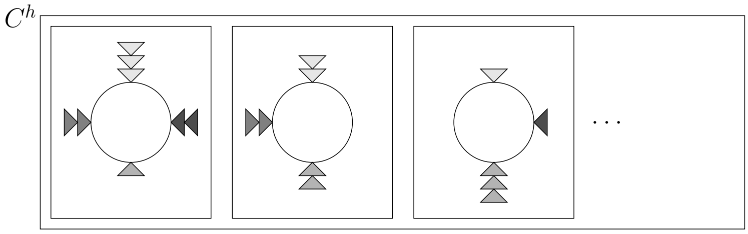


Fig. 3 The modular structure of c-spectra of equivalence classes (depicted as small rectangles) in the ordinary cluster C^h: each 'core' (depicted as a circle) has c-spectrum $\mathbf{u}_0^h$; each 'peripheral constellation' has c-spectrum $\mathbf{u}_\ell^h$ ($1 \leq \ell \leq L$, with ℓ indicated by the shading).

Summary

Having decomposed $\mathfrak{A}$ into special clusters $C^1, \ldots, C^G$ and ordinary clusters $C^{G+1}, \ldots, C^H$, let us summarize what we know about the latter. For notational convenience, we define the $\mathbb{N}^*$-vectors

$$\underline{\mathbf{u}}_\ell^h = \mathbf{U}\underline{\mathbf{w}}_\ell^h \tag{10}$$

for all h ($G < h \leq H$) and ℓ ($0 \leq \ell \leq L$). Again, if $\mathfrak{A}$ is finite, these will be $\mathbb{N}$-vectors. For any value h in this range, the following has been established.

(*i*) The cluster C^h is a union of equivalence classes, and for each of these equivalence classes, B, there exist non-negative integers $\zeta_1, \ldots, \zeta_L$ (depending on B) such that, by (8), (9) and (10),

$$\mathrm{cs}^{\mathfrak{A}}[B] =\underline{\mathbf{u}}_0^h + \zeta_1\underline{\mathbf{u}}_1^h + \cdots + \zeta_L\underline{\mathbf{u}}_L^h. \tag{11}$$

(*ii*) For all ℓ ($1 \leq \ell \leq L$), the $\mathbb{N}^*$-vectors ($\mathbb{N}$-vectors) $\underline{\mathbf{w}}_0^h$ and $\underline{\mathbf{w}}_0^h + \underline{\mathbf{w}}_\ell^h$ are solutions of $\mathcal{E}^h$. Hence, by Lemma 2.7, for all $\zeta_1, \ldots, \zeta_L \in \mathbb{N}^*$, $\underline{\mathbf{w}}_0^h + \zeta_1\underline{\mathbf{w}}_1^h, \ldots, \zeta_L\underline{\mathbf{w}}_L^h$ is also a solution of $\mathcal{E}^h$.

(*iii*) Any solution over $\mathbb{N}^*$ of the system of positive integer equations $\mathcal{E}^h$ is a solution (discarding slack variables) of $\mathcal{Q}^h$, and hence of the system of positive integer clauses $\mathcal{C}^*$.

(*iv*) The system $\mathcal{E}^h$ has singly exponentially many equations and doubly exponentially many variables, with all coefficients singly exponentially bounded. The $\mathbb{N}^*$-vectors $\underline{\mathbf{w}}_\ell^h$ ($0 \leq \ell \leq L$) are doubly exponentially finitely bounded, but with footprint of singly exponential cardinality. For $1 \leq \ell \leq L$ the $\underline{\mathbf{w}}_\ell^h$ are all finite. Moreover, if $\mathfrak{A}$ is finite, the $\underline{\mathbf{w}}_0^h$ are also all finite.

To understand the significance of the foregoing construction, fix some ordinary cluster C^h ($G < h \leq H$), and consider its c-spectrum. Evidently, $\mathrm{cs}^{\mathfrak{A}}[C^h] = \sum\{\mathrm{cs}^{\mathfrak{A}}[B] \mid B \subseteq C^h\}$. Hence, from Property (*i*) above, we have

$$\mathrm{cs}^{\mathfrak{A}}[C^h] = z_0^h\underline{\mathbf{u}}_0^h + z_1^h\underline{\mathbf{u}}_1^h + \cdots + z_L^h\underline{\mathbf{u}}_L^h, \tag{12}$$

where $z_\ell^h \in \mathbb{N}^*$ ($0 \leq \ell \leq L$). If $\mathfrak{A}$ is finite, the sum $\mathrm{cs}^{\mathfrak{A}}[C^h] = \sum\{\mathrm{cs}^{\mathfrak{A}}[B] \mid B \subseteq C^h\}$ has only finitely many terms, all of which are $\mathbb{N}$-vectors. Hence, $z_\ell^h \in \mathbb{N}$ for all h and ℓ in the given ranges.

Pictorially, we may imagine each equivalence class $B \subseteq C^h$ to be composed of various groups of elements, or 'constellations': a single 'core constellation' having c-spectrum $\underline{\mathbf{u}}_0^h$, and, for each ℓ ($1 \leq \ell \leq L$), some number (possibly zero) of 'peripheral constellations' each having c-spectrum $\underline{\mathbf{u}}_\ell^h$, as depicted in Fig. 3. The number z_0^h is simply the number of equivalence classes in $\mathcal{C}^h$, while the numbers z_ℓ^h ($1 \leq \ell \leq L$) are simply the totals obtained by summing the coefficients ζ_ℓ in (11) corresponding to all the B included in C^h. The key to our approach is that—subject to a caveat to be discussed in Sec. 4.5—we do not particularly mind how the various peripheral constellations are distributed between the equivalence classes in C^h: all that matters is the total number of constellations of each type, as given by the parameters z_ℓ^h ($G < h \leq H$, $0 \leq \ell \leq L$). And, while we have no *a priori* bound on the number of ordinary equivalence classes, we do have such a bound on L and H.

The collections of systems of Diophantine equations $\mathcal{E}^h$ and $\mathbb{N}^*$-vectors (or $\mathbb{N}$-vectors) $\underline{\mathbf{w}}_\ell^h$ will form part of the certificate for φ, which we assemble in Sec. 4.6.

4.5 Sectors and terminators

Having written Equations (12), we are tantalizingly close to constructing a certificate guaranteeing the (finite) satisfiability of φ. Observe that the bounds derived above mean that the collection of possible vectors $\mathbf{u}^h_\ell$ is computable from φ alone—while the extended natural numbers z^h_ℓ depend on the model $\mathfrak{A}$ with which we started. Recalling the discussion at the end of Sec 4.4, we can regard each $\mathbf{u}^h_\ell$ as arising from a clump of elements: if $\ell = 0$, these elements form the 'core' of some equivalence-class; if $1 \leq \ell \leq L$, they form a 'peripheral constellation', as depicted in Fig. 3. By contrast, the extended natural numbers z^h_ℓ just tell us how many of these various cores and peripheral constellations we are dealing with. The values z^h_ℓ satisfy some obvious constraints. Most notably, the total number of rays of any invertible cosmic type emitted by elements of the model must equal the total number of rays of the inverse type emitted by elements of the model, whence, for all j $(1 \leq j \leq J)$, $(\sum_{h=1}^{H}\sum_{\ell=0}^{L} z^h_\ell \mathbf{u}^h_\ell)[j] = (\sum_{h=1}^{H}\sum_{\ell=0}^{L} z^h_\ell \mathbf{u}^h_\ell)[J+j]$. Bearing in mind the results of Sec. 2.2, it is natural to wonder whether the sought-after certificate guaranteeing the (finite) satisfiability of φ is not simply provided by the $\mathbb{N}^*$-vectors $\mathbf{u}^h_\ell$. As long as there exist $z^h_\ell \in \mathbb{N}^*$ satisfying the constraints just mentioned, then, for each h, we can take z^h_ℓ of clumps of elements whose profiles are given by the corresponding vectors $\mathbf{w}^h_\ell$ (with ℓ ranging from 0 to L), and assemble them into a collection of equivalence classes to form the ordinary clusters C^h as depicted in Fig. 3. The fact that the $\mathbf{w}^h_\ell$ form a Hilbert basis for the system of positive integer equations $\mathcal{E}^h$ means (anticipating the results of Sec. 5) that *any* way of distributing peripheral constellations over cores within each cluster C^h will produce collections of elements that can be assembled into equivalence classes compatible with φ. And the fact that the number of cosmic rays of opposite invertible types balances will mean that we can join up the cosmic rays to complete the construction of the model. (The non-invertible cosmic rays, as we will see, can always be absorbed by the special clusters.)

This indeed is the basic strategy we follow in this paper. However, there is a fly in the ointment. Let us remind ourselves of the notation

$$j^* = \begin{cases} j+J & \text{if } 1 \leq j \leq J \\ j-J & \text{if } J < j \leq 2J. \end{cases}$$

Consider any j in the range $1 \leq j \leq 2J$, so that ρ_{4J+j} is an invertible cosmic ray-type, and ρ_{4J+j^*}, its inverse. By definition, a cosmic ray emitted by an element of some equivalence class B must be absorbed by an element of $A \setminus B$. Therefore, for every h $(1 \leq h \leq H)$, and every s $(0 \leq s < \mathfrak{c}(h))$ the elements of B^h_s cannot possibly emit more rays of type ρ_{4J+j} than the rest of $\mathfrak{A}$ emits rays of type ρ_{4J+j^*}. In symbols:

$$(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \leq \sum \left\{ (\mathrm{cs}^{\mathfrak{A}}[B^{h'}_{s'}])[j^*] \,\middle|\, 1 \leq h' \leq H,\ 0 \leq s' < \mathfrak{c}(h'), B^h_s \neq B^{h'}_{s'}) \right\}. \tag{13}$$

The problem is that the conditions on the $\mathbb{N}^*$-vectors $\mathbf{u}^h_\ell$ outlined above do not incorporate this observation: our proposed certificate is thus in danger of allowing us to construct equivalence classes in such a way that one of those equivalence classes emits more invertible cosmic rays of a particular type than all the others can possibly absorb. How can we ensure—in a succinct way—that this cannot happen? Not by writing (13) directly, because this requires information on the cosmic spectrum of the individual equivalence classes, and we do not know how many of those there are. In this section, we solve this problem by organizing the equivalence classes B^h_s in each cluster C^h into an alternating sequence of groups, which we refer to as sectors and terminators. Essential to the argument at this point is Lemma 2.9, which converts a large collection of inequalities—*viz* (13)—into a disjunction of triples of inequalities.

Let us write $\prec$ for the lexicographic ordering on ordered pairs of extended natural numbers: $(h, s) \prec (h', s')$ if either $h < h'$, or both $h = h'$ and $s < s'$. We transfer this ordering to equivalence classes B^h_s by writing $B^h_s \prec B^{h'}_{s'}$ if $(h, s) \prec (h', s')$, i.e. giving them the left-to-right order shown in Fig. 2. We use the symbols $\preceq$, $\succ$ and $\succeq$ in the expected way.

To make the essential ideas a little more salient, we deal first with the case where $\mathfrak{A}$ is finite. Thus, all values $\mathfrak{c}(h)$ $(1 \leq h \leq H)$ are finite, whence the sum in (13) involves only finitely many terms. (Indeed, the individual summands are also finite). Fix some j $(1 \leq j \leq 2J)$, so that ρ_{4J+j} is an invertible cosmic ray-type, and ρ_{4J+j^*}, its inverse, and consider the inequality (13). Note that $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j]$ and $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]$ are simply natural numbers. Allowing h and s to vary $(1 \leq h \leq H,\ 0 \leq s < \mathfrak{c}(h))$, we can form the list of natural numbers

$\{(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j]\}_{h,s}$, with the index-pairs (h,s) ordered lexicographically; and similarly for $\{(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]\}_{h,s}$. Applying Lemma 2.9, we see that (13) corresponds to Statement (a) of that lemma, whence, by the equivalent Statement (b), there exists an index-pair $(\mathfrak{h}(j), \mathfrak{s}(j))$ with $1 \le \mathfrak{h}(j) \le H$ and $0 \le \mathfrak{s}(j) < \mathfrak{c}(\mathfrak{h}(j))$, satisfying

$$\sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \mid B^h_s \prec B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\} \;\le\; \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \succeq B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\} \tag{14}$$

$$\sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \mid B^h_s \succ B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\} \;\le\; \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \preceq B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\} \tag{15}$$

$$\mathrm{cs}^{\mathfrak{A}}[B^{\mathfrak{h}(j)}_{\mathfrak{s}(j)}])[j] \le \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \prec B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\} + \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \succ B^{\mathfrak{h}(j)}_{\mathfrak{s}(j))}\}. \tag{16}$$

We now deal with the case where $\mathfrak{A}$ is infinite. Fixing j $(1 \le j \le 2J)$ as before, we seek an index-pair $(\mathfrak{h}(j), \mathfrak{s}(j))$ satisfying (14)–(16). Consider the collection of extended integers $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]$, with h and s varying in the ranges $1 \le h \le H$ and $0 \le s < \mathfrak{c}(h)$. If $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*] = \aleph_0$ for any values of h and s, pick $\mathfrak{h}(j) = h$ and $\mathfrak{s}(j) = s$ to be such a pair. Then (14)–(15) are trivial, and (16) follows immediately from (13). If, on the other hand, for some value of h, $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]$ is positive for infinitely many values of s, then we partition C^h into two clusters, say C' and C'', as follows. Let t be the largest value such that for some previously considered j', we have $\mathfrak{h}(j') = h$ and $\mathfrak{s}(j') = t$. We take $B^h_1, \dots, B^h_t$ to be the first equivalence classes of C', and then distribute the remaining equivalence classes of C^h among C' and C'' in such a way that each contains infinitely many B^h_s (with s varying) for which $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]$ is positive. Replace the cluster C^h by these new clusters, C', and C'', in that order, and then re-number systematically. (Thus, C'' becomes C^{h+1}.) Setting $\mathfrak{h}(j) = h+1$ and $\mathfrak{s}(j) = 0$ then makes all of the inequalities (14)–(16) trivial, because all values on the right-hand sides become infinite. Note also that, provided the previously-defined values of $\mathfrak{h}$ have been adjusted to reflect the re-numbering, the inequalities for the previously considered values of j will not be disturbed by this splitting. Thus, we are left with the case where the values $(\mathrm{ss}^{\mathfrak{A}}[B^h_s])[j^*]$ are all finite, and indeed positive for only finitely many values of s. But then we may simply ignore the zero terms, and obtain $\mathfrak{h}(j)$ and $\mathfrak{s}(j)$ exactly as for the finite case.

Let this choice of $\mathfrak{h}(j)$ and $\mathfrak{s}(j)$ be made for all j $(1 \le j \le 2J)$. Thus, we have established the inequalities (14)–(16) for all j in this range. These $2J$ inequalities will play a key role in constructing the certificate for φ—once we have re-organized them slightly. To this end, fix h $(1 \le h \le H)$, and let

$$S^h = \begin{cases} \{0\} & \text{if } 1 \le h \le G \\ \{\mathfrak{s}(j) \mid 1 \le j \le 2J,\ \mathfrak{h}(j) = h\} & \text{otherwise.} \end{cases}$$

Thus, for the special clusters, C^h $(1 \le h \le G)$, S^h simply records the index of the unique equivalence class B^h_0 in that cluster; while, for the ordinary clusters, C^h $(G < h \le H)$, S^h is guaranteed to include those indices $\mathfrak{s}(j)$ (with j varying), for which $B^{\mathfrak{h}(j)}_{\mathfrak{s}(j)}$ is included in C^h. The special treatment of special clusters in this regard is merely for notational convenience, and plays no essential role in the proof. We remark that, for ordinary clusters, C^h, the index-set S^h may be empty.

Let $\mathfrak{b}(h) = |S^h|$, and enumerate S^h as a strictly increasing sequence of extended natural numbers $s_1 \le \cdots \le s_{\mathfrak{b}(h)}$; if $\mathfrak{b}(h) = 0$, this is simply the empty sequence. Evidently, $\mathfrak{b}(h) \le 2J$, and $s_{\mathfrak{b}(h)} < \mathfrak{c}(h)$. By definition, $\mathfrak{s}(j)$ must be one of the elements s_p in the enumeration $s_1, \dots, s_{\mathfrak{b}(h)}$ of $S^{\mathfrak{h}(j)}$; and we write $\mathfrak{p}(j) = p$ to identify the index of this element. The functions

$$\mathfrak{h} : [1, 2J] \to [1, H] \qquad \mathfrak{p} : [1, 2J] \to [1, 2J]$$

will form part of the certificate for φ. Notice that $\mathfrak{p}(j) \le \mathfrak{b}(\mathfrak{h}(j)) \le 2J$. Keeping h fixed, and recalling the enumeration $s_1, \dots, s_{\mathfrak{b}(h)}$ of S^h, we define, for all p $(1 \le p \le \mathfrak{b}(h))$,

$$\dot{B}^h_p = B^h_{s_p}. \tag{17}$$

In addition, writing $s_0 = -1$, and $s_{\mathfrak{b}(h)+1} = \mathfrak{c}(h)$, we define, for all p $(1 \le p \le \mathfrak{b}(h) + 1)$,

$$\hat{B}^h_p = \bigcup\{B^h_s \mid s_{p-1} < s < s_p\}. \tag{18}$$

We refer to the (possibly empty) sets of elements $\hat{B}^h_p$ as the *sectors* of C^h, and to the sets of elements $\dot{B}^h_p$ as the *terminators* of C^h. The resulting internal organization of clusters into sectors and terminators is illustrated in

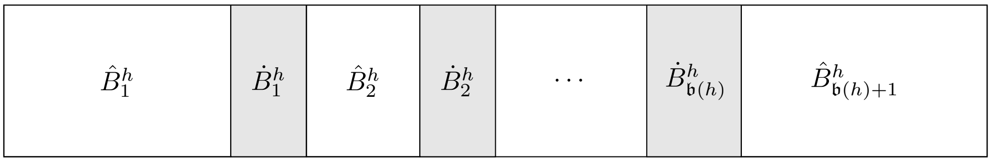

$\hat{B}^h_1$
$\dot{B}^h_1$
$\hat{B}^h_2$
$\dot{B}^h_2$
$\cdots$
$\dot{B}^h_{\mathfrak{b}(h)}$
$\hat{B}^h_{\mathfrak{b}(h)+1}$

Fig. 4 The division of C^h into $(\mathfrak{b}(h)+1)$ sectors $\hat{B}^h_p$ (unshaded) and $\mathfrak{b}(h)$ terminators (shaded).

