



Computational complexity of atomic chemical reaction networks

David Doty¹ · Shaopeng Zhu²

Published online: 8 June 2018
© Springer Nature B.V. 2018

Abstract

Chemical reaction network has been a model of interest to both theoretical and applied computer scientists, and there has been concern about its physical-realisticity which calls for study on the *atomic* property of chemical reaction networks. Informally, a chemical reaction network is “atomic” if each reaction may be interpreted as the rearrangement of indivisible units of matter. There are several reasonable definitions formalizing this idea. We investigate the computational complexity of deciding whether a given network is atomic according to each of these definitions. *Primitive atomic*, which requires each reaction to preserve the total number of atoms, is shown to be equivalent to mass conservation. Since it is known that it can be decided in polynomial time whether a given chemical reaction network is mass-conserving (Mayr and Weihmann, in: International conference on applications and theory of petri nets and concurrency, Springer, New York, 2014), the equivalence we show gives an efficient algorithm to decide primitive atomicity. *Subset atomic* further requires all atoms be species, so intuitively this type of network is endowed with a “better” property than primitive atomic (i.e. mass conserving) ones in the sense that the atoms are not just abstract indivisible units, but also actual participants of reactions. We show that deciding if a network is subset atomic is in NP, and “whether a network is subset atomic with respect to a given atom set” is strongly NP-complete. *Reachably atomic*, studied by Adleman et al. (On the mathematics of the law of mass action, Springer, Dordrecht, 2014. https://doi.org/10.1007/978-94-017-9041-3_1), and Gopalkrishnan (2016), further requires that each species has a sequence of reactions splitting it into its constituent atoms. Using a combinatorial argument, we show that there is a polynomial-time algorithm to decide whether a given network is reachably atomic, improving upon the result of Adleman et al. that the problem is decidable. We show that the reachability problem for reachably atomic networks is PSPACE-complete. Finally, we demonstrate equivalence relationships between our definitions and some cases of an existing definition of atomicity due to Gnacadja (J Math Chem 49(10):2137, 2011).

Keywords Atomic · Chemical reaction networks · Computational complexity

This paper is based on Doty and Zhu (2018), which omitted many detailed proofs to various theorems and lemmas, such as the poly-time decidability of REACHABLY-ATOMIC. These proofs have been included in the current version of the paper. We have also corrected some errors and typos that appeared in the conference version.

✉ Shaopeng Zhu
szhu@terpmail.umd.edu

David Doty
doty@ucdavis.edu

¹ 2069 Academic Surge, University of California, Davis, Davis, USA

² University of Maryland, College Park, College Park, USA

1 Introduction

A *chemical reaction network* is a set of reactions such as $A + B \rightleftharpoons C$ and $X \rightarrow 2Y$, intended to model molecular species that interact, possibly combining or splitting in the process. For 150 years (Guldberg and Waage 1864), the model has been a popular language for describing natural chemicals that react in a well-mixed solution. It is known that in theory *any* set of reactions can be implemented by synthetic DNA complexes (Soloveichik et al. 2010). Synthetically equivalent to Petri nets (Angeli et al. 2007; Bridger et al. 2016; Esparza et al. 2017), chemical reaction networks are now equally appropriate as a *programming* language that can be compiled into real chemicals. With advances in synthetic biology heralding a new era of sophisticated biomolecular engineering (Chen et al. 2013; Montagne et al. 2011; Padirac et al. 2013; Qian et al. 2011;

Qian and Winfree 2011; Seelig et al. 2006; Srinivas 2015), chemical reaction networks are expected to gain prominence as a natural high-level language for designing molecular control circuitry.

There has been a flurry of recent progress in understanding the ability of chemical reaction networks to carry out computation: computing functions (Alistarh et al. 2017; Brijder et al. 2016; Chen et al. 2014, 2013, 2014; Cummings et al. 2015; Doty 2014; Doty and Hajiaghayi 2015; Esparza et al. 2017; Salehi et al. 2016, 2015; Soloveichik et al. 2008), as well as other computational tasks such as space- and energy-efficient search (Thachuk and Condon 2012), signal processing (Hua et al. 2013; Salehi et al. 2014), linear I/O systems (Oishi and Klavins 2011), machine learning (Napp and Adams 2013), and even identifying function computation in existing biological chemical reaction networks (Cardelli and Csikász-Nagy 2012). These studies generally assume that any set of reactions is permissible, but not all are physically realistic. Consider, for example, the reaction $X \rightarrow 2X$, which appears to violate the law of conservation of mass. Typically such a reaction is a shorthand for a more realistic reaction such as $F + X \rightarrow 2X$, where F is an anonymous and plentiful source of “fuel” providing the necessary matter for the reaction to occur. The behavior of the two is approximately equal only when the number of executions of $X \rightarrow 2X$ is far below the supplied amount of F , and if F runs out then the two reactions behave completely differently. Thus, although $X \rightarrow 2X$ may be implemented approximately, to truly understand the long-term behavior of the system requires studying its more realistic implementation $F + X \rightarrow 2X$. A straightforward generalization of this “realism” constraint is that each chemical species S may be assigned a mass $m(S) \in \mathbb{R}^+$, where in each reaction the total mass of the reactants equals that of the products. Indeed, *conservative* Petri nets formalize this very idea (Edmund 1976; Mayr and Weihmann 2014), and it is straightforward to decide algorithmically if a given network is conservative by reducing to a question of linear algebra.

The focus of this paper is a more stringent condition: that the network should be *atomic*, i.e., each reaction rearranges discrete, indivisible units (atoms), which may be of different noninterchangeable types.¹ (In contrast, mass conservation requires each reaction to rearrange a conserved quantity of continuous, generic “mass”.) We emphasize that this is not intended as a study of the atoms appearing in the periodic table of the elements. Instead, we

aim to model chemical systems whose reactions rearrange certain units, but never split, create, or destroy those units. For example, DNA strand displacement systems (Soloveichik et al. 2010; Yurke et al. 2000) have individual DNA strands as indivisible components, and each reaction merely rearranges the secondary structure among the strands (i.e., which bases on the strands are hybridized to others).

Contrary to the idea of mass conservation, there is no “obviously correct” definition of what it means for a chemical reaction network to be atomic, as we will discuss. Furthermore, at least two inequivalent definitions exist in the literature (Adleman et al. 2014; Gnacadja 2011). It is not the goal of this paper to identify a single correct definition. Instead, our goal is to evaluate the choices that must be made in formalizing a definition, to place existing and new definitions in this context to see how they relate to each other, and to study the computational complexity of deciding whether a given network is atomic. This is a step towards a more broad study of the computational abilities of “physically realistic” chemical reaction networks.

1.1 Summary of results and connection with existing work

The most directly related previous work is that of Adleman et al. (2014) and of Gnacadja (2011), which we now discuss in conjunction with our results.

We identify two fundamental questions to be made in formalizing a definition of an “atomic” chemical reaction network:

1. Are atoms also species? (For example, if the only reaction is $2\text{H}_2 + \text{O}_2 \rightleftharpoons 2\text{H}_2\text{O}$; then H and O are atoms but not species that appear in a reaction.)
2. Is each species separable into its constituent atoms via reactions?

A negative answer to (1) implies a negative answer to (2). (If some atom is not a species, then it cannot be the product of a reaction.) Thus there are three consistent answers to the above two questions: no/no, yes/no, and yes/yes. We respectively call these *primitive atomic*, *subset atomic*, and *reachably atomic*, defined formally in Sect. 3. Intuitively, a network is *primitive atomic* if each species may be interpreted as composed of one or more atoms, which themselves are not considered species (a species can be composed of just a single atom, but they will have different “names”). More formally, if \mathcal{A} is the set of species, there is a set \mathcal{A} of atoms, such that each species $S \in \mathcal{A}$ has an *atomic decomposition* $\mathbf{d}_S \in \mathbb{N}^{\mathcal{A}} \setminus \{\mathbf{0}\}$ describing the atoms that constitute S , such that each reaction preserves the atoms. A network is *subset atomic* if it is primitive

¹ This usage of the term “atomic” is different from its usage in traditional areas like operating system or syntactic analysis, where an “atomic” execution is an uninterruptable unit of operation (Silberschatz et al. 2013).

atomic and the atoms are themselves considered species; i.e., if $\Delta \subseteq \mathcal{A}$. A network is *reachably atomic* if it is subset atomic, and furthermore, for each species $S \in \mathcal{A}$, there is a sequence of reactions, starting with a single copy of S , resulting in a configuration consisting only of atoms. (If each reaction conserves the atomic count, then this configuration must be unique and equal to the atomic decomposition of S .)

A long-standing open problem in the theory of chemical reaction networks is the global attractor conjecture (Craciun et al. 2009; Horn 1974), of which even the following special case remains open: is every network satisfying detailed balance *persistence*, i.e., if started with all positive concentrations, do concentrations stay bounded away from 0? Adleman et al. (2014) defined reachably atomic chemical reaction networks and proved the global attractor conjecture holds for such networks. Gnacadja (2011), attacking similar goals, defined a notion of atomicity called “species decomposition” and showed a similar result. We establish links between our definitions and those of both Gnacadja (2011) and Adleman et al. (2014) in Section 8(C) of Doty and Zhu (2017). In particular, Adleman et al. (2014) showed that it is decidable whether a given network is reachably atomic. This is not obvious since the condition of a species being separable into its constituent atoms via reactions appears to require an unbounded search. We improve this result, showing it is decidable in polynomial time.

Mayr and Weihmann (2014) proved that configuration reachability graphs for mass conserving chemical reaction networks (i.e., conservative Petri nets) are at most exponentially large in the size of the binary representation of the network, implying via Savitch’s theorem (Walter 1970) a polynomial-space algorithm for deciding reachability in mass-conserving networks. We use these results in analyzing the complexity of reachability problems in reachably atomic chemical reaction networks in Sect. 6.

It is clear that any reasonable definition of atomicity should imply mass conservation: simply assign all atoms to have mass 1, noting that any reaction preserving the atoms necessarily preserves their total count. Perhaps surprisingly, the conditions of primitive atomic and mass-conserving are in fact equivalent, so it is decidable in polynomial time whether a network is primitive atomic and what is an atomic decomposition for each species. A key technical tool is Chubanov’s algorithm (Chubanov 2015) for finding exact rational solutions to systems of linear equations with a strict positivity constraint.

In their work on autocatalysis of reaction networks (Deshpande and Gopalkrishnan 2013), Abhishek and Manoj showed that a consistent reaction network is self-replicable if and only if it is critical. Since weak-reversibility implies consistency and our definition of

reversibility implies weak-reversibility defined in Deshpande and Gopalkrishnan (2013), we obtain the following equivalence: let a chemical reaction network \mathcal{C} be reversible. Then \mathcal{C} is mass conserving if and only if there does not exist $\mathbf{c}_1 < \mathbf{c}_2 \in \mathbb{N}^{\mathcal{A}}$ such that $\mathbf{c}_1 \Rightarrow^* \mathbf{c}_2$.²

Lastly, we note that there have been other models addressing different aspects of atomicity (not necessarily using the term “atomic”). They focus on features of chemical reaction networks not modeled in this paper. For discussions on these works, please see Section 8(A) of Doty and Zhu (2017).

