Robustness of Expressivity in Chemical Reaction Networks

Peer-reviewed author version

Robustness of expressivity in chemical reaction networks[⋆]

Robert Brijder[1], David Doty[2], and David Soloveichik[3]

[1] Hasselt University, Diepenbeek, Belgium, `robert.brijder@uhasselt.be`
[2] University of California, Davis, CA, USA, `doty@ucdavis.edu`
[3] University of Texas, Austin, TX, USA, `david.soloveichik@utexas.edu`

Abstract. We show that some natural output conventions for error-free computation in chemical reaction networks (CRN) lead to a common level of computational expressivity. Our main results are that the standard definition of error-free CRNs have equivalent computational power to 1) *asymmetric* and 2) *democratic* CRNs. The former have only "yes" voters, with the interpretation that the CRN's output is yes if any voters are present and no otherwise. The latter define output by majority vote among "yes" and "no" voters.

Both results are proven via a generalized framework that simultaneously captures several definitions, directly inspired by a recent Petri net result of Esparza, Ganty, Leroux, and Majumder [CONCUR 2015]. These results support the thesis that the computational expressivity of error-free CRNs is intrinsic, not sensitive to arbitrary definitional choices.

1   Introduction

Turing machines solve exactly the same class of yes/no decision problems whether they report output via accept/reject states, or if instead they write a 1 or 0 on a worktape before halting. Similarly, finite-state transducers compute the same class of functions whether they emit output on a state (*Moore machine* [19]) or a transition (*Mealy machine* [18]). In general, if the power of a model of computation is insensitive to minor changes in the definition, this lends evidence to the claim that the model is robust enough to apply to many real situations, and that theorems proven in the model reflect fundamental truths about reality, rather than being artifacts of arbitrary definitional choices.

The theory of chemical reaction networks (CRNs) studies the general behavior of chemical reactions in well-mixed solutions, abstracting away spatial properties of the molecules. Formally, a CRN is defined as a finite set of reactions such as $2A + C \to 2B$, where A, B, and C are abstract chemical species. In a discrete CRN the state of the system is given by molecule counts of each species and the system updates by application of individual reactions.

CRNs have only recently been considered as a model of computation [21], motivated partially by the ability to implement them using a basic experimental technique called *DNA strand displacement* [22]. Discrete CRNs are Turing

---

complete if allowed an arbitrary small, but nonzero, probability of error [21], improved to probability 0 in [9]. Using a result from the theory of population protocols [3, 4], it is known that error-free CRNs decide exactly the semilinear sets [6].[4]

We study the computational robustness of error-free CRNs under different output conventions. The original output convention [3] for deciding predicates (0/1-valued functions) is that each species is classified as voting either 0 ("no") or 1 ("yes"), and a configuration (vector of nonnegative integer counts of each species) o has output $i \in \{0, 1\}$ if all species present in positive count are i-voters, i.e., there is a *consensus* on vote i. As an example, the CRN with reactions $X_1 + N \to Y$ and $X_2 + Y \to N$, with initial configuration $\{x_1 X_1, x_2 X_2, 1N\}$, where N, X_2 vote 0 and Y, X_1 vote 1, decides if $x_1 > x_2$; Y and N alternate being present as each reacts with an input, so the first input to run out determines whether we stop at Y or N. More formally, we say o is *output-stable* if every configuration o' reachable from o has the same output as o (i.e., the system need not halt, but it stops changing its output). Finally, it is required that a correct output-stable configuration is reachable not only from the initial configuration i, but also from any configuration reachable from i; under mild assumptions (e.g., conservation of mass), this implies that a correct stable configuration is actually reached with probability 1 under the standard stochastic kinetic model [14]. It has been shown in [3] that the computational power is not reduced, that is, it still decides precisely all semilinear sets, when we restrict to those CRNs where (1) each reaction has two reactants and two products (e.g., disallowing reactions such as $2A + C \to 2B$ and $A \to B + C$, a model known as a *population protocol* [3]) and (2) the system eventually halts for every possible input (see also [7]).

One can imagine alternative output conventions, i.e., ways to interpret what is the output of a configuration, while retaining the requirement that a correct output-stable configuration is reachable from any reachable configuration. Rather than requiring every species to vote 0 or 1, for example, allow the CRN to designate some species as nonvoters. It is not difficult to show that such CRNs have equivalent computational power: They are at least as powerful since one can always choose all species to be voters. The reverse direction follows by converting a CRN with a subset of voting species into one in which every species votes, by replacing every nonvoting species S with two variants S_0 and S_1, whose voting bit is swayed by reactions with the original voting species, and which are otherwise both functionally equivalent to S.