Fig. 4. Thus, C^h is thus decomposed into $\mathfrak{b}(h)+1$ sectors and $\mathfrak{b}(h)$ terminators, where the value $\mathfrak{b}(h)=0$ is allowed. For special clusters, C^g, this decomposition necessarily takes the form: $\mathfrak{b}(g)=1$, $\hat{B}^g_1=\emptyset$, $\dot{B}^g_1=C^g$; $\hat{B}^g_2=\emptyset$. Observe that the sectors $\hat{B}^h_p$ may be unions of infinitely many equivalence classes if $\mathfrak{A}$ is infinite.

Notice that, since each terminator was chosen for a particular value j $(1 \leq j \leq 2J)$ or for a special equivalence class, the total number of terminators $\sum_{h=1}^{H} \mathfrak{b}(h)$ is bounded by $2J+G$. We shall see presently that the terminator $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$ functions as a witness guaranteeing that no equivalence class emits more rays of invertible cosmic type ρ_{4J+j} than the rest of $\mathfrak{A}$ emits rays of the inverse type ρ_{4J+j^*}.

Having constructed the various terminators and sectors, let us write arithmetic expressions for their c-spectra. To reduce notational clutter, define, for all h $(1 \leq h \leq H)$

$$\underline{\dot{u}}^h_p = \mathrm{cs}^{\mathfrak{A}}[\dot{B}^h_p] \qquad (1 \leq p \leq \mathfrak{b}(h)), \qquad\qquad \underline{\hat{u}}^h_p = \mathrm{cs}^{\mathfrak{A}}[\hat{B}^h_p] \qquad (1 \leq p \leq \mathfrak{b}(h)+1).$$

From (17) and (18), we have, for h and p in the appropriate ranges,

$$\underline{\dot{u}}^h_p = \mathrm{cs}^{\mathfrak{A}}[B^h_{s_p}] \qquad\qquad \underline{\hat{u}}^h_p = \sum\{\mathrm{cs}^{\mathfrak{A}}[B^h_s] \mid s_{p-1} < s < s_p\}. \tag{19}$$

(Note that the sequence $s_1, \ldots, s_{\mathfrak{b}(h)}$ used in these equations depends on h, and recall also that, if $\mathfrak{A}$ is infinite, s_p might equal $\aleph_0$ when $p=\mathfrak{b}(h)+1$.) Considering first the special clusters, fix g in the range $[1,G]$. By construction, in this case, $\dot{B}^g_1 = B^h_0$ and $\hat{B}^g_1 = \hat{B}^g_2 = \emptyset$. Continuing to write the profile of B^g_1 as $\underline{w}^g$ (as in the proof of Lemma 4.4), the following system of positive integer equations is satisfied:

$$\{(\underline{\dot{u}}^g_1 = \mathbf{U}\underline{w}^g),\ (\underline{\hat{u}}^g_1 = \underline{0}),\ (\underline{\hat{u}}^g_2 = \underline{0}) \mid 1 \leq g \leq G\}. \tag{$\mathcal{D}_1$}$$

Turning now to the ordinary clusters, fix h in the range $[G{+}1, H]$. Since the cosmic spectrum of every equivalence class included in C^h has the form (11), we see that, for all s $(0 \leq s < \mathfrak{c}(h))$, there exist $\zeta^h_{s,1}, \ldots, \zeta^h_{s,L} \in \mathbb{N}$, such that, writing $\zeta^h_{s,0}=1$,

$$\mathrm{cs}^{\mathfrak{A}}[B^h_s] = \zeta^h_{s,0}\underline{\mathbf{u}}^h_0 + \zeta^h_{s,1}\underline{\mathbf{u}}^h_1 + \cdots + \zeta^h_{s,L}\underline{\mathbf{u}}^h_L. \tag{20}$$

Therefore, from (19), there exist nonnegative extended integers,

$$\dot{z}^h_{p,\ell} = \zeta^h_{s_p,\ell} \qquad\qquad \hat{z}^h_{p,\ell} = \sum\{\zeta^h_{s,\ell} \mid s_{p-1} < s < s_p\},$$

satisfying the following system of extended integer equations:

$$\begin{aligned} &\left\{ \left(\underline{\dot{u}}^h_p = \textstyle\sum_{\ell=0}^{L} \dot{z}^h_{p,\ell}\underline{\mathbf{u}}^h_\ell\right) \middle| \, G < h \leq H,\ 1 \leq p \leq \mathfrak{b}(h) \right\} \\ &\left\{ \left(\underline{\hat{u}}^h_p = \textstyle\sum_{\ell=0}^{L} \hat{z}^h_{p,\ell}\underline{\mathbf{u}}^h_\ell\right) \middle| \, G < h \leq H,\ 1 \leq p \leq \mathfrak{b}(h)+1 \right\}. \end{aligned} \tag{$\mathcal{D}_2$}$$

Thus, $\mathcal{D}_1$–$\mathcal{D}_2$ express the c-spectrum of every sector and every terminator in terms of the parameters $\underline{w}^g$, $\dot{z}^h_{p,\ell}$ and $\hat{z}^h_{p,\ell}$ (whose values depend on $\mathfrak{A}$), via the vectors $\underline{\mathbf{u}}^h_\ell = \mathbf{U}\underline{\mathbf{w}}^h_\ell$ (which, with some limited guessing, can be computed from φ). Notice that the equations in $\mathcal{D}_2$ really are *extended* positive integer equations, since the vectors $\underline{\mathbf{u}}^h_0$ may contain infinite values. Of course, if $\mathfrak{A}$ is finite, no infinite values occur, and we have a system of positive integer equations.

A little reflection shows that the values of the parameters $\underline{w}^g$, $\dot{z}^h_{p,\ell}$ and $\hat{z}^h_{p,\ell}$ satisfy certain conditions. First, since each terminator comprises a single equivalence class, and $\zeta^h_{s,0}$ in (20) equals 1, the following simple system of linear Diophantine equations is satisfied:

$$\{\dot{z}^h_{p,0} = 1 \mid G < h \leq H,\ 1 \leq p \leq \mathfrak{b}(h)\}. \tag{$\mathcal{E}_1$}$$

Note that the total number of equations in $\mathcal{E}_1$ is at most $2J$, since this is the maximum number of the sum $\sum_{h=G+1}^{H} \mathfrak{b}(h)$. By contrast, $\hat{z}^h_{p,0}$ equals the number of equivalence classes included in the sector $\hat{B}^h_s$ and may be any extended natural number (finite if $\mathfrak{A}$ is). Note also that the condition $\hat{z}^h_{p,0} = 0$ means that the sector $\hat{B}^h_p$ is empty, in which case $\mathrm{cs}^{\mathfrak{A}}[\hat{B}^h_p] = \underline{0}$. Therefore, we may assume that the following system of linear Diophantine clauses is satisfied.

$$\left\{ (\hat{z}^h_{p,0} \geq 1) \vee \left(\sum_{\ell=1}^{L} \hat{z}^h_{p,\ell} = 0 \right) \,\middle|\, G < h \leq H,\ 1 \leq p \leq \mathfrak{b}(h)+1 \right\}. \tag{$\mathcal{E}_2$}$$

Having expressed the c-spectra of all the sectors and terminators—i.e. the quantities $\dot{\underline{u}}^h_p$ and $\hat{\underline{u}}^h_p$—in terms of the various parameters $\underline{w}^g$, $\dot{z}^h_{p,\ell}$ and $\hat{z}^h_{p,\ell}$, we can now do the same for the the c-spectrum of the whole domain. Writing $\underline{u} = \mathrm{cs}^{\mathfrak{A}}[A]$, it is immediate by inspection of Fig. 4 that

$$\underline{u} = \sum_{h=1}^{H} \left(\sum_{p=1}^{\mathfrak{b}(h)+1} \hat{\underline{u}}^h_p + \sum_{p=1}^{\mathfrak{b}(h)} \dot{\underline{u}}^h_p \right). \tag{$\mathcal{D}_3$}$$

The vector $\underline{u}$ alerts us to a further condition on the parameters $\underline{w}^g$, $\dot{z}^h_{p,\ell}$ and $\hat{z}^h_{p,\ell}$. Since each ray of invertible cosmic type ρ_{4J+j} $(1 \leq j \leq J)$ may be paired with a ray of (distinct) inverse type, $\rho^{-1}_{4J+j} = \rho_{5J+j}$, the following equations hold:

$$\{\underline{u}[j] = \underline{u}[j+J] \mid 1 \leq j \leq J\}. \tag{$\mathcal{E}_3$}$$

Now for the promised re-organization of the inequalities (14)—(16). We begin by defining the $\mathbb{N}^*$-vectors $\underline{u}^-$, $\underline{u}^+$, $\underline{u}^\circ$, of length $2J$, as follows. For all j $(1 \leq j \leq 2J)$, $\underline{u}^-[j]$ is the number of rays of invertible cosmic type ρ_{4J+j} emitted by those equivalence classes B^h_s occurring lexicographically *before* $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$; $\underline{u}^+[j]$ is the number of rays of the same type emitted by those equivalence classes B^h_s occurring lexicographically *after* $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$; and $\underline{u}^\circ[j]$ is the number of rays of the same type emitted by $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$ itself. In symbols:

$$\begin{aligned}
\underline{u}^-[j] &= \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \mid B^h_s \prec \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} \\
\underline{u}^+[j] &= \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \mid B^h_s \succ \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} \\
\underline{u}^\circ[j] &= (\mathrm{cs}^{\mathfrak{A}}[\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}])[j].
\end{aligned}$$

In a similar vein, we define the $\mathbb{N}^*$-vectors $\underline{v}^-$, $\underline{v}^+$ and $\underline{v}^\circ$, of length $2J$, as follows. For all j $(1 \leq j \leq 2J)$, $\underline{v}^-[j]$ is the number of rays of invertible cosmic type ρ_{4J+j^*} (note the j^* instead of j) emitted by those equivalence classes B^h_s occurring lexicographically *before* $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$; $\underline{v}^+[j]$ is the number of rays of the same type emitted by those equivalence classes B^h_s occurring lexicographically *after* $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$; and $\underline{v}^\circ[j]$ is the number of rays of the same type emitted by $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$ itself. In symbols:

$$\begin{aligned}
\underline{v}^-[j] &= \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \prec \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} \\
\underline{v}^+[j] &= \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \succ \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} \\
\underline{v}^\circ[j] &= (\mathrm{cs}^{\mathfrak{A}}[\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}])[j^*].
\end{aligned}$$

Now, taking $j = 1, \ldots, 2J$ in the inequalities (14)—(16), we obtain the system of positive integer inequalities:

$$\{\underline{u}^-[j] \leq \underline{v}^\circ[j] + \underline{v}^+[j] \mid 1 \leq j \leq 2J\} \tag{$\mathcal{E}_4$}$$

$$\{\underline{u}^+[j] \leq \underline{v}^-[j] + \underline{v}^\circ[j] \mid 1 \leq j \leq 2J\} \tag{$\mathcal{E}_5$}$$

$$\{\underline{u}^\circ[j] \leq \underline{v}^-[j] + \underline{v}^+[j] \mid 1 \leq j \leq 2J\}. \tag{$\mathcal{E}_6$}$$

To see that $\mathcal{E}_4$ is a consequence of (14), suppose $1 \leq j \leq 2J$. Then

$$\underline{u}^-[j] = \sum\{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j] \mid B^h_s \prec \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} \leq \{(\mathrm{cs}^{\mathfrak{A}}[B^h_s])[j^*] \mid B^h_s \succeq \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}\} = \underline{v}^\circ[j] + \underline{v}^+[j].$$

And similarly for $\mathcal{E}_5$–$\mathcal{E}_6$.

Thus, we see that $\mathcal{E}_4$–$\mathcal{E}_6$ are nothing but a re-organized form of the inequalities (14)–(16), which constitute a succinct guarantee that, for each j ($1 \leq j \leq J$), no sector or terminator—and hence certainly no equivalence class—accounts for more rays of any invertible cosmic type than the other elements of $\mathfrak{A}$ emit of the inverse type. The point of this re-organization is that we can express the vectors $\underline{u}^-$, $\underline{u}^+$, $\underline{u}^\circ$, $\underline{v}^-$, $\underline{v}^+$ and $\underline{v}^\circ$ in terms of the vectors $\underline{\dot{u}}^h_p$ and $\underline{\hat{u}}^h_p$. We see from Equations (17) and (18) that each terminator is a single equivalence class, and each sector is a sequence of equivalence classes consecutive under the ordering $\prec$. Evidently, then, an equivalence class B^h_s occurs strictly before $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$ in the ordering $\prec$ just in case it is included in a sector $\hat{B}^h_p$ such that $(h,p) \preceq (\mathfrak{h}(j), \mathfrak{p}(j))$ or an a terminator $\dot{B}^h_p$ such that $(h,p) \prec (\mathfrak{h}(j), \mathfrak{p}(j))$; likewise, B^h_s evidently occurs strictly after $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$ in the ordering $\prec$ just in case it is included in a sector $\hat{B}^h_p$ or an a terminator $\dot{B}^h_p$ such that $(h,p) \succ (\mathfrak{h}(j), \mathfrak{p}(j))$. Thus:

$$\begin{aligned}\Big\{\underline{u}^-[j] = {} & \sum\{\underline{\dot{u}}^h_p[j] \mid (h,p) \prec (\mathfrak{h}(j), \mathfrak{p}(j))\} \\ & + \sum\{\underline{\hat{u}}^h_p[j] \mid (h,p) \preceq (\mathfrak{h}(j), \mathfrak{p}(j))\} \mid 1 \leq j \leq 2J\Big\}\end{aligned} \tag{$\mathcal{D}_4$}$$

$$\begin{aligned}\Big\{\underline{u}^+[j] = {} & \sum\{\underline{\dot{u}}^h_p[j] \mid (H, \mathfrak{b}(H)) \succeq (h,p) \succ (\mathfrak{h}(j), \mathfrak{p}(j))\} \\ & + \sum\{\underline{\hat{u}}^h_p[j] \mid (H, \mathfrak{b}(H)+1) \succeq (h,p) \succ (\mathfrak{h}(j), \mathfrak{p}(j))\} \mid 1 \leq j \leq 2J\Big\}\end{aligned} \tag{$\mathcal{D}_5$}$$

$$\Big\{\underline{u}^\circ[j] = \underline{\dot{u}}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}[j] \;\mid 1 \leq j \leq 2J\Big\}. \tag{$\mathcal{D}_6$}$$

And by parallel reasoning:

$$\begin{aligned}\Big\{\underline{v}^-[j] = {} & \sum\{\underline{\dot{u}}^h_p[j^*] \mid (h,p) \prec (\mathfrak{h}(j), \mathfrak{p}(j))\} \\ & + \sum\{\underline{\hat{u}}^h_p[j^*] \mid (h,p) \preceq (\mathfrak{h}(j), \mathfrak{p}(j))\} \mid 1 \leq j \leq 2J\Big\}\end{aligned} \tag{$\mathcal{D}_7$}$$

$$\begin{aligned}\Big\{\underline{v}^+[j] = {} & \sum\{\underline{\dot{u}}^h_p[j^*] \mid (H, \mathfrak{b}(H)) \succeq (h,p) \succ (\mathfrak{h}(j), \mathfrak{p}(j))\} \\ & + \sum\{\underline{\hat{u}}^h_p[j^*] \mid (H, \mathfrak{b}(H)+1) \succeq (h,p) \succ (\mathfrak{h}(j), \mathfrak{p}(j))\} \mid 1 \leq j \leq 2J\Big\}\end{aligned} \tag{$\mathcal{D}_8$}$$

$$\Big\{\underline{v}^\circ[j] = \underline{\dot{u}}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}[j^*] \;\Big|\; 1 \leq j \leq 2J\Big\}. \tag{$\mathcal{D}_9$}$$

Observe that the sums in $\mathcal{D}_4$–$\mathcal{D}_9$, unlike those in (14)–(16), involve only finitely—indeed, at most doubly exponentially many—terms.