2 Preliminaries

Let $\mathbb{Z}, \mathbb{N}, \mathbb{R}$ respectively denote the set of integers, non-negative integers, and reals. Let \mathcal{A} be a finite set. We write $\mathbb{N}^{\mathcal{A}}$ to denote $\{f : \mathcal{A} \rightarrow \mathbb{N}\}$. Equivalently, by assuming a “canonical” ordering on \mathcal{A} , an element $\mathbf{c} \in \mathbb{N}^{\mathcal{A}}$ can also be viewed as a $|\mathcal{A}|$ -dimensional vector of natural numbers, with each coordinate labeled by $S \in \mathcal{A}$ interpreted as the count of S . $\mathbf{c} \in \mathbb{N}^{\mathcal{A}}$ interpreted this way is called a *configuration*. We sometimes use multiset notation, e.g., $\{3A, 2B\}$ to denote the configuration with 3 copies of A , 2 of B , and 0 of all other species. $\mathbb{Z}^{\mathcal{A}}, \mathbb{R}^{\mathcal{A}}, \mathbb{N}^{\mathcal{A} \times \mathcal{A}}, \mathbb{N}^{\mathcal{A}}$ (where \mathcal{A} is also a finite set) are defined analogously.

We write $\mathbf{c} \leq \mathbf{c}'$ to denote that $(\forall X \in \mathcal{A}) \mathbf{c}(X) \leq \mathbf{c}'(X)$, and $\mathbf{c} < \mathbf{c}'$ if $\mathbf{c} \leq \mathbf{c}'$ and $\mathbf{c} \neq \mathbf{c}'$. We say \mathbf{c} and \mathbf{c}' are *incomparable* if $\mathbf{c} \not\leq \mathbf{c}'$ and $\mathbf{c}' \not\leq \mathbf{c}$.

Definition 1 Given a finite set of chemical species \mathcal{A} , a *reaction* over \mathcal{A} is a pair $\alpha = (\mathbf{r}, \mathbf{p}) \in \mathbb{N}^{\mathcal{A}} \times \mathbb{N}^{\mathcal{A}}$, specifying the stoichiometry of the reactants and products respectively.³

A *chemical reaction network* is a pair $\mathcal{C} = (\mathcal{A}, R)$, where \mathcal{A} is a finite set of chemical species, and R is a finite set of reactions over \mathcal{A} .

A chemical reaction network is *reversible* if $(\forall (\mathbf{r}, \mathbf{p}) \in R) (\mathbf{p}, \mathbf{r}) \in R$.

For configurations $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{N}^{\mathcal{A}}$, we write $\mathbf{c}_1 \Rightarrow_{\mathcal{C}}^* \mathbf{c}_2$ (read “ \mathcal{C} reaches \mathbf{c}_2 from \mathbf{c}_1 ”) if there exists a finite reaction sequence (including the empty sequence) that starts with \mathbf{c}_1

² Note that by the argument above, reversible networks are weakly reversible and hence consistent, which establishes the equivalence between self-replicability and criticality. One may further observe from Deshpande and Gopalkrishnan (2013) that this equivalence translates to the equivalence between the mass conserving property and the reachability property above.

³ There is typically a positive real-valued *rate constant* associated to each reaction, but we ignore reaction rates in this paper and consequently simplify the definition.

and ends with \mathbf{c}_2 . For simplicity, write $\mathbf{c}_1 \Rightarrow^* \mathbf{c}_2$ (read “ \mathbf{c}_2 is reachable from \mathbf{c}_1 ”) when \mathcal{C} is clear.

Definition 2 Given $\mathbf{c} \in \mathbb{N}^A$ (or $\mathbb{Z}^A, \mathbb{R}^A$ etc. analogously), the *support* of \mathbf{c} , written as $[\mathbf{c}]$, is the set $\{S \in A \mid \mathbf{c}(S) \neq 0\}$.

A few more notational conventions are listed here: write $\mathbf{e}_A \in \mathbb{N}^A$ as the unit vector that has count 1 on $A \in A$ and 0 on everything else. Given a vector $\mathbf{x} \in \mathbb{N}^A$, write $\|\mathbf{x}\| = \|\mathbf{x}\|_1 = \sum_{S \in A} \mathbf{x}(S)$. When \cdot is any data, write $\langle \cdot \rangle$ for its binary representation as a string, so $|\langle \cdot \rangle|$ is the length of the binary representation of \cdot . Given $f : A \rightarrow B$ and $C \subseteq A$, $f|_C$ is the function $C \rightarrow B$, $c \mapsto f(c)$ ($\forall c \in C$). Lastly, when \mathbf{M} is a matrix, write \mathbf{M}^T as its transposition.

3 Definitions of “atomic”

This section addresses definitions of several classes of networks, some computational complexity result of which will be exhibited later.

Intuitively, $\mathcal{C} = (A, R)$ is primitive atomic if all species can be decomposed into combinations of some atoms. Atoms are not required to be species. Each reaction conserves the total count of each type of atom in the species involved (i.e., the reaction can only rearrange atoms but not create or destroy them).

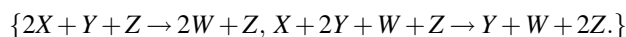
Note that the purpose of studying the primitive-atomic model (as well as all other types of atomic later) is not to analyze “real-world” atoms. Instead, we are trying to study how molecules can be interpreted as decomposable into *exchangeable parts*. In particular, if we know only the reactions but not those exchangeable parts, we are interested in whether the reactions can tell us how the molecules are composed from parts. Proposition 10 below, for example, shows that this information can be retrieved by finding a mass distribution vector.

Definition 3 (*primitive atomic*) Let Δ be a nonempty finite set and $\mathcal{C} = (A, R)$ a chemical reaction network. \mathcal{C} is *primitive atomic with respect to Δ* if for all $S \in A$, there is $\mathbf{d}_S \in \mathbb{N}^A \setminus \{\mathbf{0}\}$ such that

1. $(\forall (\mathbf{r}, \mathbf{p}) \in R)(\forall A \in \Delta) \sum_{S \in A} \mathbf{r}(S) \cdot \mathbf{d}_S(A) = \sum_{S \in A} \mathbf{p}(S) \cdot \mathbf{d}_S(A)$ (reactions preserve atoms), and
2. $(\forall A \in \Delta)(\exists S \in A) \mathbf{d}_S(A) \neq 0$. (each atom appears in the decomposition of some species)

For $S \in A$, call \mathbf{d}_S in Condition (1) the (atomic) *decomposition* of S . We say \mathcal{C} is *primitive atomic* if there is a nonempty finite set Δ such that \mathcal{C} is primitive atomic with respect to Δ . In the cases above, Δ is called the *set of atoms*.

Condition (1) embodies the intuition above. Condition (2) prescribes that each atom appears in the decomposition of at least one species. Consider the network $\mathcal{C} = (\{X, Y, W, Z\}, \{((2, 1, 0, 1)^T, (0, 0, 2, 1)^T), ((1, 2, 1, 1)^T, (0, 1, 1, 2)^T)\}$. One may write \mathcal{C} as:



\mathcal{C} is primitive-atomic with respect to, say, $\Delta = \{H, O\}$, via the decomposition vector $\mathbf{d}_X = (2, 0)^T$, $\mathbf{d}_Y = (0, 2)^T$, $\mathbf{d}_W = (2, 1)^T$, $\mathbf{d}_Z = (2, 2)^T$. Here $\mathbf{d}_X = (2, 0)^T$ means the species X is composed of 2 units of atom H and 0 unit of atom O , and $\mathbf{d}_Y, \mathbf{d}_W, \mathbf{d}_Z$ can be interpreted likewise. Observe that each of the two reactions in \mathcal{C} preserves the total count of each type of atom on both sides of reactions.

Next, we introduce the definitions of stoichiometric matrix and decomposition matrix. In particular, \mathbf{A} encodes the net change of species caused by execution of one reaction, and \mathbf{D} compiles all decomposition vectors into one data structure.

Definition 4 (*stoichiometric matrix*) The stoichiometric matrix \mathbf{A} for a chemical reaction network $\mathcal{C} = (R, A)$ is the $|R| \times |A|$ matrix where the entry $\mathbf{A}_{(\mathbf{r}, \mathbf{p}), S} = \mathbf{p}(S) - \mathbf{r}(S)$ for each $(\mathbf{r}, \mathbf{p}) \in R$ and $S \in A$.

Notation-wise, $\mathbf{A}_{(\mathbf{r}, \mathbf{p}), S}$ is the entry whose row is labeled by the reaction (\mathbf{r}, \mathbf{p}) and column by the species S . Each row of the stoichiometric matrix represents the change of count of each species via execution of 1 unit of (\mathbf{r}, \mathbf{p}) . For more illustration, see Example D.1 (Doty and Zhu 2017).

Definition 5 (*decomposition matrix*) Let $\mathcal{C} = (A, R)$ be primitive atomic with respect to Δ . The decomposition matrix, denoted as \mathbf{D}_Δ for \mathcal{C} with respect to Δ is the $|A| \times |\Delta|$ matrix whose row vectors are $(\mathbf{d}_S)^T$ ($S \in A$).

Note that the set of decomposition vectors $\{\mathbf{d}_S\}_{S \in A}$ is in general not unique for primitive atomic chemical reaction networks—for example, $A + B \rightarrow C$ is primitive atomic with respect to $\Delta = \{D\}$ via $(k, k, 2k)(\forall k \in \mathbb{N}_{>0})$. Correspondingly, \mathbf{D}_Δ 's are defined with respect to each set $\{\mathbf{d}_S\}_{S \in A}$. See Example D.2 and Remark D.2 (Doty and Zhu 2017) for more discussion on decomposition matrices.

The next definition requires all atoms to be species.

Definition 6 (*subset atomic*) Let $\mathcal{C} = (A, R)$ be a chemical reaction network and let $\Delta \subseteq A$ be nonempty. We say that \mathcal{C} is *subset- Δ -atomic* if \mathcal{C} is primitive atomic with respect to Δ and, for each $S \in A$:

1. $S \in A \cap \Delta = \Delta \implies \mathbf{d}_S = \{S\}$, and
2. $S \in A \setminus \Delta \implies \|\mathbf{d}_S\| \geq 2$.

We say \mathcal{C} is *subset atomic* if $\exists \emptyset \neq \Delta \subseteq A$ such that \mathcal{C} is subset- Δ -atomic.

By Definition 6, no two *atoms* can have the same atomic decomposition, but it is allowed that two distinct *molecular* (i.e. non-atom) species to have the same decomposition. In this case we say the two species are *isomers* (reminiscent of isomers in nature that are composed of the same atoms in different geometrical arrangements). As for the requirement that each non-atom species decompose to a vector of size at least 2, that is to incorporate the idea that generally a molecule should be composed of at least 2 atoms.