Let us gather together the above systems of extended integer clauses (or positive integer clauses, in case $\mathfrak{A}$ is finite), writing

$$\begin{aligned}\mathcal{D} &= \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_9 \\ \mathcal{E} &= \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_6.\end{aligned}$$

Let us further write $\underline{\dot{z}}$ for the tuple of values $\dot{z}^h_{p,\ell}$ in some arbitrary fixed order, and similarly for $\underline{\hat{z}}$. By regarding equations $\mathcal{D}$ as definitions of their left-hand sides, and performing the appropriate substitutions, we may regard

$\mathcal{E}$ as a system of (extended) positive integer clauses in the vector variables $\underline{w}^\dagger$, $\underline{\dot{z}}$ and $\underline{\hat{z}}$, for which we have constructed a solution over $\mathbb{N}^*$ (or over $\mathbb{N}$, if $\mathfrak{A}$ is finite). Crucially, the coefficients in these equations are all finitely bounded as a function of $\|\varphi\|$, while the constructed solution depends on the model $\mathfrak{A}$. Note also that $|\mathcal{E}_1|$ and $|\mathcal{E}_3|, \ldots, |\mathcal{E}_6|$are singly exponentially bounded, while $|\mathcal{E}_2|$ is doubly exponentially bounded, a matter to which we shall return in Sec. 6.

4.6 Certificates

Let us summarize the argument so far. From the formula φ and its expanded signature Σ, we defined the constants I and J, representing, respectively, the number of 1-types, and (approximately) one eighth of the number of polarized ray-types. Supposing φ to have a countable, 2-polarized, differentiated model $\mathfrak{A}$, interpreting Σ, we listed the star-types $\sigma_1, \ldots, \sigma_K$ realized in $\mathfrak{A}$, noting that these must be 2-polarized and compatible with φ. We identified the set $\mathcal{I} \subseteq \{1, \ldots, I\}$ of indices of 1-types uniquely realized in $\mathfrak{A}$, enumerated the special equivalence classes in $\mathfrak{A}$ as $B^1, \ldots, B^G$, and defined the sets $\mathcal{G}_i$ $(1 \leq i \leq I)$ of indices of the special equivalence classes in which π_i is realized. We observed in Sec. 4.2 that all statements in $\mathcal{B}$ hold. Denoting the profile of the special equivalence class B^g by $\underline{w}^g$, and writing $\underline{w}^\dagger = \underline{w}^1 \cdots \underline{w}^G$, we further observed in Sec. 4.3 that $\underline{w}^\dagger$ satisfies the system of positive integer clauses $\mathcal{C}^\dagger(\underline{w})$. We re-named the special equivalence classes as special clusters C^g $(1 \leq g \leq G)$; and we organized the ordinary (i.e., non-special) equivalence classes of $\mathfrak{A}$ into ordinary clusters C^h $(G < h \leq H)$. Each ordinary cluster C^h was associated with a system $\mathcal{Q}^h(\underline{w})$ of positive integer equations and inequalities that propositionally entail the clauses $\mathcal{C}^*(\underline{w})$. By adding slack variables to $\underline{w}$, we transformed $\mathcal{Q}^h$ into a system of positive integer equations $\mathcal{E}^h$, and thence obtained a sequence of $\mathbb{N}^*$-vectors $\underline{\mathbf{w}}_0^h$ and $\mathbb{N}$-vectors $\underline{\mathbf{w}}_\ell^h$ $(1 \leq \ell \leq L)$ such that $\underline{\mathbf{w}}_0^h$ and $\underline{\mathbf{w}}_0^h + \underline{\mathbf{w}}_\ell^h$ is a solution of $\mathcal{E}^h$ for all ℓ $(1 \leq \ell \leq L)$. The coefficients of $\mathcal{E}^h$ were all singly exponentially absolutely bounded, and the number L was doubly exponentially bounded; in addition, the vectors $\underline{\mathbf{w}}_\ell^h$ were all doubly exponentially finitely bounded, but had footprints of singly exponentially bounded cardinality. Finally, we computed the doubly exponentially finitely bounded vectors $\underline{\mathbf{u}}_\ell^h = \mathbf{U}\underline{\mathbf{w}}_\ell^h$. The argument of Sec. 4.4 showed that the $\underline{\mathbf{w}}_\ell^h$ $(0 \leq \ell \leq L)$ could be chosen in such a way that the c-spectrum of any equivalence class in the cluster C^h is a linear combination of the $\underline{\mathbf{u}}_\ell^h$, where the coefficients are extended natural numbers. In the case where $\mathfrak{A}$ is finite, all values obtained were finite.

We then decomposed each cluster C^h $(1 \leq h \leq H)$ into a sequence of sectors $\hat{B}_p^h$ $(1 \leq p \leq \mathfrak{b}(h) + 1)$ and terminators $\dot{B}_p^h$ $(1 \leq p \leq \mathfrak{b}(h))$, where $\mathfrak{b} : [1, H] \to [0, 2J]$ is a function. For $h > G$, we chose extended integers $\hat{z}_{p,0}^h, \ldots, \hat{z}_{p,L}^h$ such that the c-spectrum of the sector $\hat{B}_p^h$ was given by $\sum_{\ell=0}^{L} \hat{z}_{p,\ell}^h \underline{\mathbf{u}}_\ell^h$; and we chose integers $\dot{z}_{p,0}^h, \ldots, \dot{z}_{p,L}^h$, such that the c-spectrum of the terminator $\dot{B}_p^h$ was given by $\sum_{\ell=0}^{L} \dot{z}_{p,\ell}^h \underline{\mathbf{u}}_\ell^h$. (Actually, $\dot{z}_{p,0}^h = 1$.) Finally, in Sec. 4.5, we defined functions $\mathfrak{h} : [1, 2J] \to [1, H]$, $\mathfrak{p} : [1, 2J] \to [1, 2J]$ with $\mathfrak{p}(j) \leq \mathfrak{b}(\mathfrak{h}(j))$ specifying, for each j in these functions' domain, a terminator $\dot{B}_{\mathfrak{p}(j)}^{\mathfrak{h}(j)}$ witnessing the fact that no sector or terminator—and hence, no equivalence class—emits more rays of the invertible cosmic type ρ_{4J+j} than the rest of the model can absorb. Writing $\underline{\dot{z}}$ for the sequence of variables $\dot{z}_{p,\ell}^h$ $(G < h \leq H, 1 \leq p \leq \mathfrak{b}(h), 1 \leq \ell \leq L)$ in some order, and similarly for $\underline{\hat{z}}$, we showed that, under the definitions $\mathcal{D}$, the system of extended integer clauses $\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$ has a solution in $\mathbb{N}^*$. Applying Corollary 2.5, we see that it has a solution which is triply exponentially finitely bounded. We observed that, if $\mathfrak{A}$ is finite, then $\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$ is a system of Diophantine clauses, and has a solution over $\mathbb{N}$. Applying Corollary 2.3, we see that it has a solution which is triply exponentially bounded.

Let φ, Σ, $\pi_1, \ldots, \pi_I$ and $\rho_1, \ldots, \rho_{8J}$, be as described in Sec. 3, then, and let the sets of statements, constraints and definitions $\mathcal{B}$, $\mathcal{C}^*$, $\mathcal{C}^\dagger$, $\mathcal{D}$ and $\mathcal{E}$ be as described in Sec. 4. A *certificate* (for φ) is a tuple

$$\mathfrak{C} = \langle G, H, I, J, K, L, \{\sigma_k\}, \mathcal{I}, \{\mathcal{G}_i\}, \mathfrak{b}, \mathfrak{h}, \mathfrak{p}, \{\mathcal{E}^h\}, \{\underline{\mathbf{w}}_\ell^h\}, \underline{a}^\dagger, \underline{\dot{b}}, \underline{\hat{b}} \rangle \tag{21}$$

where:

$\mathfrak{C}$1: G, H, I, J, K, L are positive integers with $G \leq H$;

$\mathfrak{C}$2: for all k $(1 \leq k \leq K)$, σ_k is a 2-polarized star-type over Σ compatible with φ;

$\mathfrak{C}$3: $\mathcal{I}$ is a set of integers in the range $[1, I]$, and for all i $(1 \leq i \leq I)$, $\mathcal{G}_i$ is a set of integers in the range $[1, G]$, satisfying all the statements in $\mathcal{B}$;

$\mathfrak{C}4$: $\mathfrak{b} : [1, H] \to [0, 2J]$, $\mathfrak{h} : [1, 2J] \to [1, H]$ and $\mathfrak{p} : [1, 2J] \to [1, 2J]$ are functions such that $\mathfrak{b}(g) = 1$ for all g $(1 \leq g \leq G)$, $\mathfrak{p}(j) \leq \mathfrak{b}(\mathfrak{h}(j))$ for each j $(1 \leq j \leq 2J)$, and $\sum_{h=1}^{H} \mathfrak{b}(h) \leq 2J + G$;

$\mathfrak{C}5$: for all h $(G < h \leq H)$, $\mathcal{E}^h$ is a system of positive integer equations which (converting equations with slack variables to inequalities) propositionally entails the constraints $\mathcal{C}^*(\underline{w})$;

$\mathfrak{C}6$: for all h $(G < h \leq H)$, $\underline{\mathbf{w}}_0^h$ is a solution of $\mathcal{E}^h$ over $\mathbb{N}^*$, and, for all ℓ $(1 \leq \ell \leq L)$, $\underline{\mathbf{w}}_\ell^h$ is an $\mathbb{N}$-vector such that $\underline{\mathbf{w}}_0^h + \underline{\mathbf{w}}_\ell^h$ is a solution of $\mathcal{E}^h$;

$\mathfrak{C}7$: $\underline{a}^\dagger$, $\dot{\underline{b}}$, $\hat{\underline{b}}$ are $\mathbb{N}^*$-vectors such that $\mathcal{C}^\dagger(\underline{a}^\dagger)$ and $\mathcal{E}(\underline{a}^\dagger, \dot{\underline{b}}, \hat{\underline{b}})$.

We say that $\mathfrak{C}$ is *finite* if $\underline{a}^\dagger$, $\dot{\underline{b}}$, $\hat{\underline{b}}$ and the $\{\underline{\mathbf{w}}_0^h\}$ are all $\mathbb{N}$-vectors—equivalently, if the symbol $\aleph_0$ does not appear anywhere in $\mathfrak{C}$. (Of course, the word "finite" is being used metonymically here: all certificates, considered as data-structures, are finite objects.)

By Lemma 3.1, if φ has a (finite) model, then φ has a countable (finite) 2-polarized, differentiated, countable model, $\mathfrak{A}$, interpreting Σ. In this section, we used the model $\mathfrak{A}$ to guide the construction of a certificate $\mathfrak{C}$ for φ, with $\mathfrak{C}$ finite if $\mathfrak{A}$ is. Thus, we have shown:

Lemma 4.5 *Suppose φ is a $\mathcal{C}^2 1E$-formula in normal form. If φ is* (*finitely*) *satisfiable, then φ has a* (*finite*) *certificate of the form* (21) *in which:* (*i*) *G, I and J are all exponentially bounded as a function of $\|\varphi\|$;* (*ii*) *H, K and L are all doubly exponentially bounded;* (*iii*) *the coefficients of $\mathcal{E}^h$ are all singly exponentially absolutely bounded;* (*iv*) *the $\mathbb{N}^*$-vectors* ($\mathbb{N}$*-vectors*) *$\underline{\mathbf{w}}_\ell^h$, are all doubly exponentially finitely bounded; and* (*v*) *the $\mathbb{N}^*$-vectors* ($\mathbb{N}$*-vectors*) *$\underline{a}^\dagger$, $\dot{\underline{b}}$, $\hat{\underline{b}}$ are all triply exponentially finitely bounded.*

If $\mathfrak{C}$ is a certificate, take its *size*, denoted $\|\mathfrak{C}\|$, to be the number of bits required to write it (under some natural encoding). Thus, Lemma 4.5 states that, if φ has a finite model, then it has a certificate of doubly exponential size. In the next section, we prove the converse: if φ has a (finite) certificate, then it is (finitely) satisfiable. In section 6, we improve the size bound.

5 From certificates to models

Let a certificate $\mathfrak{C}$ of the form (21) be given. We keep $\mathfrak{C}$ fixed throught this section, using it to construct a model $\mathfrak{A}$ of φ. If $\mathfrak{C}$ is finite, then $\mathfrak{A}$ will be finite. Recall in particular that $\mathfrak{C}$ features a list of star-types $\sigma_1, \ldots, \sigma_K$.

5.1 Galaxies and the cosmos

A *set of stars* is a set A together with a mapping $\text{st} : A \to \{\sigma_1, \ldots, \sigma_K\}$. We call any element $a \in A$ a *star*, and we call $\text{st}(a)$ the *intrinsic star-type* of a. We write $\text{tp}(a)$ for $\text{tp}(\text{st}(a))$, and call this 1-type the *intrinsic* 1*-type* of a. If $\text{st}(a) = \langle \pi, (v_1, \ldots, v_{8J}) \rangle$, then we write $\text{st}_\star(a)$ for $\langle \pi, (v_1, \ldots, v_{4J}) \rangle$, and we call $\text{st}_\star(a)$ the *intrinsic galactic star-type* of a. We allow A to be infinite; however, we shall always (silently) assume sets of stars to be countable. Taking the illustration of the star-type $\sigma = \langle \pi, (v_1, \ldots, v_{8J}) \rangle$ in Fig. 1b as our cue, we informally speak of any star a with intrinsic type σ as 'emitting' a collection of 'rays' of the various types. Specifically: for each j $(1 \leq j \leq 8J)$, we shall say that a emits $\sigma[j]$ rays of type ρ_j. We employ this way of speaking throughout this section.

If $A' \subseteq A$, we define the *intrinsic profile* of A', denoted $\text{pr}(A')$, to be the $\mathbb{N}^*$-vector $(w_1, \ldots, w_K)$, where $w_k = |\{a \in A' \mid \text{st}(a) = \sigma_k\}|$. Likewise, we define the *intrinsic c-spectrum* of A', denoted $\text{cs}(A')$, to be the $\mathbb{N}^*$-vector $(u_1, \ldots, u_{2J})$, where, for all j $(1 \leq j \leq 2J)$, u_j is the total number of rays of (invertible, cosmic) type ρ_{4J+j} emitted by the stars in A'. Recalling the matrix $\mathbf{U}$ from Sec. 4.1, we see that $\text{cs}(A') = \mathbf{U} \cdot \text{pr}(A')$. Do not confuse the notation $\text{st}(a)$, where a is a star, with the notation $\text{st}^{\mathfrak{A}}[a]$, where a is an element of the domain of a polarized structure $\mathfrak{A}$; similarly for $\text{st}_\star(a)$, $\text{tp}(a)$, $\text{pr}(A')$ and $\text{cs}(A')$. So far in this section, we have not built any structures. For brevity in the sequel, if A' is a set of stars, we speak of the *rays emitted by* A' when we mean the *rays emitted by the stars contained in* A'. Thus, $\text{cs}(A')$ tells us how many cosmic rays of each type are emitted by A'.

Of course, if A is a set of stars, there is nothing to prevent us from using A as the domain of a polarized structure $\mathfrak{A}$ interpreting Σ. In that case, any $a \in A$ also has a star-type $\sigma^{\mathfrak{A}}[a]$, and indeed a galactic a star-type

$\sigma^{\mathfrak{A}}_{\star}[a]$, both defined as in Sec. 3.4. To avoid confusion, we typically refer to the star-type $\sigma^{\mathfrak{A}}[a]$ as the *extrinsic star-type of* a, and similarly for galactic star-types, 1-types, profiles, etc. However, the case we are most interested in is the one where the intrinsic notions defined above coincide with their extrinsic counterparts. If A is a set of stars, then we call a polarized structure $\mathfrak{A}$ interpreting Σ over A a *cosmos* if, for all $a \in A$, $\mathrm{st}(a) = \mathrm{st}^{\mathfrak{A}}[a]$. In this section, we shall construct a (finite) cosmos from a (finite) certificate.

We begin on a small scale. Let B be a set of stars. Say that a *galaxy* is a polarized structure $\mathfrak{B}$ interpreting Σ over B, satisfying the following properties:

(*i*) $E^{\mathfrak{B}}$ is the total relation $B \times B$;

(*ii*) for all $b \in B$, $\mathrm{st}(b)$ is compatible with φ, and $\mathrm{st}^{\mathfrak{B}}_{\star}[b] = \mathrm{st}_{\star}(b)$;

(*iii*) every (dark, galactic) 2-type realized in $\mathfrak{B}$ is compatible with φ.

By property (*i*), the qualifier ‘galactic’ can be removed from property (*iii*) without change of meaning, since all 2-types realized in $\mathfrak{B}$ are necessarily galactic. By property (*ii*), the qualifier ‘dark’ can removed from property (*iii*) without change of meaning, since all ray-types in $\mathfrak{B}$ must be compatible with φ. It is obvious that, if $\mathfrak{A}$ is a cosmos, then the equivalence classes of $\mathfrak{A}$ are galaxies. Note that, formally, we allow the set of stars B—and hence the galaxy $\mathfrak{B}$—to be empty.