For example, the network $C = \{2X + Y + Z \rightarrow 2W + Z, X + 2Y + W + Z \rightarrow Y + W + 2Z\}$ mentioned above is subset-atomic: just redefine $\Delta = \{X\}$ and $\mathbf{d}_X = (1)$, $\mathbf{d}_Y = (2)$, $\mathbf{d}_Z = (3)$, $\mathbf{d}_W = (2)$. One may verify that in the first reaction, each side has 7 atoms X , while in the second each side has 10.

The next definition further requires that decomposition of each molecular species S_i can be “realized” via a sequence of reactions, given $\{1S_i\}$ as initial state. As discussed in Sect. 1.1, this definition was originally developed in Adleman et al. (2014) to help their approach to the Global Attractor Conjecture in the field of mass action kinetics. Considering the convention for most networks, we relax their requirement of reversibility for each reaction.

Definition 7 (*reachably atomic*) A chemical reaction network $C = (A, R)$ is *reachably atomic* if

1. C is subset atomic with respect to some $\Delta \subseteq A$, and
2. for each $S \in A \setminus \Delta$, $\{1S\} \Rightarrow^* \mathbf{d}_S$.

Here and wherever necessary, with slight abuse of notation, \mathbf{d}_S , which represents the atomic decomposition of S , simultaneously represents a configuration in \mathbb{N}^A reachable from $\{1S\}$. Observe that $C = \{2X + Y + Z \rightarrow 2W + Z, X + 2Y + W + Z \rightarrow Y + W + 2Z\}$ is not reachably-atomic unless we add the following reactions: $Y \rightarrow 2X, Z \rightarrow 3X, W \rightarrow 2X$, in which case the set of atoms is $\{X\}$.

Condition 2 is a strong restriction ensuring some nice properties. For example, the atom set of a reachably atomic network is unique:

Lemma 8 *If $C = (A, R)$ is reachably atomic, then the choice of Δ with respect to which C is reachably atomic is unique. Moreover, for each $S \in A$, \mathbf{d}_S is unique, i.e., if $\{1S\} \Rightarrow^* \mathbf{c} \in \mathbb{N}^A$, then $\mathbf{c} = \mathbf{d}_S$.*

Proof The intuition is to show that should there exist $\Delta_1 \neq \Delta_2$ and without loss of generality, assume $\exists A \in \Delta_1 \setminus \Delta_2$, then the decomposition of A with respect to Δ_1 violates the preservation of atoms in Δ_2 .

Assume for the sake of contradiction that for some reachably atomic network C , there exist $\Delta_1 \neq \Delta_2$ with respect to both of which C is reachably atomic, respectively

via decomposition matrices \mathbf{D}_1 and \mathbf{D}_2 . Note that $(\Delta_1 \setminus \Delta_2) \cup (\Delta_2 \setminus \Delta_1) \neq \emptyset$. Take $A \in (\Delta_1 \setminus \Delta_2) \cup (\Delta_2 \setminus \Delta_1)$:

1. If $A \in \Delta_1 \setminus \Delta_2$, then $\{1A\}$ is decomposable into some $\mathbf{c} \in \mathbb{N}^A \mid [\mathbf{c}] \subseteq \Delta_2$ via a sequence of reactions, with $\|\mathbf{c}\|_1 \geq 2$. There is no way for this sequence of reactions to preserve atoms with respect to Δ_1 , for the initial configuration has a single atom $A \in \Delta_1$ while the final configuration has no atom A .
2. Similarly, if $A \in \Delta_2 \setminus \Delta_1$, there will be a sequence of reactions violating preservation of atoms with respect to Δ_2 .

We next prove the uniqueness of decomposition vectors \mathbf{d}_S for all $S \in A$, or equivalently, the uniqueness of decomposition matrix \mathbf{D} . Suppose not, then there exists $S \in A \setminus \Delta$ s.t. $\{1S\} \Rightarrow^* \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{N}^A$, $\mathbf{y}_1 \neq \mathbf{y}_2$ and $[\mathbf{y}_1], [\mathbf{y}_2] \subseteq \Delta$, via reaction sequences $\mathbf{s}_1, \mathbf{s}_2$ respectively. Assume without loss of generality that \mathbf{s}_1 preserves the number of atoms in each reaction, which means $\mathbf{y}_1 = \mathbf{d}_S$. Then there must be one or more actions in \mathbf{s}_2 that does (do) not preserve the number of atoms, for $\mathbf{s}_1, \mathbf{s}_2$ share the same initial configuration $\{1S_i\}$ yet reach different final count of atoms, while no atoms are allowed to be isomeric to each other. \square

Conservation laws in “-atomic” networks reminds us of a more familiar type of conservation law, which is mass conservation. The next section exhibits some observations on the relationship between these two types of conservation laws.

4 Mass-conservation and primitive atomicity

This section shows that “primitive atomic” and “mass conserving” are equivalent concepts. We first formalize what it means for a network to conserve mass:

Definition 9 (*mass conserving*) A chemical reaction network $C = (A, R)$ is *mass conserving* if

$$(\exists \mathbf{m} \in \mathbb{R}_{>0}^A)(\forall (\mathbf{r}, \mathbf{p}) \in R) \sum_{S \in A} \mathbf{r}(S) \cdot \mathbf{m}(S) = \sum_{S \in A} \mathbf{p}(S) \cdot \mathbf{m}(S)$$

Equivalently, if \mathbf{A} is the stoichiometric matrix in Definition 4, then C is mass conserving if $(\exists \mathbf{m} \in \mathbb{R}_{>0}^A) \mathbf{A} \cdot \mathbf{m} = \mathbf{0}$. We call \mathbf{m} a *mass distribution vector*.

Using our familiar example, $C = \{2X + Y + Z \rightarrow 2W + Z, X + 2Y + W + Z \rightarrow Y + W + 2Z\}$ is mass conserving with respect to $\mathbf{m} = (0.5, 1, 1.5, 1)^T$. “Mass Conserving” captures the feature that for every reaction in C , the total mass of reactants are equal to the total mass of products. Difference between the definitions of Mass Conserving and

Primitive Atomic (as well as all “-atomic” definitions descended therefrom) become clear if we compare the matrix form of their respective conservation laws: mass conservation requires a single conservation relation ($\mathbf{A} \cdot \mathbf{m} = 0^{|R|}$), while primitive atomicity requires $|\Delta|$ of them ($\mathbf{A} \cdot \mathbf{D} = \mathbf{0}$ where \mathbf{D} is a $|\Delta| \times |\Delta|$ matrix).

However, apparently these two conservation laws are closely related. In fact, the freedom of defining Δ independent of \mathcal{A} provides us a choice for making Δ a singleton, which enables us to prove the following equivalence:

Proposition 10 *For any network \mathcal{C} , \mathcal{C} is primitive atomic $\Leftrightarrow \mathcal{C}$ is mass conserving. Further, there exists an $O(|\langle \mathbf{A} \rangle|^5)$ algorithm to decide if \mathcal{C} is primitive atomic, with \mathbf{A} the stoichiometric matrix of \mathcal{C} .*

Proof Intuitively, the “ \Rightarrow ” direction is shown by assigning mass 1 to each atom, as “homogenizing” the atoms preserves the original conservation law; for the “ \Leftarrow ” direction, one may essentially create a Δ of cardinality 1 with respect to which the network is primitive atomic by assigning to each molecule the count of that atom obtained from “de-fraction-izing” its mass. Time complexity of the decision algorithm for primitive atomicity follows from the time complexity of Chubanov’s algorithm (Chubanov 2015; Roos 2015) which can be used to decide mass conserving property (and give a rational mass-assignment vector in the “yes” cases), and from the equivalence established above. The proof also reflects the difference in number of conservation relations addressed two paragraphs above. See Section 8(B) (Doty and Zhu 2017) for details and more remarks. \square

Recall that subset atomicity imposes the restriction that $\Delta \subseteq \mathcal{A}$. As we’ll show in the following section, this single restriction increases the computational complexity of the decision problem “is a network ‘(prefix)-atomic’”.

5 Complexity of subset atomic

We shall determine in this chapter the computational complexity for deciding the subset atomicity of networks. First, we define the relevant languages:

Definition 11 We define the following languages:

$$\text{SUBSET-ATOMIC} = \{ \langle \mathcal{A}, R \rangle \mid (\exists \Delta \subseteq \mathcal{A})(\mathcal{A}, R) \text{ is subset atomic with respect to } \Delta \};$$

$$\text{SUBSET-FIXED-ATOMIC} = \{ \langle \mathcal{A}, R, \Delta \rangle \mid (\mathcal{A}, R) \text{ is subset atomic with respect to } \Delta \}$$

By definition, SUBSET-ATOMIC is the language whose elements are the encoding of a *subset atomic* chemical reaction network. SUBSET-FIXED-ATOMIC, on the other hand, is the language consisting of the encoding of a (network, atom set) pair where the network is *subset atomic* with respect to the given atom set. In this section we determine the complexity classes of these languages.

5.1 SUBSET-FIXED-ATOMIC and SUBSET-ATOMIC are in NP

It is not immediately obvious that there exists a short witness for either language (which if true would imply that both languages are in NP immediately), so we reduce SUBSET-FIXED-ATOMIC to INTEGER-PROGRAMMING, which is in NP (Papadimitriou Christos 1981).

Hereinafter, the notation \leq_m^p denotes “many-one reduces to”.

Proposition 12 $\text{SUBSET-FIXED-ATOMIC} \leq_m^p \text{INTEGER-PROGRAMMING}$ (hereinafter, “IP”).

Proof The proof is done by exhibiting a polynomial time algorithm to transition the conditions in Definition 6 into a linear system. Note that the atom set \mathcal{A} is given as input. For details of the reduction, see Section 8(B) (Doty and Zhu 2017). \square

Corollary 13 $\text{SUBSET-FIXED-ATOMIC}, \text{SUBSET-ATOMIC} \in \text{NP}$.

Proof It is proved (e.g., in Papadimitriou Christos 1981) that $\text{IP} \in \text{NP}$, hence so is SUBSET-FIXED-ATOMIC.

The proof that $\text{SUBSET-ATOMIC} \in \text{NP}$ is given by a polynomial time verification algorithm using the polynomial-time verifier of SUBSET-FIXED-ATOMIC as an oracle and taking as witness both the atom set and decomposition matrix. For details, see Section B (Doty and Zhu 2017). \square

5.2 Subset-fixed-atomic is NP-hard

Our proof shall be based on reduction from MONOTONE-1-IN-3-SAT. Recall that a monotone 3-CNF C is a conjunctive normal form with no negations, and a 1-in-3 satisfying assignment for C is an assignment of Boolean values to all variables such that for each clause in C , exactly one variable is assigned true.

As a well-established result, the following language is NP-complete (Garey and Johnson 1979).

$$\text{MONOTONE-1-IN-3-SAT} = \{ \langle V, C \rangle \mid C \text{ is a monotone 3-CNF over } V = \{v_i\}_{i=1}^n, \text{ and there exists a 1-in-3 satisfying assignment for } C \}$$

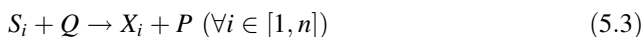
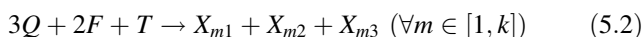
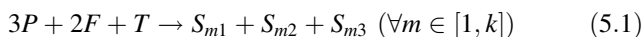
Proposition 14 MONOTONE-1-IN-3-SAT \leq_m^p SUBSET-FIXED-ATOMIC.

Proof Given an instance $\langle V, C \rangle$, we design a chemical reaction network \mathcal{C} where

1. Each molecular species consists of 2 atoms T and F (representing “True” and “False” respectively), and
2. reactions guarantee the equivalence: \mathcal{C} is subset- Δ -atomic if and only if the $\langle V, C \rangle \in$ MONOTONE-1-IN-3-SAT.

For each instance $\langle V, C \rangle$ of MONOTONE-1-IN-3-SAT, let $\Delta = \{T, F, P, Q\}$, $A = \{S_1, S_2, \dots, S_n, X_1, X_2, \dots, X_n\} \cup \Delta$.

To construct R , we denote $C = c_1 \wedge c_2 \wedge \dots \wedge c_k$. For the i -th clause $c_i \in C$, let v_{ij} denote the j -th literal of c . Same indexing convention applies for $\{S_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$, hence each S_{ij} (resp. X_{ij}) denotes an element in $\{S_i\}_{i=1}^n$ (resp. $\{X_i\}_{i=1}^n$).⁴ The set R contains the following reactions⁵:



Note that we only need to construct $4 + 2n$ species and $2k + n$ reactions whose coefficients are constant, so this transformation is polynomial in time and space. We argue that $\langle V, C \rangle \in$ MONOTONE-1-IN-3-SAT if and only if $\langle A, R, \Delta \rangle \in$ SUBSET-FIXED-ATOMIC.

\Rightarrow : If $\langle V, C \rangle \in$ MONOTONE-1-IN-3-SAT, there exists a $\phi : V \rightarrow \{0, 1\}$ under which $\exists(n_1, n_2, \dots, n_q) \subseteq (1, 2, \dots, n)$ s.t. $\phi(v_{ni}) = 1 \ (\forall i \in [1, q])$, $\phi(v_j) = 0 \ (j \in ([1, n] \setminus (n_1, n_2, \dots, n_q)))$, and for each $c_i \in C$, exactly one in three of its literals evaluates to 1. Let:

$$A_{TP} = \{S_{n_j} \mid j \in [1, q]\}$$

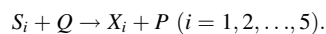
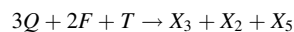
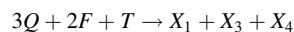
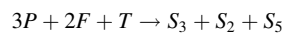
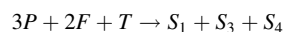
$$A_{FP} = A \setminus (\Delta \cup A_{TP} \cup \{X_i \mid i \in [1, n]\})$$

$$A_{TQ} = \{X_{n_j} \mid j \in [1, q]\}$$

$$A_{FQ} = A \setminus (\Delta \cup A_{TQ} \cup A_{TP} \cup A_{FP})$$

⁴ For example, for $V = \{v_1, v_2, \dots, v_5\}$, $C = (v_1 \vee v_3 \vee v_4) \wedge (v_3 \vee v_2 \vee v_5)$, $v_{11} = v_1, v_{12} = v_3, v_{13} = v_4, \dots, v_{23} = v_5$. Correspondingly, $S_{11} = S_1, X_{11} = X_1, \dots, S_{23} = S_5, X_{23} = X_5$.

⁵ To continue the example in the previous footnote, the set of reactions shall be:



Then $\langle A, R, \Delta \rangle \in$ SUBSET-FIXED-ATOMIC because with all atoms listed in the order: $\{T, F, P, Q\}$, one may make the following decomposition:

$$\mathbf{d}_U = (1, 0, 1, 0)^T, \quad \forall U \in A_{TP}$$

$$\mathbf{d}_V = (0, 1, 1, 0)^T, \quad \forall V \in A_{FP}$$

$$\mathbf{d}_W = (1, 0, 0, 1)^T, \quad \forall W \in A_{TQ}$$

$$\mathbf{d}_Z = (0, 1, 0, 1)^T, \quad \forall Z \in A_{FQ}$$

Because of the way $\{n_j\}_{j=1}^q$ was taken, for each reaction in (5.1), exactly one of the product species decompose to $1T$ and $1P$, with the other two decomposing to $1F$ and $1P$. Similar argument applies for reactions in (5.2), substituting X_i for S_i and Q for P . Arithmetics show that all three reactions (5.1) through (5.3) preserve the number of atoms, each atom appears in the decomposition of at least one molecular species, and the number of atoms in the decomposition of each species complies with the Definition 11. Therefore $\langle A, R, \Delta \rangle \in$ SUBSET-FIXED-ATOMIC.

\Leftarrow : If $\langle A, R, \Delta \rangle \in$ SUBSET-FIXED-ATOMIC, (5.1) ensures that each molecular species contains exactly two atoms. That is because for each $i \in [1, n]$,

$$\mathbf{d}_{S_i}(T) + \mathbf{d}_{S_i}(F) + \mathbf{d}_{S_i}(P) + \mathbf{d}_{S_i}(Q) \geq 2 \quad (5.4)$$

by Definition 11, so for each $m \in [1, k]$,

$$\sum_{j=1}^3 (\mathbf{d}_{S_{mj}}(T) + \mathbf{d}_{S_{mj}}(F) + \mathbf{d}_{S_{mj}}(P) + \mathbf{d}_{S_{mj}}(Q)) \geq 3 \times 2 = 6 \quad (5.5)$$

However the total number of atoms on the left hand side of 5.1 is exactly 6, meaning the equal sign has to be taken everywhere in (5.5) for any $m \in [1, k]$, forcing (5.4) to take equal sign as well.

Similarly, (5.2) ensures $\|\mathbf{d}_{X_i}\|_1 = 2$ for each $i \in [1, n]$.

The reaction series (5.3) implies that each S_i has at least one P and each X_i has at least one Q . Furthermore,

1. if any S_i decomposes to $2P$, its corresponding X_i shall be composed of PQ , contradicting (5.2) which says that no X_i contains any P ;
2. if any S_i decomposes to PQ , it contradicts with (5.1) which says that no S_i contains any Q .

Therefore all S_i shall decompose to either $\{1F, 1P\}$ ($(0, 1, 1, 0)^T$) or $\{1T, 1P\}$ ($(1, 0, 1, 0)^T$). Correspondingly, X_i decompose to either $(0, 1, 0, 1)^T$ or $(1, 0, 0, 1)^T$.

Construct ϕ such that $\phi(v_j) = 1$ for all $v_j \in \{v_j \mid \mathbf{d}_{S_j} = (1, 0, 1, 0)^T\}$, and $\phi(v_m) = 0$ for all $v_m \in V \setminus \{v_j \mid \mathbf{d}_{S_j} = (1, 0, 1, 0)^T\}$. Exactly one in three of the products in the right hand side of (5.1) decomposes to $(1, 0, 1, 0)$, so exactly one in three of the variables

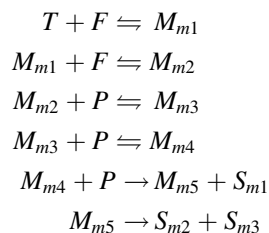
(literals) in each clause of C evaluates to 1. It follows that $\langle V, C \rangle \in \text{MONOTONE-1-IN-3-SAT}$. \square

We notice that the coefficients of all species in all the reactions (5.1)–(5.3) are constants, so the numerical parameters—entries of each $(\mathbf{r}, \mathbf{p}) \in R$ —of the instances $\langle A, R, \Delta \rangle$ constructed above are bounded by the constant 3, which is again bounded by a polynomial of the length of $\langle V, C \rangle$, presuming the encoding scheme is “reasonable and concise” (Garey and Johnson 1978). We therefore conclude that:

Corollary 15 $\text{SUBSET-FIXED-ATOMIC}$ is strongly NP-hard (and hence strongly NP-complete).

Remark 16 $\text{SUBSET-FIXED-ATOMIC}$ remains NP-complete even restricted to instances where R contains only unimolecular and bimolecular reactions.

In fact, one may convert, for each $m \in [1, k]$ (recall that k is the number of clauses in C), any reaction in the series (5.1) $(3P + 2F + T \rightarrow S_{m1} + S_{m2} + S_{m3} (\forall m \in [1, k]))$ into the following series:



And apply similar methods to the X_i species. Such conversion creates $2 \times 10k = 20k$ extra reactions and $2 \times 5k = 10k$ intermediate species, which is polynomial in both time and space.

The lower bound of the complexity of SUBSET-ATOMIC therefore remains open, but we conjecture that SUBSET-ATOMIC is NP-hard (hence NP-complete).

6 Complexity of reachably atomic

Without repeating the intuition of the definition of reachably atomic which has been explained in Sects. 1.1 and 3, we proceed with the corresponding definition of languages for deciding reachable atomicity and the reachability problem in reachably atomic networks.

Definition 17 We define the following languages

$$\begin{aligned} \text{REACHABLY-ATOMIC} &= \{ \langle A, R \rangle \mid (\exists \Delta \subseteq A) \\ &\quad \langle A, R \rangle \text{ is reachably} \\ &\quad \text{atomic with respect} \\ &\quad \text{to } \Delta \}; \end{aligned}$$

$$\begin{aligned} \text{REACHABLY-FIXED-ATOMIC} &= \{ \langle A, R, \Delta \rangle \mid \langle A, R \rangle \text{ is} \\ &\quad \text{reachably atomic with} \\ &\quad \text{respect to } \Delta \} \end{aligned}$$

Distinction between $\text{REACHABLY-FIXED-ATOMIC}$ and REACHABLY-ATOMIC is analogous to “ $\text{SUBSET-FIXED-ATOMIC}$ v.s. SUBSET-ATOMIC ”. However, by Lemma 8 there is no semantic reason to distinguish between “ $\text{REACHABLY-REACHABLY ATOMIC}$ ” and “ REACHABLY-ATOMIC ”. So we shall only consider REACHABLY-ATOMIC from now on.

6.1 REACHABLY-ATOMIC is in P

As mentioned before, the requirement that $\{1S\} \Rightarrow^* \mathbf{d}_S (\forall S \in A)$ ensures some interesting results. The complexity results in this subsection confirm this.

Lemma 18 If a network $\mathcal{C} = (A, R)$ is reachably atomic with respect to Δ via decomposition matrix \mathbf{D} (or equivalently, via the set of decomposition vectors $\{\mathbf{d}_S\}_{S \in A}$), then $\exists S \in A \setminus \Delta$ and $(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\mathbf{r} = \{1S\}$ and $\mathbf{p} = \mathbf{d}_S$.