Recall the system of positive integer clauses $\mathcal{C}^0(\underline{w})$, as well as the various constants $\underline{\mathbf{p}}_i$, $\underline{\mathbf{t}}_j$, $\underline{\mathbf{o}}^c_{i,i'}$, $\underline{\mathbf{o}}^*_{i,i'}$, $\underline{\mathbf{q}}^d_{i,i'}$, $\underline{\mathbf{q}}^*_{i,i'}$ from Sec. 4.3. The next lemma is, in effect, a converse of Lemma 4.2.

Lemma 5.1 *Suppose the* $\mathbb{N}^*$*-vector* $\underline{w} = (w_1, \ldots, w_K)$ *satisfies* $\mathcal{C}^0(\underline{w})$. *Then there exists a galaxy* $\mathfrak{B}$ *such that* $\mathrm{pr}(B) = \underline{w}$.

Proof. Let B be a set of stars with intrinsic profile $\underline{w}$. For each $a \in B$, and each galactic ray—say, of type ρ—emitted by a, we show how to select some $b \in B \setminus \{a\}$ for which $\mathrm{tp}^{\mathfrak{B}}[a,b]$ has not yet been defined, so that we can set $\mathrm{tp}^{\mathfrak{B}}[a,b] = \rho$. We say in this case that b *absorbs* the ray in question. We then complete the construction of $\mathfrak{B}$ by setting any 2-types not defined by this process to be dark galactic 2-types compatible with φ. Clearly, following this construction, $E^{\mathfrak{B}} = B \times B$. Moreover, since every galactic ray emitted by any star in $a \in B$ is found an absorption site in B, and all other 2-types are dark, $\mathrm{st}^{\mathfrak{B}}_{\star}[a] = \mathrm{st}_{\star}(a)$. The construction proceeds in three stages.

Stage 1: Consider first the *invertible* galactic ray-types $\rho_1, \ldots, \rho_{2J}$, and fix j ($1 \leq j \leq J$) for the moment. Recall that the chosen enumeration of the ray-types ensures that $\rho_j^{-1} = \rho_{J+j}$. The total number of rays of type ρ_j emitted by the elements of B is $\underline{\mathbf{t}}_j \cdot \underline{w}$; and the total number of rays of type ρ_j^{-1} emitted by the elements of B is $\underline{\mathbf{t}}_{J+j} \cdot \underline{w}$. By $\mathcal{C}^0_1$, these are equal, so let the two sets of rays be put in 1–1 correspondence. Do this for all j ($1 \leq j \leq J$). Now take and $a, b \in B$ such that a and b emit invertible galactic rays that have been paired in this process. Suppose the type of the ray emitted by a is ρ. Let $\pi = \mathrm{tp}(a)$ and $\pi' = \mathrm{tp}(b)$. Then $\pi = \mathrm{tp}_1(\rho)$ and $\pi' = \mathrm{tp}_2(\rho)$. Recalling that the certificate $\mathfrak{C}$ given in (21) is subject to the condition $\mathfrak{C}2$, the star-types $\sigma_1, \ldots, \sigma_K$ are 2-polarized. Since, therefore, all ray-types considered are polarized, $\pi \neq \pi'$, whence $a \neq b$. And since all the intrinsic star-types of a and b are 2-polarized, there can be no other invertible ray emitted by a with absorption-type π' and no other invertible ray emitted by b with absorption-type π; hence a cannot be chosen to absorb any other invertible galactic ray emitted by b, and b cannot be chosen to absorb any other invertible galactic ray emitted by a. Therefore, we may set $\mathrm{tp}^{\mathfrak{B}}[a,b] = \rho$ for any such pair a, b, without danger of clashes. At the end of this stage, absorption sites have been found for all the invertible galactic rays emitted by the stars in B.

Stage 2: Fix i and i' ($1 \leq i, i' \leq I$), and write $\pi = \pi_i$ and $\pi' = \pi_{i'}$. We proceed to find absorption sites for all non-invertible galactic rays of absorption-type π' emitted by stars having intrinsic 1-type π, and, simultaneously, absorption sites for all non-invertible galactic rays of absorption-type π emitted by stars having intrinsic 1-type π'. The number of stars of B having intrinsic 1-type π is $\underline{\mathbf{p}}_i \cdot \underline{w}$, and the number of stars of B having intrinsic 1-type π' is $\underline{\mathbf{p}}_{i'} \cdot \underline{w}$. By $\mathcal{C}^0_2$, each of these numbers is either at most 1 or at least Z.

If $\underline{\mathbf{p}}_i \cdot \underline{w} = 0$, then, by $\mathcal{C}^0_3$, putting $c = 1$, we have $\underline{\mathbf{o}}^1_{i',i} \cdot \underline{w} = 0$. That is to say, no star of B having intrinsic 1-type π' emits any ray with absorption-type π. Hence, for the pair of 1-types π and π', there is no work to do. An exactly similar argument applies if $\underline{\mathbf{p}}_{i'} \cdot \underline{w} = 0$. Henceforth, then, we may assume that these quantities are both positive.

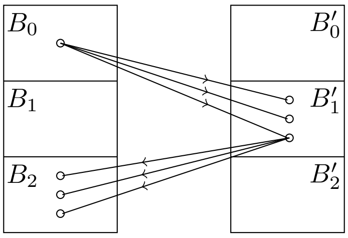


Fig. 5 Finding absorption sites for non-invertible galactic rays

Now suppose $\underline{\mathbf{p}}_i \cdot \underline{w} = 1$, and let a be the unique star of B having intrinsic 1-type π. By $\mathcal{C}^0_3$, putting $c = 2$, $\underline{\mathbf{o}}^2_{i',i} \cdot \underline{w} = 0$: that is to say, none of the stars of intrinsic 1-type π' in B emits more than one galactic ray (invertible or non-invertible) with absorption-type π. Suppose, then, b has intrinsic 1-type π' and emits exactly one non-invertible galactic ray with absorption-type π (and therefore no invertible galactic rays with absorption-type π). Let ρ be the type of this ray. We note first that that $a \neq b$; for otherwise, we have $i = i'$, and therefore $\underline{\mathbf{p}}_i \cdot \underline{w} = 1$, $\underline{\mathbf{o}}^1_{i,i'} \cdot \underline{w} > 0$ and $\underline{\mathbf{o}}^*_{i',i} \cdot \underline{w} = 0$, contradicting $\mathcal{C}^0_4$ (with $c = 1$). Since b emits no invertible galactic rays with absorption-type π, $\mathrm{tp}^{\mathfrak{B}}[b, a]$ is currently undefined, so we may set $\mathrm{tp}^{\mathfrak{B}}[b, a] = \rho$ without undoing any of the work of Stage 1. Clearly, we can carry out these assignments for all stars of B having intrinsic 1-type π' and emitting exactly one non-invertible galactic ray with absorption-type π. It remains to deal with any galactic rays emitted by a and having absorption-type π'. Let the number of such rays be c. Certainly, $c \leq Mm$, and if $c = 0$, there is nothing to do; so we may assume $1 \leq c \leq Mm$. Since a has 1-type π, we have $\underline{\mathbf{o}}^c_{i,i'} \cdot \underline{w} > 0$, whence by $\mathcal{C}^0_4$, $\underline{\mathbf{o}}^*_{i',i} \cdot \underline{w} \geq c$. That is, we can find at least c stars of B having intrinsic 1-type π' that do not emit any non-invertible galactic rays with absorption-type π—hence which do not emit any rays which were assigned to be absorbed by a in this stage. Of these c stars, up to one may have been chosen in Stage 1 to absorb an invertible galactic ray emitted by a: if so, it has already been dealt with in Stage 1; and therefore, for each non-invertible ray—say of type ρ—emitted by a and having absorption-type π', we can find a fresh star b of 1-type π' that does not send any galactic ray to a, and for which $\mathrm{tp}^{\mathfrak{B}}[a, b]$ was not defined in Stage 1. Hence we may set $\mathrm{tp}^{\mathfrak{B}}[a, b] = \rho$ without undoing any work of Stage 1 or any work previously done in this stage. Thus, we have again found absorption sites for all non-invertible galactic rays of absorption-type π' emitted by the unique star having intrinsic 1-type π, and absorption sites for all non-invertible galactic rays of absorption-type π emitted by stars having intrinsic 1-type π'. An exactly similar argument applies if $\underline{\mathbf{p}}_{i'} \cdot \underline{w} = 1$. Henceforth, then, we may assume that $\underline{\mathbf{p}}_i \cdot \underline{w}$ and $\underline{\mathbf{p}}_{i'} \cdot \underline{w}$ are both at least 2.

By $\mathcal{C}^0_2$, B contains at least $Z \geq 3Mm$ stars having intrinsic 1-type π, and at least $Z \geq 3Mm$ stars having intrinsic 1-type π'. Partition the former set into subsets B_0, B_1, B_2 of cardinality at least Mm; and partition the latter into subsets B'_0, B'_1, B'_2 of cardinality at least Mm. No star in B emits more than Mm rays in total. For each ℓ ($0 \leq \ell < 3$), and for for each non-invertible ray—say of type ρ—emitted by any star $a \in B_\ell$ and having absorption-type π', we may choose a fresh star $b \in B'_{\ell+1}$ (addition in subscripts modulo 3) not chosen to absorb any other galactic ray emitted by a, and set $\mathrm{tp}^{\mathfrak{B}}[a, b] = \rho$. Likewise, for each ℓ ($0 \leq \ell < 3$), and for for each non-invertible ray—say of type ρ—emitted by any star $b \in B'_\ell$ and having absorption-type π', we may choose a fresh star $a \in B_{\ell+1}$ not chosen to absorb any other galactic ray emitted by b, and set $\mathrm{tp}^{\mathfrak{B}}[b, a] = \rho$. The arrangement is depicted in Fig. 5; by inspection, none of these type assignments clashes with any other, even when $\pi = \pi'$. Once again, we have found absorption sites for all non-invertible galactic rays of absorption-type π' emitted by each star having intrinsic 1-type π, and absorption sites for all non-invertible galactic rays of absorption-type π emitted by each star having intrinsic 1-type π'.

Stage 3: We have now found absorption sites for all the galactic rays emitted by the stars in B, and it remains to set any remaining 2-types. Suppose, then, a and b are distinct stars such that $\mathrm{tp}^{\mathfrak{B}}[a, b]$ has not yet been assigned. Let $\mathrm{tp}(a) = \pi_i = \pi$ and $\mathrm{tp}(b) = \pi_{i'} = \pi'$. Suppose first that π and π' are not galactically coupled. This means that there exists a dark galactic 2-type τ compatible with φ, such that $\mathrm{tp}_1(\tau) = \pi$ and $\mathrm{tp}_2(\tau) = \pi'$. Then we set $\mathrm{tp}^{\mathfrak{B}}[a, b] = \tau$. We will complete the proof by showing that, if π and π' are galactically coupled, then $\mathrm{tp}^{\mathfrak{B}}[a, b]$ was in fact defined in Stages 1 or 2, so there is nothing to do in this case.

Suppose, then $\pi \overset{g}{\sim} \pi'$. By $\mathcal{C}^0_9$, either a is the unique star in B having intrinsic 1-type π, or b is the unique star of B having intrinsic 1-type π'. Without loss of generality, suppose the former. Obviously, if b emits a galactic ray with absorption-type π, then a will have been chosen to absorb that ray in Stage 1 or Stage 2 (it was the only candidate), and thus $\mathrm{tp}^{\mathfrak{B}}[a,b]$ is already defined. So assume b emits no such ray, and let $c' \leq Mm$ be the number of galactic rays (invertible or non-invertible) with absorption type π' emitted by a. Since a is the unique star of B having intrinsic 1-type π, we have $\underline{\mathbf{o}}^{c'+1}_{i,i'} \cdot \underline{w} = 0$, and hence, putting $c = c'+1$ in $\mathcal{C}^0_5$, $\underline{\mathbf{o}}^{*}_{i',i} \cdot \underline{w} \leq c'$. Let B' be the set of stars of B having intrinsic 1-type π' that do not send any non-invertible galactic ray to a. Thus, $|B'| \leq c'$. Yet the elements of B' are the only stars that can serve as absorption sites for the c' galactic rays emitted by a and having absorption-type π'. Thus, every element of $b \in B'$ was chosen as such an absorption site in Stages 1 and 2; and therefore $\mathrm{tp}^{\mathfrak{B}}[a,b]$ is already defined, as required. □

5.2 Constructing the cosmos: the stars

Recalling that the certificate $\mathfrak{C}$ given in (21) is subject to the condition $\mathfrak{C}7$, we see that the system of positive integer clauses $\mathcal{C}^\dagger(\underline{w}^\dagger)$ has a solution $\underline{a}^\dagger$. For ease of comparison with Sec. 4, we shall henceforth write $\underline{w}^\dagger = \underline{w}^1, \ldots, \underline{w}^G$ to denote this solution. For the same reason, we write $\underline{\dot{z}}$ and $\underline{\hat{z}}$ for the vectors $\underline{\dot{b}}$ and $\underline{\hat{b}}$, respectively. Furthermore, we take the various vectors $\underline{\dot{u}}^h_p$, $\underline{\hat{u}}^h_p$, $\underline{\dot{v}}^h_p$, $\underline{\hat{v}}^h_p$, $\underline{u}^-$, $\underline{u}^+$, $\underline{u}^\circ$, $\underline{v}^-$, $\underline{v}^+$ and $\underline{v}^\circ$ to be defined in terms of $\underline{w}^\dagger$, $\underline{\dot{z}}$ and $\underline{\hat{z}}$ by the equations in $\mathcal{D}$. Thus, for all g ($1 \leq g \leq G$), we have $\mathcal{C}^0(\underline{w}^g)$; and we have $\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$. Note that, if $\mathfrak{C}$ is finite, then all the vectors mentioned here are $\mathbb{N}$-vectors.

We begin by constructing the stars that will become special elements of our model. By Lemma 5.1, let $\mathfrak{B}^g$ be a galaxy with profile $\underline{w}^g$, for all g ($1 \leq g \leq G$). By $\mathcal{C}^g_4$, these galaxies are all non-empty. The galaxies $\mathfrak{B}^1, \ldots, \mathfrak{B}^G$ will form the special equivalence classes in our model of φ. As in Section 4.4, we impose on each special equivalence class B^g a degenerate decomposition into the empty sectors $\hat{B}^g_1 = \hat{B}^g_2 = \emptyset$ and the single terminator $\dot{B}^g_1 = B^g$. To maintain consistency with notation to be introduced shortly, for all g ($1 \leq g \leq G$), we write

$$\underline{\dot{w}}^g_1 = \underline{w}^g \qquad\qquad \underline{\hat{w}}^g_1 = \underline{\hat{w}}^g_2 = \underline{0}.$$

Thus, $\mathrm{pr}(\dot{B}^g_1) = \underline{\dot{w}}^g_1$, $\mathrm{pr}(\hat{B}^g_1) = \underline{\hat{w}}^g_1$ and $\mathrm{pr}(\hat{B}^g_2) = \underline{\hat{w}}^g_2$.