Proof The claim is saying that if a network is reachably atomic, then there exists a molecular species that can be decomposed into its atomic decomposition in one single reaction. Proof is done by assuming otherwise and chasing the decomposition sequence to find an infinite descending chain of species ordered by the size of their decomposition vectors, contradicting the finiteness of species set. Suppose not, then for all reactions with $\mathbf{r} = \{1S\}$ for some $S \in A \setminus \Delta$, either $[\mathbf{p}] \cap (A \setminus \Delta) \neq \emptyset$, or $[\mathbf{p}] \subseteq \Delta$ but $\mathbf{p} \neq \mathbf{d}_S$. The last case cannot happen, due to the uniqueness of atomic decomposition for reachably atomic networks (Recall Lemma 8). Hence for all (\mathbf{r}, \mathbf{p}) with $\mathbf{r} = \{1S\}$ for some $S \in A \setminus \Delta$, $[\mathbf{p}] \cap (A \setminus \Delta) \neq \emptyset$ [*].

[*], together with the reachable-atomicity, implies that for each $S \in A \setminus \Delta$ one may find a S' s.t. $\|\mathbf{d}_S\|_1 > \|\mathbf{d}_{S'}\|_1$ [**]. To see this, consider an arbitrary $S_i \in A \setminus \Delta$: any (\mathbf{r}, \mathbf{p}) with $\mathbf{r} = \{1S_i\}$ has either $\|\mathbf{p}\|_1 = 1$, or $\|\mathbf{p}\|_1 \geq 2$. In the second case we are done, for any $S_j \in [\mathbf{p}] \cap (A \setminus \Delta)$ satisfies $\|\mathbf{d}_{S_j}\|_1 < \|\mathbf{d}_{S_i}\|_1$; in the first case, we have found some S_{i+1} s.t. $\mathbf{d}_{S_{i+1}} = \mathbf{d}_{S_i}$ (and we call such (\mathbf{r}, \mathbf{p}) an isomerization reaction), so we recursively inspect into all $(\mathbf{r}_1, \mathbf{p}_1)$ with $\mathbf{r}_1 = \{1S_{i+1}\}$. Such a recursion must finally terminate with some S_{i+m} that satisfies $(\exists (\mathbf{r}_m, \mathbf{p}_m) \mid \mathbf{r}_m = \{1S_{i+m}\}) \|\mathbf{p}_m\|_1 \geq 2$, for otherwise \mathbf{d}_{S_i} would not be reachably decomposable into \mathbf{d}_{S_i} via any

reaction sequence. It follows that any $S_{i+m+1} \in [\mathbf{p}_m] \cap (\mathcal{A} \cap \mathcal{A})$ satisfies $\|\mathbf{d}_{S_i}\|_1 > \|\mathbf{d}_{S_{i+m+1}}\|_1$.

We have argued that our initial assumption (for the sake of contradiction) implies $[**]$. But $[**]$ would imply that there exists no molecular species with minimal size, contradicting the finiteness of \mathcal{A} . \square

Theorem 19 REACHABLY-ATOMIC \in P.

Proof We need to exhibit a polynomial time algorithm that decides whether there exists a separation of \mathcal{A} into two non-empty, disjoint sets M (molecules) and \mathcal{A} (atoms), with elements in M decomposable via sequences of reactions into combination of elements in \mathcal{A} .

To achieve this goal, we set $M = \{S \in \mathcal{A} \mid (\exists(\mathbf{r}, \mathbf{p}) \in R) \mathbf{r} = \{1S\}\}$, the subset of species which are the single reactant of some reaction; apparently M is non-empty for reachably-atomic networks, by Lemma 18. Then recursively, we check if there exist elements in M that can be decomposed into combination of atoms via a reaction sequence of length $i = 1, 2, \dots$, and reject if we succeed to do so at $i = k$ but fails at $i = k + 1$ while not all elements in M have been examined. When this process terminates (note that M is finite) finding (candidate) atomic decomposition for all molecules, we verify if the necessary conservation laws hold. We describe this polynomial time algorithm in more details, argue for its correctness, and then exhibit the pseudo-code.

Our algorithm will compose of the following steps:

1. Enumerate all reactions in R : for each reaction starting with $\{1S\}$ for some $S \in \mathcal{A}$, put S in the set M of molecules. After the enumeration, define $\mathcal{A} = \mathcal{A} \setminus M$. If $\mathcal{A} = \emptyset$ or $M = \emptyset$, reject.
2. Find if there exists some molecular species $S \in M$ such that S is decomposable into some \mathbf{d}_S with $[\mathbf{d}_S] \subseteq \mathcal{A}$ by execution of a single reaction. If any of such \mathbf{d}_S is of size 0 or 1, reject. Keep track of each decomposition vector \mathbf{d}_S ;

Construct a subset M' of M s.t. M' contains all molecular species which *cannot* be decomposed into \mathbf{d}_S with $[\mathbf{d}_S] \subseteq \mathcal{A}$ by execution of one reaction. If $M' = M$, then reject.

3. While M' is not empty, iterate and try to find an $S \in M'$ that satisfies this condition: $\exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\mathbf{r} = \{1S\}$ and $[\mathbf{p}] \subseteq (M \setminus M') \cup \mathcal{A}$. Note that all elements S'' in $M \setminus M'$ satisfies $\{1S''\} \Rightarrow^* \mathbf{d}_{S''}$, hence if S can be decomposed into complexes consisting solely of elements in $(M \setminus M') \cup \mathcal{A}$ via execution of one reaction, then S itself satisfies $\{1S\} \Rightarrow^* \mathbf{d}_S$ as well. Keep track of \mathbf{d}_S and exclude such S from M' .

If in some iteration we cannot find any such $S \in M'$, then reject; else, the iteration will finally halt excluding all such S 's and making M' empty.

4. We have obtained \mathbf{d}_S for each $S \in \mathcal{A}$ (note that $\mathbf{d}_A = \mathbf{e}_A$ for all $A \in \mathcal{A}$) s.t. $[\mathbf{d}_S] \subseteq \mathcal{A}$ and $\{1S\} \Rightarrow^* \mathbf{d}_S$. By Lemma 8, such set of decomposition is unique. Now, check if this decomposition conserves atoms by composing stoichiometric matrix \mathbf{A} and decomposition matrix \mathbf{D} and examine if $\mathbf{A} \cdot \mathbf{D} = \mathbf{0}$, and finally check if each atom A appears at least once in some \mathbf{d}_S for some $S \in M$.

We first prove that if reachably atomic then the algorithm will halt in ACCEPT:

if \mathcal{C} is reachably atomic, then we claim that the set $\{S \in \mathcal{A} \mid \exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\{1S\} = \mathbf{r}\}$ is exactly the set of molecules M , with its complement $\mathcal{A} = \mathcal{A} \setminus M$. To see this, recall that we prescribed there being no " $\mathbf{r} \rightarrow \mathbf{r}$ " reactions in R , so all reactions $(\mathbf{r}, \mathbf{p}) \in R \mid \|\mathbf{r}\| = 1$ is either an isomerization reaction ($\|\mathbf{p}\|_1 = 1, \mathbf{p} \neq \mathbf{r}$) or disassociation reaction ($\|\mathbf{p}\|_1 \geq 2$). Both types of reactions can only happen when $S \in [\mathbf{r}]$ is a molecule; it follows that $\{S \in \mathcal{A} \mid \exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\{1S\} = \mathbf{r}\} \subseteq M$. Conversely, when $S \in M$, reachably atomicity gives $S \in \{S \in \mathcal{A} \mid \exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\{1S\} = \mathbf{r}\}$.⁶

Neither M nor \mathcal{A} would be empty, for $(\mathcal{A} = \emptyset) \Rightarrow (\mathcal{C}$ is not reachably atomic) and $(M = \emptyset) \Rightarrow (R = \emptyset)$. Hence the algorithm passes Step 1, correctly identifying the partition (M, \mathcal{A}) of \mathcal{A} .

By Lemma 18, reachable-atomicity implies that the algorithm will find at least one molecular species S that directly decomposes to its atomic decomposition \mathbf{d}_S and grantedly, $\|\mathbf{d}_S\| \geq 2$, so Step (2) will be passed.

Further, applying the same argument in Lemma 18 on the set M' , the while loop must shrink the cardinality of M' by at least 1 per iteration, and finally exit by making M' empty,⁷ passing Step (3);

Finally, the decomposition must preserve atoms for all reactions, and all atoms must appear in the decomposition of at least one molecule, which make both tests in Step (4) passed.

It remains to show that if \mathcal{C} is not reachably atomic then our algorithm will halt in REJECT. We claim that: if \mathcal{C} is not

⁶ We point out that the set of atoms $M \neq \{S \in \mathcal{A} \mid \exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $(\{1S\} = \mathbf{r}) \wedge (\|\mathbf{p}\| \geq 2)\}$, so we have to test the $\|\mathbf{d}_S\| \geq 2$ condition in later steps. This is because it might be the case that the only reaction (\mathbf{r}, \mathbf{p}) with $\mathbf{r} = \{1S\}$ turns out to be an isomerization reaction. A counter example would be:



By our definition $M = \{S \in \mathcal{A} \mid \exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\{1S\} = \mathbf{r}\}$, we shall correctly identify $M = \{A, B\}$, yet the added condition $\|\mathbf{p}\| \geq 2$ would make $M = \{B\}$, a mis-identification.

⁷ That is, if $(\forall S \in M')(\forall(\mathbf{r}, \mathbf{p}) \in R) (\mathbf{r} = \{1S\} \Rightarrow [\mathbf{p}] \cap M' \neq \emptyset)$, then for each species S in M' there will be $S' \in M'$ s.t. $\|S\|_1 > \|S'\|_1$, contradicting the finiteness of M' .

reachably atomic, then exactly one of the following will be true:

1. There is no valid separation of A into M and Δ . That is, either $\{S \in A \mid \exists(\mathbf{r}, \mathbf{p}) \in R \text{ s.t. } \{1S\} = \mathbf{r}\} = \emptyset$ (no species is the single reactant of an isomerization or disassociation reaction, so no species S is decomposable from $\{1S\}$), or $\{S \in A \mid \exists(\mathbf{r}, \mathbf{p}) \in R \text{ s.t. } \{1S\} = \mathbf{r}\} = A$ (every species is the single reactant of some isomerization or disassociation reaction, which contradicts the definition of subset atomicity for atoms should be neither isomerizable nor decomposable). An example where $\{S \in A \mid \exists(\mathbf{r}, \mathbf{p}) \in R \text{ s.t. } \{1S\} = \mathbf{r}\} = \emptyset$ would be $(A = \{A, B, C\}, R = \{2A + 3B \rightarrow 4C\})$, while $(A' = \{A, B\}, R' = \{A \rightarrow B, B \rightarrow A\})$ would be an instance where $\{S \in A \mid \exists(\mathbf{r}, \mathbf{p}) \in R \text{ s.t. } \{1S\} = \mathbf{r}\} = A$.

Observe such a valid separation (M, Δ) of A , if existing, is unique for a certain $\mathcal{C} = (A, R)$, since $S \in M$ if and only if $\exists(\mathbf{r}, \mathbf{p}) \in R$ s.t. $\mathbf{r} = \{1S\}$, and this property is uniquely decided by R .

2. There exists the unique valid separation (M, Δ) of A , but there exists no molecular species directly decomposable into its atomic decomposition via execution of one single reaction. That is, $(\forall S \in M)(\forall(\mathbf{r}, \mathbf{p}) \in R)(\mathbf{r} = \{1S\} \Rightarrow [\mathbf{p}] \cap M \neq \emptyset)$. An example of this is $(A = \{A, B, C\}, R = \{A \rightarrow B + 5C, B \rightarrow A + 5C\})$. We would successfully identify $M = \{A, B\}$, $\Delta = \{C\}$, but for all reactions we $(\mathbf{r}, \mathbf{p}) \in R$ have $[\mathbf{p}] \cap M \neq \emptyset$.

3. There exists the unique valid separation (M, Δ) of A and $\{S \in M \mid (\exists(\mathbf{r}_S, \mathbf{p}_S) \in R) ((\mathbf{r}_S = \{1S\}) \wedge ([\mathbf{p}_S] \subseteq \Delta))\} \neq \emptyset$, but for some $S \in \{S \in M \mid (\exists(\mathbf{r}_S, \mathbf{p}_S) \in R)((\mathbf{r} = \{1S\}) \wedge ([\mathbf{p}_S] \subseteq \Delta))\}$, $\|\mathbf{p}\| \leq 1$. That is, we have some reaction $S_1 \rightarrow A_1$ with $S_1 \in M$ and $A_1 \in \Delta$, or $S_1 \rightarrow \emptyset$. In this case, either a molecule decomposes to a single atom, or it vanishes.

Typical examples are: $\mathcal{C}_1 = (\{A, B, C\}, \{A \rightarrow B, B \rightarrow C\})$, $\mathcal{C}_2 = (\{A, B, C\}, \{A \rightarrow 2C, B \rightarrow \emptyset\})$. In both cases one would identify $M_{\mathcal{C}_1} = M_{\mathcal{C}_2} = \{A, B\}$; for both networks, $\{S \in M \mid (\exists(\mathbf{r}_S, \mathbf{p}_S) \in R)((\mathbf{r} = \{1S\}) \wedge ([\mathbf{p}_S] \subseteq \Delta))\} = \{A, B\}$. But in \mathcal{C}_1 , B decomposes to a single atom C ; in \mathcal{C}_2 , B vanishes. This disqualifies both sets from being reachably atomic by placing them in the third case.

4. There exists the unique valid separation (M, Δ) of A , and $\{S \in M \mid (\exists(\mathbf{r}_S, \mathbf{p}_S) \in R)((\mathbf{r}_S = \{1S\}) \wedge ([\mathbf{p}_S] \subseteq \Delta))\} \neq \emptyset$; further, each $S \in \{S \in M \mid (\exists(\mathbf{r}_S, \mathbf{p}_S) \in R)((\mathbf{r}_S = \{1S\}) \wedge ([\mathbf{p}_S] \subseteq \Delta))\}$ satisfies $\|\mathbf{p}\|_1 \geq 2$. However, there exists some indecomposable molecular species. That is, $\exists S' \in M$ s.t. $\forall \mathbf{c} \in \mathbb{N}^A$ where $\{1S'\} \Rightarrow^* \mathbf{c}$, $[\mathbf{c}] \cap M \neq \emptyset$.

An example for this case is $\mathcal{C} = (\{A, B, C, D, E\}, \{A \rightarrow B, B \rightarrow C, D \rightarrow 3E\})$. One may identify $M =$

$\{A, B, D\}$ and find $\{D\} = \{S \in M \mid (\exists(\mathbf{r}, \mathbf{p}) \in R)((\mathbf{r} = \{1S\}) \wedge ([\mathbf{p}] \subseteq \Delta))\}$. Further, the reaction $D \rightarrow 3E$ where $\mathbf{r} = \{1D\}$ satisfies $\|\mathbf{p}\| = 3$. This network does not belong to any of the first few cases but it does belong to Case 4, for $\forall \mathbf{c}$ where $\{1A\} \Rightarrow^* \mathbf{c}$, $[\mathbf{c}] \subseteq \{B, C\} \subseteq M$.

5. There is a unique valid decomposition (M, Δ) of A and $(\forall S \in M) (\exists \mathbf{c}_S$ with $[\mathbf{c}_S] \subseteq \Delta) (\{1S\} \Rightarrow^* \mathbf{c}_S) \wedge (\|\mathbf{c}_S\|_1 \geq 2)$, but the decomposition does not preserve atoms for some reaction. That is, with \mathbf{A} the stoichiometric matrix and \mathbf{D} the decomposition matrix (row vectors being the \mathbf{c}_S 's restricted to Δ), $\mathbf{A} \cdot \mathbf{D} \neq \mathbf{0}$. Note that for reachably atomic networks, atomic decomposition vectors (or equivalently, matrix) should be unique.

One example of this would be $(\{A, B, C, D\}, \{A \rightarrow B, B \rightarrow 3C, A + B \rightarrow 5C + D\})$. Note that here we have $M = \{A, B\}$ and $\{3C\} \stackrel{B \rightarrow 3C}{=} \mathbf{d}_B \stackrel{A \rightarrow B}{=} \mathbf{d}_A$

$\stackrel{A+B \Rightarrow 5C+D, B \rightarrow 3C}{=} \{2C + D\}$, contradicting the preserva-

tion of atoms. Note that this happens to be another example where a network is mass conserving (Just set $\mathbf{m}(A) = \mathbf{m}(B) = 3\mathbf{m}(C) = 3\mathbf{m}(D) = 3$) but not subset atomic (and hence not reachably atomic).

6. There is a unique valid decomposition (M, Δ) of A and $(\forall S \in M) (\exists \mathbf{c}_S$ with $[\mathbf{c}_S] \subseteq \Delta) (\{1S\} \Rightarrow^* \mathbf{c}_S) \wedge (\|\mathbf{c}_S\|_1 \geq 2)$, and the decomposition preserves atoms ($\mathbf{A} \cdot \mathbf{D} = \mathbf{0}$), but some atoms are "redundant": $\exists A \in \Delta$ s.t. $\forall S \in M, A \notin [\mathbf{d}_S] = [\mathbf{c}_S]$.

One may modify the last example into this case: $(\{A, B, C, D\}, \{A \rightarrow B, B \rightarrow 3C, A + B \rightarrow 6C\})$. Here we have $M = \{A, B\}$ and $\{3C\} \stackrel{B \rightarrow 3C}{=} \mathbf{d}_B \stackrel{A \rightarrow B}{=} \mathbf{d}_A$

$\stackrel{A+B \Rightarrow 6C, B \rightarrow 3C}{=} \{3C\}$, so every condition for reachably

atomic is satisfied, except that $(\forall S)D \notin [\mathbf{d}_S]$.

All six cases exclude each other, so *at most* one case could hold; on the other hand, the negation of the disjunction of all six cases says that there exists a non-empty proper subset of A and a decomposition matrix \mathbf{D} , such that all three conditions of primitive atomicity holds with respect to A via \mathbf{D} , and $\{1S\} \Rightarrow^* \mathbf{d}_S$ for all S . This is exactly the definition of reachably atomicity. So taking contraposition, non-reachable-atomicity implies *at least* one of the six cases hold.

Instances satisfying Case (1) will be rejected in Step (1), while Cases (2) and (3) will get rejected in Step (2). In case (4), the loop for finding decomposition vectors must terminate before M' is emptied, so it will get rejected by Step (3); Cases (5) and (6) triggers rejection in Step (4).

We present the pseudocode for the decider in Algorithm 1:

```

1 Initialize global set  $M, M', \Delta, D = \emptyset$  //  $D$ : {decomposition
  vectors}.
2 for  $(\mathbf{r}, \mathbf{p}) \in R$  do
3   if  $(\exists S \in \mathcal{A}) \mathbf{r} = \{1S\}$  then
4      $M \leftarrow M \cup \{S\}$ 
5   end
6 end
7  $\Delta \leftarrow \mathcal{A} \setminus M$ 
8  $M' \leftarrow M$ 
9 if  $M = \emptyset$  or  $\Delta = \emptyset$  then
10  REJECT
11 end
12  $D \leftarrow D \cup \{\mathbf{e}_A\}_{A \in \Delta}$ 
13 for  $(\mathbf{r}, \mathbf{p}) \in R$  where  $(\exists S \in M) \mathbf{r} = \{1S\}$  do
14   if  $[\mathbf{p}] \subseteq \Delta$  then
15     if  $\|\mathbf{p}\|_1 \leq 1$  then
16       REJECT
17     end
18   end
19    $D \leftarrow D \cup \{\langle \mathbf{d}_S = \mathbf{p} \rangle\}$ 
20    $M' \leftarrow M' \setminus \{S\}$ 
21 end
22 if  $M' = M$  then
23  REJECT
24 end
25 while  $M' \neq \emptyset$  do
26   if  $(\forall S \in M') (\forall (\mathbf{r}, \mathbf{p}) \in R \mid \mathbf{r} = \{1S\}) ([\mathbf{p}] \cap M' \neq \emptyset)$ 
27     then
28       REJECT
29     end
30   else
31     for  $(S \in M' \mid (\exists (\mathbf{r}, \mathbf{p}) \in R \mid \mathbf{r} = \{1S\}$ 
32       and  $[\mathbf{p}] \cap M' = \emptyset))$  do
33        $D \leftarrow D \cup \{\langle \mathbf{d}_S = \sum_{S' \in [\mathbf{p}]} \mathbf{d}_{S'} \rangle\}$ 
34        $M' \leftarrow M' \setminus \{S\}$ 
35     end
36   end
37 end
38 Compose  $\mathbf{A}$  (stoichiometric matrix) and  $\mathbf{D}$  (decomposition
  matrix)
39 if  $\mathbf{A} \cdot \mathbf{D} \neq \mathbf{0}$  then
40  REJECT
41 end
42 if  $(\exists A \in \Delta) (\forall S \in M) A \notin [\mathbf{d}_S]$  then
43  REJECT
44 end
45 ACCEPT

```

Algorithm 1: Decider for REACHABLY-ATOMIC

Let us briefly mention that this is a polynomial time algorithm. The first for-loop takes $O(|R|)$ time; the second for-loop takes at most $O(|R|)$ iterations, and each iteration takes $O(|\mathcal{A}|^3)$ time; as for the while loop, note that it either shrinks the size of M' by 1 per iteration, or REJECTS. Hence the while loop takes at most $O(|\mathcal{A}|)$ to exit. The if-statement inside the while-loop takes $O(|\mathcal{A}| \cdot |R| \cdot |\mathcal{A}|^2)$ to evaluate. Lastly, composing and multiplying $\mathbf{A} \cdot \mathbf{D}$ takes $O(|R||\mathcal{A}| \cdot |\mathcal{A}|^2)$ time, and verifying each $A \in \Delta$ is “used”

by the decomposition of some molecule is $O(|\mathcal{A}| \cdot |\mathcal{A}| \cdot |\mathcal{A}|)$. The times complexity is therefore dominated by the while, which is $O(|R||\mathcal{A}|^4)$. No input, output or intermediate encoding takes more than polynomial space to record, so REACHABLY-ATOMIC $\in \mathbf{P}$, as desired. \square

6.2 REACHABLE-REACH is PSPACE-complete

We shall first introduce the definition of configuration reachability graphs, followed by a result proved in Mayr and Weihmann (2014) (see also Sect. 1.1), based on which we prove REACHABLE-REACH (see Definition 23), a problem motivated by restricting relevant problems such as “exact reachability” (Leroux 2011), is PSPACE-complete.