We turn now to the ordinary stars. Fix some h ($G < h \leq H$). Define

$$\begin{aligned} \underline{\dot{w}}^h_p &= \dot{z}^h_{p,0}\underline{\mathbf{w}}^h_0 + \dot{z}^h_{p,1}\underline{\mathbf{w}}^h_1 + \cdots + \dot{z}^h_{p,L}\underline{\mathbf{w}}^h_L && (1 \leq p \leq \mathfrak{b}(h)) \\ \underline{\hat{w}}^h_p &= \hat{z}^h_{p,0}\underline{\mathbf{w}}^h_0 + \hat{z}^h_{p,1}\underline{\mathbf{w}}^h_1 + \cdots + \hat{z}^h_{p,L}\underline{\mathbf{w}}^h_L && (1 \leq p \leq \mathfrak{b}(h)+1). \end{aligned}$$

(Remember that $\mathfrak{b}(h)$ may be zero.) If $1 \leq p \leq \mathfrak{b}(h)$, then, by $\mathcal{E}_1$, $\dot{z}^h_{p,0} = 1$, whence, by $\mathfrak{C}6$ and Lemma 2.7, $\underline{\dot{w}}^h_p$ is a solution of $\mathcal{E}^h$, and thus, by $\mathfrak{C}5$, satisfies the system of positive integer clauses $\mathcal{C}^*$. By Lemma 5.1, then, let $\dot{\mathfrak{B}}^h_p$ be a galaxy with intrinsic profile $\underline{\dot{w}}^h_p$ over a domain $\dot{B}^h_p$. Now consider $\underline{\hat{w}}^h_p$, for any p ($1 \leq p \leq \mathfrak{b}(h)+1$). We have two cases. If $\hat{z}^h_{p,0} = 0$, then, by $\mathcal{E}_2$, $\underline{\hat{w}}^h_p = \underline{0}$; in that case, we let $\hat{B}^h_p$ be the empty set of stars, with intrinsic profile $\underline{0}$. Otherwise, $\hat{z}^h_{p,0} > 0$, and we proceed as follows. By $\mathfrak{C}6$ and Lemma 2.7 again, both $\underline{\mathbf{w}}^h_0$ and also the vector $\underline{\tilde{w}} = \underline{\mathbf{w}}^h_0 + \dot{z}^h_{p,1}\underline{\mathbf{w}}^h_1 + \cdots + \dot{z}^h_{p,L}\underline{\mathbf{w}}^h_L$ are solutions of $\mathcal{E}^h$, and thus by $\mathfrak{C}5$ satisfy the system of positive integer clauses $\mathcal{C}^*$. Again, if $\mathfrak{C}$ is finite, then the $\underline{\mathbf{w}}^h_0$ are all $\mathbb{N}$-vectors. By Lemma 5.1, then, there exist galaxies $\mathfrak{B}$, with intrinsic profile $\underline{\mathbf{w}}^h_0$, and $\tilde{\mathfrak{B}}$, with intrinsic profile $\underline{\tilde{w}}$. Now take $\hat{z}^h_{p,0} - 1$ copies of $\mathfrak{B}$ and a single copy of $\tilde{\mathfrak{B}}$ (all disjoint), and let $\hat{B}^h_p$ be the union of their domains. (Here, if $\hat{z}^h_{p,0} = \aleph_0$, then we take $\hat{z}^h_{p,0} - 1 = \aleph_0$.) Evidently, $\mathrm{pr}(\hat{B}^h_p) = \underline{\hat{w}}^h_p$. The sets $\dot{B}^h_p$ and $\hat{B}^h_p$ will form, respectively, the sectors and terminators of the cluster C^h; we may imagine them to be arranged as in Fig. 4. Let this construction be carried out for all values of h ($G < h \leq H$). Putting the special and ordinary clusters together, we see that, for all h ($1 \leq h \leq H$):

$$\mathrm{pr}(\dot{B}^h_p) = \underline{\dot{w}}^h_p \qquad (1 \leq p \leq \mathfrak{b}(h)), \qquad\qquad \mathrm{pr}(\hat{B}^h_p) = \underline{\hat{w}}^h_p \qquad (1 \leq p \leq \mathfrak{b}(h)+1). \tag{22}$$

Let

$$A = (B^1 \cup \cdots B^G) \cup \bigcup_{h=G+1}^{H} \left(\bigcup_{p=1}^{\mathfrak{b}(h)} \dot{B}^h_p \cup \bigcup_{p=1}^{\mathfrak{b}(h)+1} \hat{B}^h_p \right) = \bigcup_{h=1}^{H} \left(\bigcup_{p=1}^{\mathfrak{b}(h)} \dot{B}^h_p \cup \bigcup_{p=1}^{\mathfrak{b}(h)+1} \hat{B}^h_p \right).$$

The set of stars A will form the domain of the model being constructed. Evidently, if $\mathfrak{C}$ is finite, then A is finite. For convenience, we write $A^\dagger = B^g_1 \cup \cdots \cup B^G$ and $A^* = A \setminus A^\dagger$.

Let us pause to review the construction of some non-empty sector $\hat{B}^h_p$ outlined above (where $G < h \leq H$). Our aim was to build a collection of stars with intrinsic profile $\hat{w}^h_p = z^h_{p,0}\hat{\mathbf{w}}^h_0 + z^h_{p,1}\hat{\mathbf{w}}^h_1 + \cdots + z^h_{p,L}\hat{\mathbf{w}}^h_L$, that could be organized into galaxies. Essentially, we proceeded as follows. For each ℓ ($0 \leq \ell \leq L$), we took $z^h_{p,\ell}$ sets of stars, each with intrinsic profile $\hat{\mathbf{w}}^h_\ell$. It helps to imagine that, for $\ell = 0$, the sets of stars in question formed 'galactic cores', while for $1 \leq \ell \leq L$, they formed 'peripheral constellations'. Thus, we had at out disposal $z^h_{p,0}$ galactic cores, and for each ℓ ($1 \leq \ell \leq L$), $z^h_{p,\ell}$ peripheral constellations of the ℓth type. The question was: how to distribute the various peripheral constellations between the $z^h_{p,0}$ galactic cores so that each core, together with its allotted constellations, could be used to manufacture a galaxy using Lemma 5.1? According to conditions $\mathfrak{C}_5$ and $\mathfrak{C}_6$, the answer is: any way we like. Any combination of one galactic core and some collection of peripheral constellations is a set of stars whose intrinsic profile satisfies $\mathcal{C}^*$, and hence which may be the domain of a galaxy. Therefore, we made the simplest possible choice, and associated all the peripheral constellations with a single galactic core (to form a galaxy with profile $\underline{\tilde{w}}$), leaving all the other galactic cores, if any, to form galaxies on their own.

We shall require the intrinsic c-spectra of these sets of stars. It follows from (22) and $\mathcal{D}_1$ and $\mathcal{D}_2$ that, for all h ($1 \leq h \leq H$):

$$\mathrm{cs}(\dot{B}^h_p) = \underline{\dot{u}}^h_p \qquad (1 \leq p \leq \mathfrak{b}(h)), \qquad\qquad \mathrm{cs}(\hat{B}^h_p) = \underline{\hat{u}}^h_p \qquad (1 \leq p \leq \mathfrak{b}(h)+1). \tag{23}$$

5.3 Constructing the cosmos: the invertible cosmic rays

We wish to define a model $\mathfrak{A} \models \varphi$ over the set of stars A. If $\mathfrak{B}$ is any of the galaxies formed in the construction of A, then we set $\mathfrak{A}_{|B} = \mathfrak{B}$, thus guaranteeing that, for all $a \in A$, $\mathrm{st}^{\mathfrak{A}}_\star[a] = \mathrm{st}_\star(a)$. That is, all galactic rays emitted by the stars in A have been found absorption sites in the same galaxy as the star emitting them, and all remaining pairs of elements from the same galaxy of A have been assigned a dark galactic 2-type compatible with φ. It remains to specify the 2-types of pairs of elements from different galaxies in such a way that $\mathrm{st}^{\mathfrak{A}}[a] = \mathrm{st}(a)$ for all $a \in A$, and that all dark cosmic types are compatible with φ. We begin with the *invertible* cosmic rays, i.e. those of any of the types $\rho_{4J+1}, \ldots, \rho_{6J}$.

We impose a standard ordering $\prec$ on the sectors and terminators as defined as follows. If $(h,p) \prec (h',p')$, then we take both $\hat{B}^h_p$ and $\dot{B}^h_p$ to precede both $\hat{B}^{h'}_{p'}$ and $\dot{B}^{h'}_{p'}$; and in addition, we take $\hat{B}^h_p$ to be the immediate predecessor of $\dot{B}^h_p$. (For a single value of h, this is the left-to-right ordering illustrated in Fig. 4.) We again use the symbol $\prec$ to denote this ordering, since no confusion should result. Of course, $\prec$ is intended to mirror the ordering on equivalence classes employed in Sec 4.5. Let the collection $\{\hat{B}^h_p, \dot{B}^h_p\}_{h,p}$ of sectors and terminators be enumerated as $B_1, \ldots B_n$, under the standard ordering, $\prec$.

Fix j in the range $1 \leq j \leq 2J$. For all i ($1 \leq i \leq n$), let U_i be the set of rays of invertible cosmic type ρ_{4J+j} emitted by B_i, and let $u_i = |U_i|$. Similarly, let V_i be the set of rays of invertible cosmic type ρ_{4J+j^*} emitted by B_i, and let $v_i = |V_i|$. Let $U = \bigcup_{i=1}^n U_i$ and $V = \bigcup_{i=1}^n V_i$. Our goal is to pair the rays in U 1–1 with those in V so that no resulting pair is emitted by stars from the same B_i. We remark that the sets U_i, V_i, U and V, as well as the extended natural numbers u_i and v_i depend on j; we suppress this dependency in the notation to avoid clutter. Notice, incidentally, that replacing j by j^* systematically exchanges the U_is and V_is.

Keeping the value of j fixed, we observe that $u_1, \ldots, u_n$ is simply a list of the extended natural numbers $\underline{\dot{u}}^h_p[j]$ and $\underline{\hat{u}}^h_p[j]$ in some order, whence $\mathcal{D}_3$ implies $u[j] = \sum_{i=1}^n u_i$. Likewise, $v_1, \ldots, v_n$ is a list of the extended natural numbers $\underline{\dot{u}}^h_p[j^*]$ and $\underline{\hat{u}}^h_p[j^*]$ in some order, whence $u[j^*] = \sum_{i=1}^n v_i$. It follows from $\mathcal{E}_3$ that $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$. Moreover, let k ($1 \leq k \leq n$) be such that $B_k = \dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$. Since we have chosen the standard ordering $\prec$ to list the sectors and terminators $B_1, \ldots, B_n$, it follows from (23) and $\mathcal{D}_4$–$\mathcal{D}_9$ that

$$\begin{aligned} \underline{u}^-[j] &= u_1 + \cdots + u_{k-1} & \qquad \underline{v}^-[j] &= v_1 + \cdots + v_{k-1} \\ \underline{u}^+[j] &= u_{k+1} + \cdots + u_n & \qquad \underline{v}^+[j] &= v_{k+1} + \cdots + v_n \\ \underline{u}^\circ[j] &= u_k & \qquad \underline{v}^\circ[j] &= v_k. \end{aligned}$$

Thus, $\mathcal{E}_4$–$\mathcal{E}_6$ just amount to saying: (*i*) $\sum_{i=1}^{k-1} u_i \leq \sum_{i=k}^{n} v_i$, (*ii*) $\sum_{i=k+1}^{n} u_i \leq \sum_{i=1}^{k} v_i$ and (*iii*) $u_k \leq \sum_{i=1}^{k-1} v_i + \sum_{i=k+1}^{n} v_i$. This is statement (b) of Lemma 2.9; by the equivalent statement (a), $u_i \leq \sum_{i' \neq i} v_{i'}$ for all i $(1 \leq i \leq n)$. Actually, we have a little more. Since we can repeat the same argument with j replaced by the value j^* (which simply exchanges the u's and v's), we have, for all i $(1 \leq i \leq n)$, $v_i \leq \sum_{i' \neq i} u_{i'}$, and, indeed, $(u_i + v_i) \leq \sum_{i' \neq i}(u_{i'} + v_{i'})$ for all i $(1 \leq i \leq n)$. Thus, we have established that, for all j $(1 \leq j \leq 2J)$, no sector or terminator emits more rays of any invertible cosmic type ρ_{4J+j} than the other sectors or terminators (taken together) emit rays of the inverse type ρ_{4J+j^*}; and indeed, for all j $(1 \leq j \leq J)$, no sector or terminator emits more rays of any symmetrized cosmic type $(\rho_{4J+j}, \rho_{5J+j})$ than the other sectors or terminators taken together.

Having established these facts, we proceed with the pairing of the invertible cosmic rays emitted throughout A. We begin with the case where $\mathfrak{C}$—and hence the set of stars A—is finite. (This is actually the more difficult case.) Here, it is easier to work with sets of symmetrized rays. Fix j again, this time in the range $1 \leq j \leq J$. Thus, j specifies a symmetrized ray-type $(\rho_{4J+j}, \rho_{5J+j})$. Define the U_i, V_i, u_i, v_i as above, for this chosen value of j. For all i $(1 \leq i \leq n)$, let $W_i = U_i \cup V_i$, and let $w_i = |W_i|$. Thus, W_i is the set of rays of symmetrized type $(\rho_{4J+j}, \rho_{5J+j})$ emitted by B_i. Let $W = \bigcup_{i=1}^{n} W_i$. We showed above that $\sum_{i=1}^{n} u_i = \sum_{i=1}^{n} v_i$, Moreover, we also showed above that, for all i $(1 \leq i \leq n)$, $(u_i + v_i) \leq \sum_{i' \neq i}(u_{i'} + v_{i'})$. That is, $|W| = \sum_{i=1}^{n} w_i$ is an even number (remember: if $\mathfrak{C}$ is finite, all values in question are finite); and $w_i \leq \sum_{i' \neq i} w_{i'}$ for all i $(1 \leq i \leq n)$. Therefore, by Lemma 2.8, the rays in the various sets W_i can be paired up so that all paired rays belong to different sets in the family—in other words, are emitted by stars lying in different sectors or terminators, and hence certainly by stars lying in different galaxies. Let such a pairing be chosen.

Consider two rays that have been paired up in this process, emitted by stars a and b, say. Either the rays in question are of opposite types (i.e. one ρ_{4J+j} and one ρ_{5J+j}) or of the same type (i.e. both ρ_{4J+j} or both ρ_{5J+j}). In the former case, we say that the pair is *good*, in the latter case, *bad*. We now show how to modify this pairing so that all pairs are good. Observe first that the number of bad pairs in which both rays are of type ρ_{4J+j} equals the number of bad pairs in which both rays are of type ρ_{5J+j}, since, otherwise, the total number of cosmic rays of type ρ_{4J+j} and the number of cosmic rays of type ρ_{5J+j} would be different, contradicting $\mathcal{E}_3$. (In this step, we are using the assumption that the number of rays in question is finite.) Let the set of bad pairs of type ρ_{4J+j} be matched 1–1 with the set of bad pairs of type ρ_{5J+j}. Take any bad pair of type ρ_{4J+j}, and let the stars emitting these (paired) rays be a and b; take the bad pair of type ρ_{5J+j} that is matched with it, and let the stars emitting these (paired) rays be a' and b'. The fact that the relevant star-types are 2-polarized ensures that a, b, a' and b' are all distinct; moreover, by construction, a and b are from different galaxies, as are a' and b'. Suppose first that a and a' are from the same galaxy. Then a and b' are from different galaxies, as are a' and b. Hence we may replace the bad pairs (a, b) and (a', b') with the good pairs (a, b') and (a', b). If b and b' are from the same galaxy, a symmetric argument applies. If a and a' are from different galaxies and b and b' are also from different galaxies, then we may replace the bad pairs (a, b) and (a', b') with the good pairs (a, a') and (b, b'). By doing this for all matched bad pairs, we obtain a pairing in which all pairs are good. Thus, we may pair the invertible cosmic rays emitted by the stars of A of type ρ_{4J+j} 1–1 with the invertible cosmic rays of type ρ_{5J+j} emitted by the stars of A, such that any two rays paired in this way are emitted by stars in different galaxies. This completes the pairing in the case where $\mathfrak{C}$ is finite.

Now consider the case where $\mathfrak{C}$ is infinite. Again, fix j in the range $1 \leq j \leq J$, and define the U_i, V_i, u_i, v_i as above, for this chosen value of j. We showed above that for all i $(i \leq i \leq n)$, $u_i \leq \sum_{i' \neq i} v_{i'}$ and $v_i \leq \sum_{i' \neq i} u_{i'}$. In addition, we showed that $\sum_{i=1}^{n} u_i = \sum_{i=1}^{n} v_i$. If this common sum is finite, we can proceed as in the case where $\mathfrak{C}$ is finite, and there is nothing more to do. Thus, we may assume that there exist i and i' such that $u_i = v_{i'} = \aleph_0$. If $i = i'$, we have the following:

1. B_i emits $\aleph_0$ rays of type ρ_{4J+j} and $\aleph_0$ rays of type ρ_{4J+j^*};
2. $A \setminus B_i$ emits $\aleph_0$ rays of type ρ_{4J+j^*} and $\aleph_0$ rays of type ρ_{4J+j}.

But then we can pair the $\aleph_0$ rays of type ρ_{4J+j} emitted by B_i with the $\aleph_0$ rays of type ρ_{4J+j^*} emitted by $A \setminus B_i$; and similarly with ρ_{4J+j} and ρ_{4J+j^*} interchanged. This is the required pairing. Hence we may assume that $i \neq i'$, and indeed that v_i and $u_{i'}$ are both finite. Moreover, by replacing j by j^* if necessary (which, remember, simply exchanges the U_is and the V_is), we may assume without loss of generality that $v_i \leq u_{i'}$. Now observe the following:

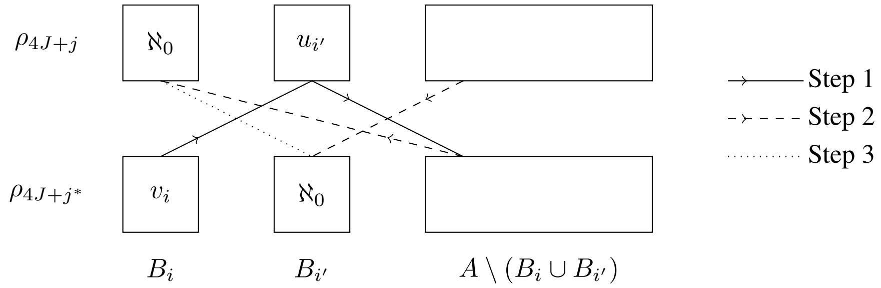


Fig. 6 Fixing the invertible cosmic rays of type ρ_{4J+j} and ρ_{5J+j} (infinite case).

1. B_i emits $\aleph_0$ rays of type ρ_{4J+j} and v_i rays of type ρ_{4J+j^*};
2. $B_{i'}$ emits $\aleph_0$ rays of type ρ_{4J+j^*} and $u_{i'}$ rays of type ρ_{4J+j};
3. $A \setminus B_i$ emits $\aleph_0$ rays of type ρ_{4J+j^*} and at least v_i rays of type ρ_{4J+j};
4. $A \setminus B_{i'}$ emits $\aleph_0$ rays of type ρ_{4J+j} and at least $u_{i'}$ rays of type ρ_{4J+j^*}.

We create the pairing in three steps (Fig. 6). Step 1: pair up all v_i rays of type ρ_{4J+j^*} emitted by B_i with v_i of the rays of type ρ_{4J+j} are emitted by $B_{i'}$; then pair up the $(u_{i'} - v_i)$ remaining rays of type ρ_{4J+j} emitted by $B_{i'}$ with rays of type ρ_{4J+j^*} are emitted by $A \setminus (B_i \cup B_{i'})$. Notice that these are guaranteed to exist, because $A \setminus B_{i'}$ emits at least $u_{i'}$ rays of type ρ_{4J+j^*} and B_i emits only v_i rays of type ρ_{4J+j^*}. Step 2: pair any rays of type ρ_{4J+j^*} emitted by $A \setminus (B_i \cup B_{i'})$ with rays of type ρ_{4J+j} emitted by B_i; this is possible since there are infinitely many of the latter, and indeed, we may perform this pairing in such a way that B_i still emits infinitely many unpaired rays of type ρ_{4J+j}. Likewise, pair any rays of type ρ_{4J+j} emitted by $A \setminus (B_i \cup B_{i'})$ with rays of type ρ_{4J+j^*} emitted by $B_{i'}$, again in such a way that $B_{i'}$ still emits infinitely many unpaired rays of type ρ_{4J+j^*}. Step 3: pair the (infinitely many) unpaired rays of type ρ_{4J+j} emitted by B_i with the (infinitely many) unpaired rays of type ρ_{4J+j^*} emitted by $B_{i'}$. This completes the pairing in the case where $\mathfrak{C}$ is infinite.

Let us carry out this process independently for all j $(1 \leq j \leq J)$. Suppose $a \in A$ emits a ray of type ρ_{4J+j}, paired with ray of type ρ_{5J+j} emitted by b, and let the intrinsic 1-type of a be π and the intrinsic 1-type of b be π'. By the above construction, a and b are from different galaxies; moreover, since the intrinsic star-types of a and b are 2-polarized, a emits no other invertible ray with absorption-type π', and b emits no other invertible ray with absorption-type π. Thus we may simply set $\text{tp}^{\mathfrak{A}}[a, b] = \rho_{4J+j}$, without danger of clashes. In this way, we obtain an absorption site for every invertible cosmic ray emitted by every star in A. This is all thanks to the functions $\mathfrak{h}$ and $\mathfrak{p}$ in $\mathfrak{C}$, which allow us to pick a small collection of equivalence classes—namely, the $\dot{B}^{\mathfrak{h}(j)}_{\mathfrak{p}(j)}$—witnessing the fact that no equivalence class emits more rays of invertible cosmic type than the rest of the cosmos can absorb.

5.4 Constructing the cosmos: the non-invertible cosmic rays

Recall that the set of stars A is partitioned into $A^\dagger$ and A^* (though the latter set may be empty). We consider first the non-invertible cosmic rays emitted by the stars in A^*. The plan is to find stars in $A^\dagger$ to absorb them. Fix i $(1 \leq i \leq I)$. Let a be a star in A^*: we consider any non-invertible cosmic rays with absorption type π_i emitted by a. Note that a belongs to a galaxy whose intrinsic profile satisfies the linear Diophantine clauses $\mathcal{C}^*$. Suppose first that $\mathcal{G}_i = \emptyset$. Then, by $\mathcal{C}^0_8$, a emits no cosmic rays with absorption type π_i, and so there is nothing to do. Suppose now that $i \in \mathcal{I}$. By $\mathfrak{C}_3$, all the statements in $\mathcal{B}$ hold; and from $\mathcal{B}_2$, $\mathcal{G}_i = \{g\}$ for some g $(1 \leq g \leq G)$. By $\mathcal{C}^g_2$, B^g_2 contains a star b with 1-type π_i. On the other hand, by $\mathcal{C}^0_7$, a cannot emit more than one cosmic ray with absorption type π_i. Thus, if a emits a non-invertible cosmic ray, say of type ρ_j, having absorption type π_i, it cannot emit any invertible cosmic ray with absorption type π_i, whence $\text{tp}^{\mathfrak{A}}[a, b]$ must currently be undefined. Thus, we may set $\text{tp}^{\mathfrak{A}}[a, b] = \rho_j$, and we have found an absorption site for the ray in question. Henceforth, then, we may assume $i \notin \mathcal{I}$. Suppose that, nevertheless, $|\mathcal{G}_i| = 1$, and let $\mathcal{G}_i = \{g\}$, where $1 \leq g \leq G$. By $\mathcal{C}^g_3$ and $\mathcal{C}^0_2$, $B^g_1 \subseteq A^\dagger$ contains at least $Z \geq mM$ stars having intrinsic 1-type π_i. Since a cannot emit more than this number of rays in total, we can certainly find fresh stars in $A^\dagger$ (i.e., stars b for which $\text{tp}^{\mathfrak{A}}[a,b]$ has not been defined) to absorb all the non-invertible cosmic rays with absorption type π_i emitted by a. If b is chosen to absorb such a ray of type ρ_j, we set $\text{tp}^{\mathfrak{A}}[a,b] = \rho_j$. Thus, we may assume $|\mathcal{G}_i| > 1$, and hence, by $\mathcal{B}_1$, $|\mathcal{G}_i| \geq Z \geq mM$. But then, by $\mathcal{C}_2^g$ (where $g \in \mathcal{G}_i$), $A^\dagger$ contains at least mM stars of 1-type π_i, and we may proceed as in the previous case. Evidently this procedure can be executed for all $a \in A^*$, and for all i $(1 \leq i \leq I)$, with none of these 2-type assignments overwriting any other. At the end of this process, every ray (galactic or cosmic) emitted by any star $a \in A^*$, has been found an absorption site $b \in A^\dagger$; and the relevant 2-type $\text{tp}^{\mathfrak{A}}[a,b]$ has been set to be the type of that ray.

We consider next the non-invertible cosmic rays emitted by the stars in $A^\dagger$. Fix integers i and i' $(1 \leq i, i' \leq I)$, and write $\pi = \pi_i$ and $\pi' = \pi_{i'}$. (We allow the possibility that $i = i'$.) We proceed to find absorption sites for all non-invertible cosmic rays of absorption-type π' emitted by stars in $A^\dagger$ having intrinsic 1-type π, and, simultaneously, absorption sites for all non-invertible galactic rays of absorption-type π emitted by stars in $A^\dagger$ having intrinsic 1-type π'. If $\mathcal{G}_i$ is empty, then, by the constraints $\mathcal{C}_1^g$ (where $1 \leq g \leq G$), there are no cosmic rays with emission-type π, and by $\mathcal{C}_8^0$, there are no cosmic rays with absorption-type π, whence there is nothing to do. A symmetrical argument applies if $\mathcal{G}_{i'}$ is empty. Henceforth then, we assume that $\mathcal{G}_i$ and $\mathcal{G}_{i'}$ have cardinality at least 1. We have three cases to consider.

Case 1: $i \in \mathcal{I}$. By $\mathcal{B}_2$, $\mathcal{G}_i = \{g\}$ for some g $(1 \leq g \leq G)$. By $\mathcal{C}_2^g$, B_1^g contains at least one star with intrinsic 1-type π, and by $\mathcal{C}_6^0$, exactly one. Denote that star by a. Now let b be any star of $A^\dagger$ with intrinsic 1-type π' emitting a non-invertible cosmic ray with absorption-type π. By $\mathcal{C}_5^g$, $b \notin B_1^g$, and indeed $i \neq i'$; moreover, by $\mathcal{C}_7^0$, b emits exactly one such ray—say of type ρ_j, and indeed emits no invertible cosmic ray with absorption-type π. It follows that $\text{tp}^{\mathfrak{A}}[b,a]$ is currently undefined; so we can select a to absorb the non-invertible cosmic ray emitted by b with absorption type π, setting $\text{tp}^{\mathfrak{A}}[b,a] = \rho_j$. In the other direction, let a emit d cosmic rays (invertible or non-invertible) having absorption type π', where $d \leq mM$. Thus, $\underline{\mathbf{q}}^d_{i,i'} \cdot \underline{w}^g > 0$, whence, by $\mathcal{C}_6^g$, $\sum\{\underline{\mathbf{q}}^*_{i',i} \cdot \underline{w}^h \mid 1 \leq h \leq G,\ h \neq g\} \geq d$, i.e. there exist at least d stars in $b \in A^\dagger \setminus B_1^g$ having 1-type π', and which emit no non-invertible cosmic ray with absorption type π. For any such b, $\text{tp}^{\mathfrak{A}}[a,b]$, either is currently undefined, or b already absorbs an invertible ray sent by a. Now, if a emits any invertible cosmic ray with absorption type π', then this will have been accounted for previously, and so it (and the star chosen to absorb it) may be disregarded. Considering the remaining (non-invertible) cosmic rays emitted by a with absorption type π', then, we can find, for each such ray—and having type, say, ρ_j—a fresh absorption site b of type π emitting no ray with absorption type π; we then assign $\text{tp}^{\mathfrak{A}}[a,b] = \rho_j$.

Case 2: $i' \in \mathcal{I}$. We proceed symmetrically to Case 1.

Case 3: $i, i' \notin \mathcal{I}$. By $\mathcal{B}_1$, either $|\mathcal{G}_i| = 1$ or $|\mathcal{G}_i| \geq Z \geq 2mM$, and either $|\mathcal{G}_{i'}| = 1$ or $|\mathcal{G}_{i'}| \geq Z$. We consider the four resulting sub-cases in turn. Observe first however, the combined effect of $\mathcal{C}_2^0$ and $\mathcal{C}_3^g$: if $\mathcal{G}_i = \{g\}$ for some g $(1 \leq g \leq G)$, then, as we are now supposing $i \notin \mathcal{I}$, it follows that B_1^g contains at least $Z \geq 3mM$ stars with intrinsic 1-type π; and similarly for π'.

Sub-case 3(*i*): $|\mathcal{G}_i| = |\mathcal{G}_{i'}| = 1$. Let $\mathcal{G}_i = \{g\}$, $\mathcal{G}_{i'} = \{g'\}$, If $g = g'$, then by the constraints $\mathcal{C}_1^{g''}$ (where $g'' \neq g$), no elements of $A^\dagger \setminus B_1^g$ can have 1-type π or π', and by $\mathcal{C}_5^g$, no elements of B_1^g can emit any cosmic rays with absorption type π or π', whence there are no rays to consider, and nothing to do. If, on the other hand, $g \neq g'$, let D_i be the set of stars (in B_1^g) having intrinsic 1-type π, and partition D_i into two subsets, $D_{i,0}$ and $D_{i,1}$, each of cardinality at least mM. Do the same for i'. Each star $a \in D_{i,\ell}$ $(\ell = 0, 1)$, emits a collection of at most mM cosmic rays with absorption type π'; for each non-invertible ray in this collection, having type, say, ρ_j, we may choose a fresh absorption site $b \in D_{i',\ell}$, and assign $\text{tp}^{\mathfrak{A}}[a,b] = \rho_j$. (If a emits an invertible cosmic ray with absorption type π', this will already have been found an absorption site; and we choose other absorption sites in $D_{i',\ell}$ for the non-invertible rays.) Likewise, for each non-invertible cosmic ray with absorption type π_i emitted by $b \in D_{i',\ell}$ $(\ell = 0, 1)$, and having type, say, ρ_j, we may choose a fresh absorption site $a \in D_{i,\ell+1}$ (addition in subscript modulo 2), and assign $\text{tp}^{\mathfrak{A}}[b,a] = \rho_j$. The partitions $\{D_{i,0}, D_{i,1}\}$ and $\{D_{i',0}, D_{i',1}\}$ ensure that these assignments do not overwrite each other (Fig. 7). We remark in this context that D_i and $D_{i'}$ are necessarily disjoint (and in fact $i \neq i'$).

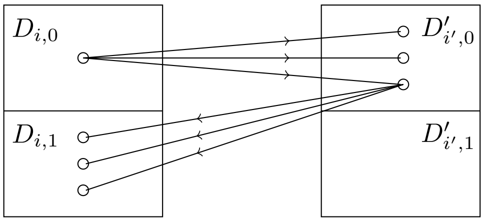


Fig. 7 Finding absorption sites for non-invertible cosmic rays

Sub-case 3(*ii*): $|\mathcal{G}_i| = 1$, but $|\mathcal{G}_{i'}| \geq Z$. Let D_i be the set of stars having intrinsic 1-type π, and write $\mathcal{G}_i = \{g\}$. By the constraints $\mathcal{C}_1^{g'}$ ($g' \neq g$), we have $D_i \subseteq B_1^g$. By $\mathcal{C}_5^g$, no star in B_1^g—and in particular, no such star with intrinsic 1-type π'—can emit any cosmic rays with absorption type π; so let $D_{i'}$ be the the stars in $A^\dagger \setminus B_1^g$ having intrinsic 1-type π'. By the constraints $\mathcal{C}_2^{g'}$ ($g' \neq g$), we have $|D_{i'}| \geq (Z-1) \geq 2mM$. Now proceed as for the first sub-case.

Sub-case 3(*iii*): $|\mathcal{G}_{i'}| = 1$, but $|\mathcal{G}_i| \geq Z$. Symmetric to Sub-case 3(*ii*):

Sub-case 3(*iv*): $|\mathcal{G}_i| \geq Z$ and $|\mathcal{G}_{i'}| \geq Z$. Partition $\mathcal{G}_i$ into three sets, $\mathcal{G}_{i,0}$, $\mathcal{G}_{i,1}$ and $\mathcal{G}_{i,2}$, each with cardinality exactly mM. Partition $\mathcal{G}_{i'}$ into $\mathcal{G}_{i',0}$, $\mathcal{G}_{i',1}$, $\mathcal{G}_{i',2}$, similarly. It is easy to see that these partitions can be chosen such that $g \in \mathcal{G}_{i,\ell}$ and $g \in \mathcal{G}_{i'}$ implies $g \in \mathcal{G}_{i',\ell}$. For $\ell = 0, 1, 2$, let $D_{i,\ell}$ be the set of stars with intrinsic 1-type π lying in some B_1^g with $g \in \mathcal{G}_{i,\ell}$, and let $D_{i',\ell}$ be the set of stars with intrinsic 1-type π' lying in some B_1^g with $g \in \mathcal{G}_{i',\ell}$. By the various constraints $\mathcal{C}_2^g$ ($1 \leq g \leq G$), we have $|D_{i,\ell}| \geq mM$, for $\ell = 0, 1, 2$; similarly for $D_{i',\ell}$. Notice that, by the way in which $\mathcal{G}_{i'}$ and $\mathcal{G}_{i'}$ are partitioned, $\ell \neq \ell'$ implies $D_{\ell,i} \cap D_{\ell',i'} = \emptyset$. Now, each star $a \in D_{i,\ell}$ ($\ell = 0, 1, 2$), emits a collection of at most mM cosmic rays with absorption type π'; for each non-invertible ray in this collection, having type, say, ρ_j, we may choose a fresh absorption site $b \in D_{i',\ell+1}$ (addition in subscripts modulo 3), and assign $tp^{\mathfrak{A}}[a, b] = \rho_j$. Similarly, each $b \in D_{i',\ell}$ ($\ell = 0, 1, 2$), emits a collection of at most mM cosmic rays with absorption type π; for each non-invertible ray in this collection, having type, say, ρ_j, we may choose a fresh absorption site $a \in D_{i,\ell+1}$ (addition in subscripts modulo 3), and assign $tp^{\mathfrak{A}}[b, a] = \rho_j$. By arranging the various sets $D_{i,\ell}$ and $D_{i',\ell}$ similarly to the sets B_ℓ and B'_ℓ in Fig. 5, we see that these assignments do not overwrite each other, and that no ray is emitted and absorbed by stars in the same galaxy.

Clearly, we may carry out the above assignments for all pairs of indices i, i', without danger of overwriting. At the end of this process, every ray (galactic or cosmic, invertible or non-invertible), emitted by any star $a \in A$, has been found an absorption site $b \in A$; and the relevant 2-type $\text{tp}^{\mathfrak{A}}[a, b]$ has been set to be the type of this ray.

5.5 Constructing the cosmos: the dark cosmic 2-types

To complete the construction of $\mathfrak{A}$, we must set any two-types not considered in the above process to be a dark cosmic 2-type compatible with φ. (We require a dark cosmic 2-type, because, having found absorption sites for all rays, we want to secure $\text{st}^{\mathfrak{A}}[a] = \text{st}(a)$ for any $a \in A$.) Suppose, then, that a and b are stars from different galaxies, and let a have 1-type $\pi = \pi_i$, and b, 1-type $\pi' = \pi_{i'}$. If π and π' are not c-coupled, and $\text{tp}^{\mathfrak{A}}[a, b]$ has not yet been defined, we can simply choose any dark cosmic 2-type τ compatible with φ and satisfying $\text{tp}_1(\tau) = \pi$, $\text{tp}_2(\tau) = \pi'$, and set $\text{tp}^{\mathfrak{A}}[a, b] = \tau$. It therefore suffices to show that, if π and π' *are* c-coupled, then either a has been chosen to absorb a ray emitted by b or vice versa, and therefore that $\text{tp}^{\mathfrak{A}}[a, b]$ has already been defined.