Definition 20 (Configuration Reachability Graph) An *i*-initiated Configuration Reachability Graph $G_{\mathcal{C}, \mathbf{i}}$ of the chemical reaction network $\mathcal{C} = (\mathcal{A}, R)$ is a directed graph (V, E) , where:

1. each $v_{\mathbf{c}} \in V$ ($\mathbf{c} \in \mathbb{N}^{\mathcal{A}}$) is labeled by a reachable configuration \mathbf{c} of \mathcal{C} ;
2. $v_{\mathbf{i}} \in V$ ($\mathbf{i} \in \mathbb{N}^{\mathcal{A}}$) is the vertex labeled by the initializing configuration \mathbf{i} ;
3. the ordered pair $(v_{\mathbf{c}_1}, v_{\mathbf{c}_2}) \in E$ if and only if $\mathbf{c}_1 \Rightarrow^1 \mathbf{c}_2$.

Remark 21 For the sake of simplicity, we use $G_{\mathcal{C}, \mathbf{i}}$ as shorthand for $G_{\mathcal{C}, v_{\mathbf{i}}}$.

For the same \mathcal{C} , Configuration Reachability Graphs can be far from isomorphic due to parameterization by different initialization vectors. We have included an example (Example D.3) in Doty and Zhu (2017).

We will soon prove the conclusion on the complexity of the reachability problem for reachably atomic networks. But first, we point out that the following is a straightforward translation of a finding in Mayr and Weihmann (2014), giving the complexity class of reachability problems for mass-conserving chemical reaction networks.

Observation 22 (Result from Mayr and Weihmann 2014) For all mass conserving chemical reaction networks \mathcal{C} and initial configuration \mathbf{i} of \mathcal{C} , $|G_{\mathcal{C}, \mathbf{i}}| \in O(2^{\text{poly}(|\mathcal{C}, \mathbf{i}|)})$. That is, the binary size of the encoding of the configuration reachability graph $G_{\mathcal{C}, \mathbf{i}}$ is at most exponential to the binary size of the encoding of the pair $(\mathcal{C}, \mathbf{i})$.

Furthermore, the reachability problem for mass conserving networks is PSPACE-complete. That is, it is PSPACE-complete to decide if an instance is in the following language:

$$\{\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \mid (A, R) \text{ is mass conserving; } \mathbf{c}_1, \mathbf{c}_2 \in \mathbb{N}^A; \mathbf{c}_1 \Rightarrow^* \mathbf{c}_2 \rangle\}$$

Built on Observation 22, we now exhibit the proof that the decision problem “Given a Reachably Atomic network, is \mathbf{c}_2 reachable from \mathbf{c}_1 ” is PSPACE-Complete.

Definition 23 (REACHABLE-REACH) We define the language REACHABLE-REACH = $\{\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle \mid (A, R) \text{ is reachably atomic; } \mathbf{c}_1, \mathbf{c}_2 \in \mathbb{N}^A; \mathbf{c}_1 \Rightarrow^* \mathbf{c}_2\}$

Proposition 24 REACHABLE-REACH is PSPACE-Complete.

Proof REACHABLE-REACH \in PSPACE is a direct application of Observation 22—note that all reachably-atomic chemical reaction networks are primitive atomic, and hence mass conserving (Proposition 10). Hardness is shown by simulating polynomial space Turing Machines via reactions.

Let $\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle$ be an instance, and let $Z := |\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle|$. In Theorem 19 we proved that REACHABLY-ATOMIC $\in P$ so we may run the polynomial decider on $\langle A, R \rangle$ and REJECT if $\langle A, R \rangle \notin$ REACHABLY-ATOMIC.

If the REACHABLY-ATOMIC decider halts in ACCEPT, we would obtain $\Delta \subseteq A$ with respect to which (A, R) is reachably atomic, as well as the set $\{\mathbf{d}_S\}_{S \in \Delta}$ of decomposition vectors. Further, we would have confirmed that $\mathcal{C} = (A, R)$ is mass-conserving, for this is implied by reachably atomicity. Recall from Mayr and Weihmann (2014) that the number of vertices in a configuration reachability graph $G_{\mathcal{C}, \mathbf{c}_1}$ for mass-conserving network \mathcal{C} is at most exponential to the binary size of the input. Now, let $n = |V_{G_{\mathcal{C}, \mathbf{c}_1}}|$, then by Savitch’s Theorem (Papadimitriou 2003),

$$\begin{aligned} \text{REACHABLE-REACH} &\in \text{SPACE}((\log n)^2) \\ &= \text{SPACE}((\log(O(2^{\text{poly}(Z)})))^2) \\ &= \text{SPACE}(O(\text{poly}(L))^2) \end{aligned}$$

It follows that REACHABLE-REACH \in PSPACE.

As for the PSPACE hardness, we shall prove by simulating a polynomial-space Turing Machine. That is, consider the language

$$L := \{\langle M, x, 0^{|x|^c} \mid M \text{ is an } O(|x|^c) \text{ - space, clocked Turing Machine, } x \in \{0, 1\}^* : M(x) \rightarrow 1\}$$

Just to clarify the notation, “ $M(x) \rightarrow 1$ ” means M on the input x runs for $O(|x|^c)$ time and ACCEPTS. We shall construct a REACHABLE-REACH instance $\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle$ by a polynomial time reduction from an instance $\langle M, x, 0^{|x|^c} \rangle$, and show that $\langle M, x, 0^{|x|^c} \rangle \in L$ if and only if $\langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle \in$ REACHABLE-REACH.

Without loss of generality, assume the initial configuration of M is $q_1 \in Q_M = \{q_1, q_2, \dots, q_{t-2}, q_A, q_R\}$, where $t := |Q_M|$, q_A is the accept state and q_R is the reject state, and assume that the Turing Machine blank the tape cells and return the tape head to the leftmost position before halting. Let p denote the maximum number of tape cells that M may use on input x (Note that $p \in O(|x|^c)$). Define the following set of species:

$$A = \{A, \underbrace{Q_1, \dots, Q_{t-2}, Q_A, Q_R}_{\text{imitate machine states}}, \underbrace{P_1, \dots, P_p}_{\text{imitate head positions}}, \underbrace{T_1^0, T_1^1, \dots, T_p^0, T_p^1}_{\text{imitate tape contents}}\}$$

and configurations:

$$\begin{aligned} \mathbf{c}_1 &= \{P_1, Q_1, T_1^{x[1]}, \dots, T_{|x|}^{x[|x|]}, T_{|x|+1}^\sqcup, \dots, T_p^\sqcup\} \\ \mathbf{c}_2 &= \{1Q_A, T_1^\sqcup, \dots, T_p^\sqcup, P_1\} \end{aligned}$$

Recalling that the Turing Machine transition function $\delta_M : Q_M \setminus \{q_A, q_R\} \times \Gamma \rightarrow Q_M \times \Gamma \times \{-1, +1\}$, we construct the set R of reactions in as in Algorithm 2.⁸

```

1 for  $(\forall q_i \in Q_M)(\forall b \in \{0, 1, \sqcup\})(\forall k \in \{1, 2, \dots, p\})$  do
2   if  $\delta(q_i, b) \mapsto (q_j, b', m)$  then
3     Add Reaction
4      $Q_i + T_k^b + P_k \rightarrow Q_j + T_k^{b'} + P_{k+m}$ 
5     //  $m \in \{\pm 1\}$  : tape head moving direction.
6   end
7 for  $S \in A \setminus \{A\}$  do
8   Add Reaction  $S \rightarrow 2A$ 
9 end

```

Algorithm 2: Construction of R

Observe that (A, R) is a reachably atomic network with respect to $\Delta = \{A\} \subseteq A$, for any molecular species can be decomposed to $\{2A\}$ explicitly via Lines 11–12, A appears in the decomposition of all molecular species, and all reactions preserve the number of atoms.

Further,

$$\begin{aligned} \langle M, x, 0^{|x|^c} \rangle \in L &\Leftrightarrow M(x) \rightarrow 1 \\ &\Leftrightarrow \exists \text{computation path } (q_1, (x[1], x[2], \dots, \\ &\quad x[|x|], \underbrace{\sqcup, \dots, \sqcup}_{p-|x|})) \Rightarrow^* \\ &\quad (q_A, \underbrace{\sqcup, \sqcup, \dots, \sqcup}_p) \\ &\Leftrightarrow \mathbf{c}_1 \Rightarrow^* \mathbf{c}_2 \\ &\Leftrightarrow \langle A, R, \mathbf{c}_1, \mathbf{c}_2 \rangle \in \text{REACHABLE-REACH} \end{aligned}$$

⁸ We organize the construction procedure as an algorithm to make the description more concise.

Finally, $|A| = 1 + t + 3p$; $|R| \in O(3pq + |A|)$, $\|\mathbf{c}\|_1 = \|\mathbf{c}\|_2 = p + 2$. All coefficients of reactions are constant. Hence this reduction is polynomial in Z both timewise and spacewise. \square

Remark 25 The fact that the coefficients of all reactions involved in the proof of Proposition 24 are constant also implies that REACHABLE-REACH is PSPACE-hard (and hence complete) in the strong sense. Another side remark on the irreversibility of reactions may be found in Section 8(D) (Doty and Zhu 2017).

We also found connections between our definitions of “-atomic” and the concept of “core composition”, addressed by Gnacadja (2011) and detailed in Section 8(C) (Doty and Zhu 2017). Some interesting results are:

1. **Lemma C.12** states that a network is subset atomic if and only if it admits a “near-core composition” with certain restrictions;
2. **Lemma C.20** in the same section says reachable-atomicity implies admitting a core composition;
3. **Theorem C.15** gives the equivalence between “reversibly-reachable atomic” and “explicitly-reversibly constructive with no isomeric elementary species”.

7 Open problems

Conjecture 26 SUBSET-ATOMIC \in NP-complete.

One may note that there are two sources of indeterminacy in the problem SUBSET-ATOMIC: the choice of Δ and \mathbf{D} . For example, the network constructed in the proof of NP-hardness of SUBSET-FIXED-ATOMIC would remain subset atomic if we define $\Delta = \{T, F\}$, and let $\mathbf{d}_P = \mathbf{d}_Q = \{kT, sF\}$ for any $k, s \geq 2$.

There is a formal sense in which chemical reaction networks have been shown to be able to compute functions $f : \mathbb{N}^k \rightarrow \mathbb{N}$ (Chen et al. 2013) and predicates $\mathbb{N}^k \rightarrow \{0, 1\}$ (Angluin et al. 2006). A function/predicate can be computed “deterministically” (i.e., regardless of the order in which reactions occur) \iff it is semilinear (see Ginsburg and Spanier 1966 for a definition).