Suppose, then, π and π' are c-coupled. Since π and π' are by assumption both realized, the sets of cluases $\mathcal{C}_1^*$ and $\mathcal{C}_1^g$ guarantee that both $\mathcal{G}_i$ and $\mathcal{G}_{i'}$ are non-empty; indeed we cannot have $\mathcal{G}_i = \mathcal{G}_{i'}$ with this set a singleton, since a and b are, by hypothesis, from different galaxies. By $\mathcal{B}_3$, then, either $i \in \mathcal{I}$ or $i' \in \mathcal{I}$. Suppose the former. By $\mathcal{B}_2$, we have $\mathcal{G}_i = \{g\}$ for some g ($1 \leq g \leq G$), and by the various constraints $\mathcal{C}_1^{g''}$ ($g'' \neq g$), $\mathcal{C}_6^0$ and $\mathcal{C}_1^*$, $a \in B_1^g$ is the only star in A with intrinsic 1-type π. Now if $b \in A^*$, then, by $\mathcal{C}_2^*$, b emits a ray with absorption-type π. Certainly, then, a will have been chosen to absorb this ray, since it is the only candidate with the correct intrinsic 1-type. Moreover, if $b \in A^\dagger$ emits any cosmic ray with absorption-type π, then that ray must likewise have been absorbed by a. Thus we may suppose that b belongs to the set D of stars in $A^\dagger$ with 1-type π' that do not emit any cosmic rays with absorption type π. By construction of $A^\dagger$,

$$|D| = \sum\{\underline{q}^*_{i',i} \cdot \underline{w}^h \mid 1 \leq h \leq G,\ h \neq g\}.$$

Now suppose a emits $d' \leq mM$ non-invertible cosmic rays with absorption type π'. Thus, $\underline{q}^{d'+1}_{i,i'} \cdot \underline{w}^g = 0$, so that, setting the parameter d in $\mathcal{C}^g_7$ to $d'+1$, we have $|D| \leq d'$; and each such ray will have been chosen a fresh absorption site among the elements of D, thus covering that set. Hence, $\mathrm{tp}^{\mathfrak{A}}[a,b]$ has already been defined, as required.

This completes the construction of $\mathfrak{A}$. Since every ray emitted by every star has been found an absorption site, and all other 2-types assigned are dark, we have $\mathrm{st}^{\mathfrak{A}}[a] = \mathrm{st}(a)$, for all $a \in A$. But each of the star-types $\sigma_1, \ldots, \sigma_K$ is compatible with φ, as are all the assigned dark 2-types, whence $\mathfrak{A} \models \varphi$. We have shown:

Lemma 5.2 *Suppose φ is a $\mathcal{C}^2 1E$-formula in the form* (1). *If a* (*finite*) *certificate for φ exists, then φ is* (*finitely*) *satisfiable.*

6 Proof of main result

Lemma 4.5 states that, if a formula φ of the form (1) is satisfiable, there exists a certificate

$$\mathfrak{C} = \langle G, H, I, J, K, L, \{\sigma_k\}, \mathcal{I}, \{\mathcal{G}_i\}, \mathfrak{b}, \mathfrak{h}, \mathfrak{p}, \{\mathcal{E}^h\}, \{\underline{\mathbf{w}}^h_\ell\}, \underline{a}^\dagger, \underline{\dot{b}}, \underline{\hat{b}}\rangle \tag{24}$$

satisfying conditions $\mathfrak{C}1$–$\mathfrak{C}7$, with $\|\mathfrak{C}\|$ doubly exponentially bounded. We proceed to show that, in fact, such a $\mathfrak{C}$ exists with $\|\mathfrak{C}\|$ singly exponentially bounded.

We begin with the numbers H and L. Condition $\mathfrak{C}7$ states that $\underline{a}^\dagger$ is a solution of the system of positive integer clauses $\mathcal{C}^\dagger(\underline{w}^\dagger)$, and that $\underline{a}^\dagger, \underline{\dot{b}}, \underline{\hat{b}}$ is a solution of the system of extended integer equations and inequalities $\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$. We proceed to construct solutions in which at most singly exponentially many values are non-zero. Our basic tools will be Proposition 2.4 and Corollary 2.6. However, both $\mathcal{C}^\dagger$ and $\mathcal{E}_2$ are systems of positive integer *clauses*; moreover, $|\mathcal{E}_2|$ is at best doubly exponentially bounded. For this reason, we must proceed with caution.

We begin with the set of clauses $\mathcal{C}^\dagger(\underline{w}^\dagger)$. For each clause in this set, select some disjunct of that clause satisfied by $\underline{a}^\dagger$; and let $\tilde{\mathcal{C}}(\underline{w}^\dagger)$ be the set of selected equations and inequalities. Thus, $\tilde{\mathcal{C}}(\underline{w}^\dagger)$ is also satisfied by $\underline{a}^\dagger$, propositionally entails $\mathcal{C}^\dagger(\underline{w}^\dagger)$, and has the same cardinality as $\mathcal{C}^\dagger(\underline{w}^\dagger)$. Let $\underline{w}$ be the tuple of variables $\underline{w}^\dagger, \underline{\dot{z}}$, and let $\underline{a}$ be the corresponding tuple of constants $\underline{a}^\dagger, \underline{\dot{b}}$. Thus, $\underline{a}$, $\underline{\hat{b}}$ is a solution of the system of extended integer equations and inequalities

$$\mathcal{F}(\underline{w}, \underline{\hat{z}}) = \tilde{\mathcal{C}}(\underline{w}^\dagger) \cup \mathcal{E}_1(\underline{w}, \underline{\hat{z}}) \cup \mathcal{E}_3(\underline{w}, \underline{\hat{z}}) \cup \mathcal{E}_4(\underline{w}, \underline{\hat{z}}) \cup \mathcal{E}_5(\underline{w}, \underline{\hat{z}}) \cup \cup \mathcal{E}_6(\underline{w}, \underline{\hat{z}}).$$

And, of course, $\underline{\hat{b}}$ a solution of $\mathcal{E}_2(\underline{\hat{z}})$. (Note that $\mathcal{E}_2$ does not feature the variables $\underline{w}^\dagger, \underline{\dot{z}}$.) Let $n = |\mathcal{F}|$, let N_v be the maximum value of any finite *variable* coefficient in $\mathcal{F}$, and let N be the maximum value of any finite coefficient in $\mathcal{F}$. Thus, we see that n is exponentially bounded, and N (and hence N_v) is doubly exponentially bounded.

Let us consider now the set of positive integer clauses $\mathcal{E}_2$; we repeat it here for ease of reference:

$$\left\{ (\hat{z}^h_{p,0} \geq 1) \vee \left(\sum_{\ell=1}^{L} \hat{z}^h_{p,\ell} = 0 \right) \;\middle|\; G < h \leq H,\ 1 \leq p \leq \mathfrak{b}(h) + 1 \right\}. \tag{$\mathcal{E}_2$}$$

The tuple $\underline{\hat{z}}$ is simply an ordering of the collection of variables $\{\dot{z}^h_{p,\ell}\}$, with h, p and ℓ varying over their usual ranges (i.e., $G < h \leq H$, $1 \leq p \leq \mathfrak{b}(h) + 1$, $0 \leq \ell \leq L$). Thus, we may write $\hat{b}^h_{p,\ell}$ to denote that element of the tuple $\underline{\hat{b}}$ corresponding to $\hat{z}^h_{p,\ell}$. Reordering $\underline{\hat{z}}$ if necessary, let us write $\underline{\hat{z}} = \underline{z}_1, \underline{z}_2, \underline{z}_3$, where: (*i*) $\underline{z}_1$ is the tuple of all variables $\hat{z}^h_{p,0}$; (*ii*) $\underline{z}_2$ is the tuple of all variables $\hat{z}^h_{p,\ell}$, with $\ell \geq 1$, for which $\hat{b}^h_{p,\ell} > 0$; (*iii*) $\underline{z}_3$ is the tuple of all variables $\hat{z}^h_{p,\ell}$, with $\ell \geq 1$, for which $\hat{b}^h_{p,\ell} = 0$. Re-ordering the $\underline{\hat{b}} = \underline{b}_1, \underline{b}_2, \underline{0}$ correspondingly, we see that $\mathcal{F}(\underline{w}, \underline{z}_1, \underline{z}_2, \underline{z}_3)$ has solution $\underline{a}, \underline{b}_1, \underline{b}_2, \underline{0}$.

Now freeze the values $\underline{a}$, $\underline{b}_1$ and $\underline{0}$ for the moment. That is, consider the system of extended integer equations and inequalities

$$\mathcal{F}'(\underline{z}_2) = \mathcal{F}(\underline{a}, \underline{b}_1, \underline{z}_2, \underline{0}).$$

Notice that $|\mathcal{F}'| = |\mathcal{F}| = n$, and that the largest finite variable coefficient in $\mathcal{F}'$ is still at most N_v. Since $\mathcal{F}'(\underline{z}_2)$ has a solution over $\mathbb{N}^*$, by Corollary 2.6, it has a solution $\underline{b}'_2$, in which at most exponentially many values are positive. Reordering the variables $\underline{z}_2$ if necessary, let us write $\underline{z}_2 = \underline{z}_{2,1}, \underline{z}_{2,2}$ and $\underline{b}'_2 = \underline{b}_{2,1}, \underline{0}$, where the tuple $\underline{b}_{2,1}$ is of at most exponential length, and $\mathcal{F}(\underline{w}, \underline{z}_1, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$ has solution $\underline{a}, \underline{b}_1, \underline{b}_{2,1}, \underline{0}, \underline{0}$. Notice also, however, that the tuple $\underline{b}_1, \underline{b}_{2,1}, \underline{0}, \underline{0}$ satisfies $\mathcal{E}_2(\underline{z}_1, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$, since the values of all the variables $z^h_{p,0}$ remain unchanged from $\hat{\underline{b}}$, as do the values of all the variables $z^h_{p,\ell}$ ($\ell \geq 1$) such that $\hat{b}^h_{p,\ell} = 0$.

Reordering the variables $\underline{z}_1$ if necessary (which are all of the form $\hat{z}^h_{0,p}$), let us write $\underline{z}_1 = \underline{z}_{1,1}, \underline{z}_{1,2}$, where $\underline{z}_{1,1}$ consists of those variables $\hat{z}^h_{0,p}$ such that, for some ℓ ($1 \leq \ell \leq L$), $\hat{z}^h_{\ell,p}$ is in $\underline{z}_{2,1}$ (i.e. was assigned a non-zero value in the tuple $\underline{b}'_2$). Since $\underline{z}_{2,1}$ has at most exponential length, so has $\underline{z}_{1,1}$. Reordering $\underline{b}_1 = \underline{b}_{1,1}, \underline{b}_{1,2}$ correspondingly, we see that $\mathcal{F}(\underline{w}, \underline{z}_{1,1}, \underline{z}_{1,2}, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$ has solution $\underline{a}, \underline{b}_{1,1}, \underline{b}_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}$, and $\mathcal{E}_2(\underline{z}_{1,1}, \underline{z}_{1,2}, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$ has solution $\underline{b}_{1,1}, \underline{b}_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}$.

Now freeze the values $\underline{b}_{1,1}$ and $\underline{b}_{2,1}$, together with both the $\underline{0}$s. That is, consider the system of extended integer equations and inequalities

$$\mathcal{F}''(\underline{w}, \underline{z}_{1,2}) = \mathcal{F}(\underline{w}, \underline{b}_{1,1}, \underline{z}_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}).$$

The number of equations in $\mathcal{F}''$ is still n, and the largest finite variable coefficient is still at most N_v. Since $\mathcal{F}''(\underline{w}, \underline{z}_{1,2})$ has a solution over $\mathbb{N}^*$, by Corollary 2.6, it has a solution $\underline{a}', \underline{b}'_{1,2}$, in which at most exponentially many values are positive. We claim that $\underline{b}_{1,1}, \underline{b}'_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}$ is a solution of $\mathcal{E}_2(\underline{z}_{1,1}, \underline{z}_{1,2}, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$. For the only variables whose values have changed (in comparison to the solution $\underline{b}_{1,1}, \underline{b}_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}$) are those in $\underline{z}_{1,2}$, i.e. those $z^h_{0,p}$ for which all the corresponding variables $z^h_{\ell,p}$ ($\ell \geq 1$) are in $\underline{z}_{2,2}$, which guarantees that the sums $\sum_{\ell=1}^{L} \hat{z}^h_{p,\ell}$ occurring in the relevant clauses of $\mathcal{E}_2$ are zero.

Hence $(\mathcal{F} \cup \mathcal{E}_2)(\underline{w}, \underline{z}_{1,1}, \underline{z}_{1,2}, \underline{z}_{2,1}, \underline{z}_{2,2}, \underline{z}_3)$ has the solution $\underline{a}', \underline{b}_{1,1}, \underline{b}'_{1,2}, \underline{b}_{2,1}, \underline{0}, \underline{0}$, featuring only exponentially many positive values. By renumbering if necessary, this means that H and L may be assumed to be singly exponentially bounded, since all terms involving zero-valued variables can be simply deleted from the equations and inequalities in $\mathcal{E}$.

We next deal with the value K. Observe first that, as we have just shown, we may assume that only exponentially many of the values in the tuple $\underline{a}^\dagger$ are non-zero. (Model-theoretically speaking, this means that, among the special elements of the constructed cosmos, only exponentially many star-types are realized.) Moreover, as we earlier established, each of the vectors $\underline{\mathbf{w}}^h_\ell$ has at most exponentially large footprint, and we have just shown (via the bounds on H and L) that there are only exponentially many of them. (Model-theoretically speaking, this means that, among the ordinary elements of the constructed cosmos, only exponentially many star-types are realized.) Thus, we may safely ignore all but at most singly exponentially many star-types in our enumeration $\sigma_1, \ldots, \sigma_K$. By projecting out components of vectors corresponding to unrealized star-types, we may assume K too is exponentially bounded.

Finally, we consider the vectors $\underline{a}^\dagger$, $\dot{\underline{b}}$ and $\hat{\underline{b}}$. Having established singly exponential bounds on H, K and L, these are simply required as solutions over $\mathbb{N}^*$ to a system of extended integer equations and inequalities in exponentially many variables, with all coefficients doubly exponentially finitely bounded. By Corollary 2.5, such a solution exists which is doubly exponentially finitely bounded, yielding a $\mathfrak{C}$ of singly exponential size.

The argument the case where $\mathfrak{C}$ is finite proceeds in exactly the same way, noting that we are dealing only with systems of Diophantine integer clauses with solutions over $\mathbb{N}$. The only change of wording required is that we use Proposition 2.4 instead of Corollary 2.6, and Corollary 2.3, instead of Corollary 2.5.

We have thus strengthened Lemma 4.5.

Lemma 6.1 *Suppose φ is a $\mathcal{C}^2 1E$-formula in normal form. If φ is (finitely) satisfiable, then φ has a (finite) certificate $\mathfrak{C}$, with $\|\mathfrak{C}\|$ singly exponentially bounded as a function of $\|\varphi\|$.*

We can now prove the main theorem of this paper.

Theorem 6.2 *The (finite) satisfiability problem for $\mathcal{C}^2 1E$ is* NEXPTIME*-complete.*

Proof. It is obvious that we can check whether a (finite) certificate $\mathfrak{C}$ satisfies the conditions $\mathfrak{C}1$–$\mathfrak{C}7$ given in Sec. 4.6 in time bounded by a polynomial function of $\|\mathfrak{C}\|$. Equally obviously, given a $\mathcal{C}^2$-formula φ with ceiling M, and a structure $\mathfrak{A}$ with $|A| \leq M$, we can determine in time bounded by a polynomial function of $\|\varphi\|$ and M (and hence by a singly exponential function of $\|\varphi\|$) whether $\mathfrak{A} \models \varphi$. The upper bound then follows from Lemmas 2.1, 5.2 and 6.1. The lower bound follows from the well-known fact that the (finite) satisfiability problem for the two-variable fragment of first-order logic is NExpTime-hard (see, e.g. [26], p. 255). □

Corollary 6.3 *If φ is a finitely satisfiable $\mathcal{C}^2$1E-formula, then φ has a model of size bounded by a doubly exponential function of* $\|\varphi\|$.

Proof. Immediate from the bounds on $\|\mathfrak{C}\|$ established above together with the model-construction process in Sec. 5. □

7 Two equivalence relations

In this section, we show that the satisfiability and finite satisfiability problems for $\mathcal{C}^2$2E are both undecidable.

A *deterministic* 2*-counter machine* $\mathbf{M}$ has a finite set of states $s_0, \ldots, s_L$ and two counters, c_1 and c_2, each holding a non-negative integer. We regard s_0 as a start state and s_L as a stop state. The basic operations of $\mathbf{M}$ are: test whether c_i holds the value 0; and increment/decrement c_i (where attempting to decrement zero yields zero). The program of $\mathbf{M}$ associates with each state s_ℓ other than s_L a basic operation (i.e. a zero-test, increment or decrement), together with a specification of the next state of the machine (depending, in the case of of zero-tests, on the outcome). No action is specified for the stop state. A *configuration* for $\mathbf{M}$ is a triple comprising a state together with the values of c_1 and c_2. The *run* of $\mathbf{M}$ is the (finite or infinite) sequence of configurations starting with $\langle s_0, 0, 0\rangle$, where each configuration is obtained from its predecessor as specified by the program of $\mathbf{M}$, in the obvious way. We allow this sequence to stop if a configuration featuring the stop state, s_L, is encountered, in which case we say that the machine $\mathbf{M}$ *terminates*. It is well-known that deterministic Turing machines may be effectively simulated by deterministic 2-counter machines. Hence, the problem of deciding whether a given deterministic 2-counter machine terminates is r.e.-complete.

We proceed to show how runs of deterministic 2-counter machines can be encoded using the logic $\mathcal{C}^2$2E. Recall that, in $\mathcal{C}^2$2E, the distinguished binary predicates E_1 and E_2 must be interpreted as equivalences. Where a structure $\mathfrak{A}$ is clear from context, we refer to the equivalence classes of $E_1^{\mathfrak{A}}$ as E_1*-classes*, and similarly for E_2. Note that the coarsest common refinement $E_1^{\mathfrak{A}} \cap E_2^{\mathfrak{A}}$ of these two equivalences is also an equivalence; to aid intuition, we refer to its equivalence classes as *configurations*. We write $E_{12}(x, y)$ as an abbreviation for the formula $E_1(x, y) \wedge E_2(x, y)$. We employ unary predicates d_1, d_2 to partition the universe, in such a way that, within any E_1- or E_2-class, the elements satisfying them form configurations:

$$\forall x((d_1(x) \vee d_2(x)) \wedge (\neg d_1(x) \vee \neg d_2(x))) \tag{25}$$

$$\bigwedge_{k=1}^{2} \forall x \forall y(E_{12}(x, y) \wedge d_k(x) \rightarrow d_k(y)) \tag{26}$$

$$\bigwedge_{k=1}^{2} \bigwedge_{j=1}^{2} \forall x \forall y(E_k(x, y) \wedge d_j(x) \wedge d_j(y) \rightarrow E_{3-k}(x, y)). \tag{27}$$

We call a configration whose elements satisfy d_k a d_k*-configuration*. It follows that each equivalence class contains at most one d_1-configuration, and at most one d_2-configuration. Where two different configurations, B and B', lie in some E_k-class ($k \in \{1, 2\}$), then we say that B' is the *successor of* B if B is a d_k-configuration and B' a d_{3-k}-configuration. Thus, for B and B' as described, one is the successor of the other. Successors, where they exist, are obviously unique.

We employ unary predicates $s_1, \ldots, s_L$, and refer to them as *states*; we also employ an additional unary predicate s to stand for their disjunction. We require that every configuration contains a unique element satisfying

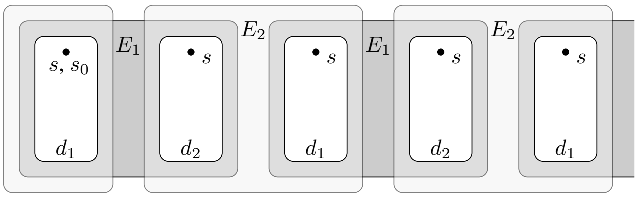


Fig. 8 Initial segment of a chain of configurations: each configuration (white region) contains a unique s-element determining its state; the first configuration is in the start state, and forms an E_2-class on its own.

s, which will be in exactly one state:

$$\forall x \exists_{[=1]} y (E_{12}(x,y) \wedge s(y)) \tag{28}$$

$$\forall x \left(s(x) \rightarrow \bigvee_{\ell=1}^{L} s_\ell(x) \right) \wedge \bigwedge_{1 \leq \ell < \ell' \leq L} \forall x (s_\ell(x) \rightarrow \neg s_{\ell'}(x)). \tag{29}$$

A configuration whose s-element satisfies s_ℓ will be said to be *in state* s_ℓ. We call s_0 the *start state* and s_L, the *stop state*. We employ a binary predicate t, and we require that $t(x,y)$ holds only between s-elements of configurations one of which is the successor of the other:

$$\forall x \forall y (t(x,y) \rightarrow (s(x) \wedge s(y) \wedge \bigvee_{k=1}^{2} (E_k(x,y) \wedge \neg E_{3-k}(x,y) \wedge d_k(x)))). \tag{30}$$

We require that there exists a d_1-configuration in the start state, that this configuration is the only one in its E_2-class (i.e., is not the successor of any configuration), and that every configuration in a state other than the stop state has a successor:

$$\exists x (d_1(x) \wedge s_0(x) \wedge \forall y (E_2(x,y) \rightarrow E_1(x,y))) \tag{31}$$

$$\bigwedge_{\ell=0}^{L-1} \forall x (s_\ell(x) \rightarrow \exists y . t(x,y)). \tag{32}$$

It follows that, in any model of (25)–(32), there is a chain, $B_0, B_1, \ldots$, (possibly infinite) of distinct configurations, where B_0 is in the start state, and where each B_{i+1} is the successor of B_i. Moreover, if this chain is finite and maximal (i.e. cannot be extended), then its final configuration must be in the stop state. Notice that this condition must obtain if the model is finite. The situation is illustrated in Fig. 8.

Recall that, if B is any configuration, then B contains exactly one element satisfying s. We employ two further unary predicates c_1 and c_2: we refer to the set of elements of B satisfying c_i ($1 \leq i \leq 2$) as the *the c_i-counter in B*, and we refer to the cardinality of this set as the *value of* that counter. It helps to assume that the sets of elements of B satisfying the respective predicates s, c_1 and c_2 partition B; however, this is not formally a requirement.

We now consider any deterministic 2-register machine, $\mathbf{M}$, and proceed to describe the run of $\mathbf{M}$ using $\mathcal{C}^2$2E-formulas. We first define, for $i = 1, 2$, a 1-place formula $c_i^\circ(x)$, which, in effect, states that the c_i-register in the configuration containing x is zero:

$$\neg \exists y (E_{12}(x,y) \wedge c_i(y)).$$

Using these formulas, we fix these register values for any d_1-configuration that is not a successor to be zero:

$$\forall x (d_1(x) \wedge \forall y (E_2(x,y) \rightarrow E_1(x,y)) \rightarrow c_1^\circ(x) \wedge c_2^\circ(x)).$$

We next define a formula $c_i^=(x,y)$ with the following property. Suppose b and b' are elements of configurations B and B', respectively, where B' is the successor of B: if the pair $\langle b, b' \rangle$ satisfies $c_i^=(x,y)$ then the c_i-counter of

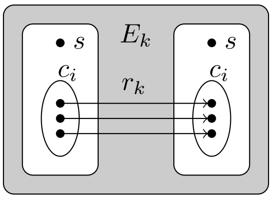


Fig. 9 Successive configurations whose elements satisfy the formula $c_i^=(x,y)$: the c_i counters are in 1–1 correspondence under r_k, and so are equinumerous.

B and the c_i-counter of B' contain the same value. To construct $c_i^=(x,y)$, we employ a pair of binary predicates r_1, r_2, denoting relations contained within the equivalences E_1, E_2, respectively, but disjoint from the other:

$$\bigwedge_{k=1}^{2} \forall x \forall y (r_k(x,y) \rightarrow E_k(x,y) \wedge \neg E_{3-k}(x,y)). \tag{33}$$

Recall that, under our assumptions concerning b and b', if b satisfies d_k then b and b' lie in a common E_k class. The formula $c_i^=(x,y)$ then simply states that, in that case, every element in the c_i-register of B is related by r_k to exactly one element in the c_i-register of B', and that every element in the c_i-register of B' is related by the inverse of r_k to exactly one element in the c_i-register of B.

$$\bigwedge_{k=1}^{2} (d_k(x) \rightarrow \forall y(E_{12}(x,y) \wedge c_i(y) \rightarrow \exists_{[=1]} x(r_k(y,x) \wedge c_i(x))) \wedge \\ \forall x(E_{12}(y,x) \wedge c_i(x) \rightarrow \exists_{[=1]} y(r_k(y,x)) \wedge c_i(y))).$$

Note how the variables x and y are 're-used' by quantifiers. This formula relies on the sentence (33) to have its advertised effect: the relation r_k holds only between elements in the same E_k-class but different E_{3-k}-classes. The situation is illustrated in Fig. 9.

Similarly, we can define a formula $c_i^+(x,y)$ entailing that, if the configuration B' containing y is the successor of the configuration B containing x, then the c_i-register of B' is one greater than that of B, and a formula $c_i^-(x,y)$ entailing that the c_i-register of B' is one less than that of B (or that both are zero).

Using the formulas $c_i^\circ(x)$, $c_i^+(x,y)$ and $c_i^-(x,y)$, we may then encode the program of $\mathbf{M}$ in the expected way. For example, if the basic operation of $\mathbf{M}$ associated with state s_i is to increment counter c_1 and move to state s_j, then we require:

$$\forall x \forall y (s_i(x) \wedge t(x,y) \rightarrow (c_1^+(x,y) \wedge c_2^=(x,y) \wedge s_j(y))).$$

Writing such formulas for all states s_i ($0 \leq i < L$), we can effectively construct a $\mathcal{C}^2$2E-formula $\varphi_{\mathbf{M}}$ any model of which contains a sequence of configurations $B_0, B_1, \ldots$, encoding the run of $\mathbf{M}$. Indeed, $\varphi_{\mathbf{M}}$ has a finite model if and only if $\mathbf{M}$ has a terminating run. Hence:

Theorem 7.1 *The finite satisfiability problem for $\mathcal{C}^2$2E is r.e.-complete.*

Bearing in mind that $\mathbf{M}$ terminates just in case its run encounters the stop state, we see that $\varphi_{\mathbf{M}} \wedge \forall x \neg s_L(x)$ has an (infinite) model if and only if $\mathbf{M}$ is non-terminating. Hence:

Theorem 7.2 *The satisfiability problem for $\mathcal{C}^2$2E is co-r.e.-complete.*

Table of symbols

The following table lists mathematical symbols used with the same meanings over several sections of the paper. Symbols with only local use are not listed here.

Symbol	Explanation	Pages
φ	a $\mathcal{C}^2$1E-formula in normal form	8
Z	a number exponentially large in $\lVert\varphi\rVert$	8
Σ	signature of φ with polynomially many spare predicates	8
π_i $(1 \leq i \leq I)$	the 1-types of Σ	9
ρ_j $(1 \leq j \leq 8J)$	the ray-types of Σ	9
$\mathfrak{A}$	a structure interpreting Σ, often a model of φ	12, 32
σ_k $(1 \leq k \leq K)$	star-types occurring in the structure $\mathfrak{A}$ or certificate $\mathfrak{C}$	13, 28
$B^g = C^g$ $(1 \leq g \leq G)$	a special equivalence class, special cluster	14, 18
$\mathcal{I}$	the 1-types uniquely realized in $\mathfrak{A}$	14
$\mathcal{G}_i$	the indices of the special equivalence classes realizing π_i	14
$\underline{w}^\dagger = \underline{w}^1, \ldots, \underline{w}^G$	profiles of special equivalence classes	18
C^h $(1 \leq h \leq H)$	a cluster	20
$\mathcal{E}^h$ $(G < h \leq H)$	equations satisfied by the profile of C^h	20
$\underline{\mathbf{w}}^h_\ell$ $(1 \leq \ell \leq L)$	a basis vector for solutions of $\mathcal{E}^h$	20
B^h_s $(0 \leq s < \mathfrak{c}(h))$	an equivalence class/galaxy in C^h ($\mathfrak{c}(h)$ may be $\aleph_0$)	20
$\dot{B}^h_p$ $(1 \leq p \leq \mathfrak{b}(h))$	a terminator in C^h ($\mathfrak{b}(h)$ may be 0)	23
$\hat{B}^h_p$ $(1 \leq p \leq \mathfrak{b}(h) + 1)$	a sector in C^h	23
$\dot{z}^h_{\ell,p}$ $(\underline{\dot{z}})$	number of occurrences of ℓth constellation in $\dot{B}^h_p$	24
$\hat{z}^h_{\ell,p}$ $(\underline{\hat{z}})$	number of occurrences of ℓth constellation in $\hat{B}^h_p$	24
$\mathcal{D}$	the equations $\mathcal{D}_1$–$\mathcal{D}_9$	26
$\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$	the equations and inequalities $\mathcal{E}_1$–$\mathcal{E}_6$, using $\mathcal{D}$ as definitions	26
$\underline{a}^\dagger, \underline{\dot{b}}, \underline{\hat{b}}$	a solution for $\mathcal{E}(\underline{w}^\dagger, \underline{\dot{z}}, \underline{\hat{z}})$	27
$\mathfrak{C}$	a certificate	27

Acknowledgements

This work is supported by the EPSRC, grant number EP/K017438/1. The author would like to thank Mr. Y. Guskov, Dr. E. Kieroński, Prof. L. Tendera and Prof. T. Tinchev for valuable comments and suggestions.

References

[1] M. Mortimer, On languages with two variables, Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **21**, 135–140 (1975).

[2] 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).

[3] E. Grädel, M. Otto, and E. Rosen, Two-variable logic with counting is decidable, in: Logic in Computer Science, (IEEE, 1997), pp. 306–317.

[4] L. Pacholski, W. Szwast, and L. Tendera, Complexity of two-variable logic with counting, in: Logic in Computer Science, (IEEE, 2007), pp. 318–327.

[5] I. Pratt-Hartmann, Complexity of the two-variable fragment with counting quantifiers, Journal of Logic, Language and Information **14**(3), 369–395 (2005).

[6] E. Kieroński, Results on the guarded fragment with equivalence or transitive relations, in: Computer Science Logic, edited by L. Ong, LNCS Vol. 3634 (Springer, 2005), pp. 309–324.

[7] E. Kieroński, J. Michalyszyn, I. Pratt-Hartmann, and L. Tendera, Two-variable first-order logic with equivalence closure, SIAM Journal on Computing **43**(3), 1012–1063 (2014).

[8] E. Kieroński and M. Otto, Small substructures and decidability issues for first-order logic with two variables., in: Logic in Computer Science, (IEEE, 2005), pp. 448–457.

[9] W. Szwast and L. Tendera, FO^2 with one transitive relation is decidable, in: Proc. of STACS, edited by N. Portier and T. Wilke, LIPIcs Vol. 20 (Schloß Dagstuhl - Leibniz-Zentrum für Informatik, 2013), pp. 317–328.

[10] Y. Kazakov, Saturation-based decision procedures for extensions of the guarded fragment, PhD thesis, Universität des Saarlandes, Saarbrücken, Germany, 2006.

[11] E. Kieroński and L. Tendera, On finite satisfiability of two-variable first-order logic with equivalence relations, in: Logic in Computer Science, (IEEE, 2009), pp. 123–132.

[12] Y. Kazakov, U. Sattler, and E. Zolin, How many legs do I have? Non-simple roles in number restrictions revisited, in: Proc. of LPAR, edited by N. Dershowitz and A. Voronkov (2007), pp. 303–317.

[13] L. Tendera, Counting in the two-variable guarded fragment with transitivity, in: Proc. of STACS, edited by V. Diekert and B. Durand, LNCS Vol. 3404 (Springer, Berlin, 2005), pp. 83–96.

[14] I. Horrocks, U. Sattler, and S. Tobies, Practical reasoning for very expressive description logics, Logic Journal of the IGPL **8**(3), 239–263 (2000).

[15] I. Horrocks and U. Sattler, A tableaux decision procedure for $\mathcal{SHOIQ}$, Journal of Automated Reasoning **39**(3), 249–276 (2007).

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

[17] E. Grädel, On the restraining power of guards, Journal of Symbolic Logic **64**, 1719–1742 (1999).

[18] Y. Kazakov, A polynomial translation from the two-variable guarded fragment with number restrictions to the guarded fragment, in: Proc. of JELIA, edited by J. J. Alferes and J. Leite, LNCS Vol. 3229 (Springer, Berlin, 2004), pp. 372–384.

[19] I. Pratt-Hartmann, Complexity of the guarded two-variable fragment with counting quantifiers, Journal of Logic and Computation **17**(1), 133–155.

[20] A. Janiczak, Undecidability of some simple formalized theories, Fundamenta Mathematicae **40**(1), 131–139 (1953).

[21] D. Scott, A decision method for validity of sentences in two variables, Journal of Symbolic Logic **27**, 477 (1962).

[22] L. Pottier, Minimal solutions of linear Diophantine systems : bounds and algorithms, in: Proc. Rewriting Techniques and Applications, edited by R. Book, LNCS Vol. 488 (Springer, Berlin, 1991), pp. 162–173.

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

[24] C. Papadimitriou, On the complexity of integer programming, Journal of the ACM **28**(4), 765–768 (1981).

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

[26] E. Börger, E. Grädel, and Y. Gurevich, The classical decision problem (Springer, Berlin, 1997).