Problem 27 What semilinear functions or predicates can atomic chemical reaction networks compute *deterministically*, and how efficiently? What general functions or predicates can atomic chemical reaction networks compute *with high probability*, and how efficiently?

Remark 28 A partial answer for Problem 27 based on results in Chen et al. (2013) says that primitive atomic networks and subset atomic networks can stably compute

any semilinear functions, but it is not obvious how to modify the subset-atomic network into reachably-atomic with the stably-computation property maintained, or whether it is even possible to do so.

In fact, in their proof of the lemma that any semilinear function $f : \mathbb{N}^k \rightarrow \mathbb{N}$ can be stably computed by a chemical reaction network, Chen et al. (2013) designed a chemical reaction network which can be made primitive-atomic by a slight modification:



The modified chemical reaction network (7.1)–(7.6) still stably computes the same function f , as the waste product W_j 's do not participate in any other reactions. This network is mass-conserving, via the mass distribution function

$$\begin{aligned} \mathbf{m} : \{T_i, F_i, Y_j, K_j, W_j, Y_{ij}^P, \widehat{Y}_{ij}^P, Y_{ij}^C, \widehat{Y}_{ij}^C, M_{ij}\}_{ij} &\rightarrow \mathbb{R} \\ T_i, F_i, Y_j, Y_{ij}^P, M_{ij}, Y_{ij}^C, \widehat{Y}_{ij}^C &\mapsto 2, \\ \widehat{Y}_{ij}^P, K_j &\mapsto 4, \\ W_j &\mapsto 6 \end{aligned}$$

And by setting $\Delta = \{A\}$ and $\mathbf{d}_S(A) = m(S)$ ($\forall S \in \{T_i, F_i, Y_j, K_j, W_j, Y_{ij}^P, \widehat{Y}_{ij}^P, Y_{ij}^C, \widehat{Y}_{ij}^C, M_{ij}\}_{ij} = A$), we find that the network above is also primitive atomic. This shows that any semilinear function can be stably computed by a primitive-atomic chemical reaction network.

Redefining $A \leftarrow A \cup \Delta$, we obtain a subset-atomic network that stably computes f , which implies that the computation power of subset-atomic chemical reaction networks are no weaker than primitive-atomic chemical reaction networks.

8 Conclusion

We have established the computational complexity to decide different atomic properties of chemical reaction networks: polynomial for primitive atomic, NP for subset atomic (conjectured NP-complete), and polynomial for reachably atomic. We also determined the computational complexity for membership-decision of some relevant

languages, such as SUBSET-FIXED-ATOMIC (NP-complete) and REACHABLE-REACH (PSPACE-complete).

Acknowledgements The authors are thankful to Manoj Gopalkrishnan, Gilles Gnacadja, Javier Esparza, Sergei Chubanov, Matthew Cook, and anonymous reviewers for their insights and useful discussion.

Funding Funding was provided by NSF (Grant No. 1619343).

References

- Adleman L, Gopalkrishnan M, Huang M-D, Moisset P, Reishus D (2014) On the mathematics of the law of mass action. Springer, Dordrecht, pp 3–46. https://doi.org/10.1007/978-94-017-9041-3_1
- Alistarh D, Aspnes J, Eisenstat D, Gelashvili R, Rivest R (2017) Time-space trade-offs in molecular computation. In: Proceedings of the twenty-eighth annual ACM-SIAM symposium on discrete algorithms. pp 2560–2579
- Angeli D, De Leenheer P, Sontag ED (2007) A Petri net approach to the study of persistence in chemical reaction networks. *Math Biosci* 210:598–618
- Angluin D, Aspnes J, Diamadi Z, Fischer M, Peralta R (2006) Computation in networks of passively mobile finite-state sensors. *Distrib Comput* 18:235–253. <https://doi.org/10.1007/s00446-005-0138-3> (Preliminary version appeared in PODC 2004)
- Brijder R, Doty D, Soloveichik D (2016) Robustness of expressivity in chemical reaction networks. In: DNA 2016: proceedings of the 22th international meeting on DNA computing and molecular programming
- Cardelli L, Csikász-Nagy A (2012) The cell cycle switch computes approximate majority. *Scientific Reports*, 2
- Chen H-L, Cummings R, Doty D, Soloveichik D (2014) Speed faults in computation by chemical reaction networks. *Distrib Comput* (2015, to appear, special issue of invited papers from DISC)
- Chen H-L, Doty D, Soloveichik D (2013) Deterministic function computation with chemical reaction networks. *Nat Comput* 13(4):517–534 (Special issue of invited papers from DNA 2012)
- Chen H-L, Doty D, Soloveichik D (2014) Rate-independent computation in continuous chemical reaction networks. In ITCS 2014: proceedings of the 5th conference on innovations in theoretical computer science. pp 313–326
- Chen Y-J, Dalchau N, Srinivas N, Phillips A, Cardelli L, Soloveichik D, Seelig G (2013) Programmable chemical controllers made from DNA. *Nat Nanotechnol* 8(10):755–762
- Chubanov S (2015) A polynomial projection algorithm for linear feasibility problems. *Math Program* 153(2):687–713
- Craciun G, Dickenstein A, Shiu A, Sturmfels B (2009) Toric dynamical systems. *J Symb Comput* 44(11):1551–1565
- Cummings R, Doty D, Soloveichik D (2015) Probability 1 computation with chemical reaction networks. *Nat Comput* 1–17. <https://doi.org/10.1007/s11047-015-9501-x> (Special issue of invited papers from DNA 2014)
- Deshpande A, Gopalkrishnan M (2013) Autocatalysis in reaction networks. arXiv preprint [arXiv:1309.3957](https://arxiv.org/abs/1309.3957)
- Doty David (January 2014) Timing in chemical reaction networks. In: SODA 2014: proceedings of the 25th annual ACM-SIAM symposium on discrete algorithms, pp 772–784
- Doty D, Hajiaghayi M (2015) Leaderless deterministic chemical reaction networks. *Nat Comput* 14(2):213–223 (Preliminary version appeared in DNA 2013)
- Doty D, Zhu S (2017) Computational complexity of atomic chemical reaction networks. arXiv preprint [arXiv:1702.05704](https://arxiv.org/abs/1702.05704)
- Doty D, Zhu S (2018) Computational complexity of atomic chemical reaction networks. In: International conference on current trends in theory and practice of informatics. pp 212–226. Springer, New York
- Esparza J, Ganty P, Leroux J, Majumdar R (2017) Verification of population protocols. *Acta Inform* 54(2):191–215
- Garey MR, Johnson DS (1978) “strong” NP-completeness results: motivation, examples, and implications. *JACM* 25(3):499–508
- Garey MR, Johnson DS (1979) Computers and intractability. W. H. Freeman, New York
- Ginsburg S, Spanier EH (1966) Semigroups, Presburger formulas, and languages. *Pac J Math* 16(2):285–296. <http://projecteuclid.org/euclid.pjm/1102994974>
- Gnacadja G (2011) Reachability, persistence, and constructive chemical reaction networks (part II): a formalism for species composition in chemical reaction network theory and application to persistence. *J Math Chem* 49(10):2137
- Gopalkrishnan M (2016) Private communication. Email
- Guldberg CM, Waage P (1864) Studies concerning affinity. Forhandlinger: Videnskabs-Selskabet i Christiania. In: Norwegian Academy of Science and Letters, 35
- Horn FJM (1974) The dynamics of open reaction systems. In SIAM-AMS proceedings VIII, pp 125–137
- Hua J, Ahmad SS, Riedel Marc D, Parhi Keshab K (2013) Discrete-time signal processing with DNA. *ACS Synth Biol* 2(5):245–254
- Leroux J (2011) Vector addition system reachability problem: a short self-contained proof. In: ACM SIGPLAN notices. ACM, vol 46, pp 307–316
- Lien YE (1976) A note on transition systems. *Inform Sci* 10(2):347–362
- Mayr EW, Weihmann J (2014) A framework for classical Petri net problems: conservative petri nets as an application. In: International conference on applications and theory of petri nets and concurrency. Springer, New York, pp 314–333
- Montagne K, Plasson R, Sakai Y, Fujii T, Rondelez Y (2011) Programming an in vitro DNA oscillator using a molecular networking strategy. *Mol Syst Biol* 7(1):466
- Napp NE, Adams RP (2013) Message passing inference with chemical reaction networks. In: Advances in neural information processing systems. pp 2247–2255
- Oishi K, Klavins E (2011) Biomolecular implementation of linear I/O systems. *IET Syst Biol* 5(4):252–260
- Padirac A, Fujii T, Rondelez Y (2013) Nucleic acids for the rational design of reaction circuits. *Curr Opin Biotechnol* 24(4):575–580
- Papadimitriou Christos H (1981) On the complexity of integer programming. *JACM* 28(4):765–768
- Papadimitriou CH (2003) Computational complexity. Wiley, New York
- Qian L, Winfree E, Bruck J (2011) Neural network computation with dna strand displacement cascades. *Nature* 475(7356):368–372
- Qian L, Winfree E (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science* 332(6034):1196
- Roos K (2015) An improved version of Chubanov’s method for solving a homogeneous feasibility problem. *Optim Method Softw* 33(1):26–44
- Salehi SA, Parhi KK, Riedel MD (2016) Chemical reaction networks for computing polynomials. *ACS Synth Biol* 6(1):76–83
- Salehi SA, Riedel MD, Parhi KK (2014) Asynchronous discrete-time signal processing with molecular reactions. In: 2014 48th Asilomar conference on signals, systems and computers, pp 1767–1772

- Salehi SA, Riedel MD, Parhi KK (2015) Markov chain computations using molecular reactions. In: 2015 IEEE international conference on digital signal processing (DSP), pp 689–693
- Savitch WJ (1970) Relationships between nondeterministic and deterministic tape complexities. *J Comput Syst Sci* 4(2):177–192
- Seelig G, Soloveichik D, Zhang DY, Winfree E (2006) Enzyme-free nucleic acid logic circuits. *Science* 314(5805):1585–1588. <https://doi.org/10.1126/science.1132493>
- Silberschatz A, Galvin PB, Gagne G, Silberschatz A (2013) Operating system concepts, vol 4. Addison-Wesley, Reading
- Soloveichik D, Cook M, Winfree E, Bruck J (2008) Computation with finite stochastic chemical reaction networks. *Nat Comput* 7(4):615–633. <https://doi.org/10.1007/s11047-008-9067-y>
- Soloveichik D, Seelig G, Winfree E (2010) DNA as a universal substrate for chemical kinetics. *Proc Natl Acad Sci* 107(12):5393 **(Preliminary version appeared in DNA 2008)**
- Srinivas N (2015) Programming chemical kinetics: engineering dynamic reaction networks with DNA strand displacement. PhD thesis, California Institute of Technology
- Thachuk C, Condon A (2012) Space and energy efficient computation with DNA strand displacement systems. In: DNA 2012: proceedings of the 18th international meeting on DNA computing and molecular programming, pp 135–149
- Yurke B, Turberfield AJ, Mills AP Jr, Simmel FC, Neumann JL (2000) A DNA-fuelled molecular machine made of DNA. *Nature* 406(6796):605–608