We investigate two output conventions that are not so easily seen to be convertible to the original convention. The first convention is the *asymmetric* model, in which there are only 1-voters, whose presence or absence indicates a configuration-wide output of 1 or 0, respectively. It is not obvious how to con-

---

[4] We use the term "error-free" in this section to refer to a specific requirement of "stability" defined formally in Section 2.2. When the set of configurations reachable from an initial configuration is always finite (for instance, with population protocols, or more generally mass-conserving CRNs), then stability coincides with probability 0 of error. See [9] for an in-depth discussion of how these notions can diverge when the set of configurations reachable from an initial configuration is infinite.

vert an asymmetric CRN into a symmetric CRN, since this appears to require producing 0-voters if and only if 1-voters are absent. The second convention is the *democratic* model, in which there are 0- and 1-voters, but the output of a configuration is given by the majority vote rather than being defined only with consensus. Intuitively, the difficulty in converting a democratic CRN into a symmetric consensus CRN is that, although the democratic CRN may stabilize on a majority of, for example, 1-voters over 0-voters, the exact numerical gap between them may never stabilize. A straightforward attempt to convert a democratic CRN into a consensus CRN results in a CRN that changes the output every time a new 0- or 1-voter appears. For instance, suppose we use the previously described CRN for computing whether $x_1 > x_0$, where x_1 and x_0 respectively represent the count of 1- and 0-voters. If the original democratic CRN repeatedly increments x_0 and then x_1, the resulting CRN flips between Y and N indefinitely — thus never stabilizing in the consensus model — even if $x_1 > x_0$ remains true indefinitely.

We show that these conventions have equivalent power as the original definition. Our techniques further establish that the class of predicates computable by CRNs is robust to two additional relaxations of the classical notion of stable computation [3]: (1) a correct output configuration need not be reachable from *every* reachable configuration, only the initial configuration, and (2) the set of output configurations need not be "stable" (i.e., closed under application of reactions), so long as each initial configuration can reach only a correct output.

After defining existing notions of computation by CRNs in Section 2, we introduce in Section 3 a very general computational model for CRNs, called a *generalized chemical reaction decider* (gen-CRD). Its definition is directly inspired by a recent powerful result from Petri net theory [13], restated here as Theorem 3.2. Using this result we show that under mild conditions, gen-CRDs decide only semilinear sets. We then show that the original symmetric consensus model, the asymmetric consensus model, and the symmetric majority model all fit into this framework, establishing their common expressivity.

2 Chemical reaction networks and deciders

2.1 Chemical reaction networks

Let $\mathbb{Z}$ and $\mathbb{N}$ denote the integers and nonnegative integers, respectively. Let Λ be a finite set. The set of vectors over $\mathbb{N}$ indexed by Λ (i.e., the set of functions $\boldsymbol{c} : \Lambda \to \mathbb{N}$) is denoted by $\mathbb{N}^\Lambda$. The zero vector is denoted $\boldsymbol{0}$. For $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{N}^\Lambda$ we write $\boldsymbol{c} \leq \boldsymbol{c}'$ if and only if $\boldsymbol{c}(S) \leq \boldsymbol{c}'(S)$ for all $S \in \Lambda$. For $\boldsymbol{c} \in \mathbb{N}^\Lambda$ and $\Sigma \subseteq \Lambda$, the *projection* of $\boldsymbol{c}$ to Σ, denoted by $\boldsymbol{c} \restriction_\Sigma$, is an element in $\mathbb{N}^\Sigma$ such that $\boldsymbol{c} \restriction_\Sigma (S) = \boldsymbol{c}(S)$ for all $S \in \Sigma$. Let $\|\boldsymbol{c}\| = \|\boldsymbol{c}\|_1 = \sum_{S \in \Lambda} \boldsymbol{c}(S)$ denote the L_1 norm of $\boldsymbol{c}$. We sometimes use multiset notation, e.g., $\boldsymbol{c} = \{1A, 2C\}$ to denote $\boldsymbol{c}(A) = 1, \boldsymbol{c}(C) = 2, \boldsymbol{c}(S) = 0$ for $S \notin \{A, C\}$, or when defining reactions, additive notation, i.e., $A + 2C$.

A *reaction* α over Λ is a pair $(\boldsymbol{r}, \boldsymbol{p})$ with $\boldsymbol{r}, \boldsymbol{p} \in \mathbb{N}^\Lambda$ and $\boldsymbol{r} \neq \boldsymbol{p}$, where $\boldsymbol{r}$ and $\boldsymbol{p}$ are the *reactants* and *products* of α, respectively. We write $\boldsymbol{r} \to \boldsymbol{p}$ to denote a reaction $(\boldsymbol{r}, \boldsymbol{p})$, e.g., $A + B \to 2A + C$ denotes the reaction $(\{A, B\}, \{2A, C\})$.

Definition 2.1. *A* chemical reaction network (CRN) *is an ordered pair* $\mathcal{N} = (\Lambda, R)$ *with* Λ *a finite set and* R *a finite set of reactions over* Λ.

The elements of Λ are called the *species* of $\mathcal{N}$. The elements of $\mathbb{N}^\Lambda$ are called the *configurations* of $\mathcal{N}$. Viewing $\boldsymbol{c}$ as a multiset, each element of $\boldsymbol{c}$ is called a *molecule*. For $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{N}^\Lambda$, we write $\boldsymbol{c} \Rightarrow_\mathcal{N} \boldsymbol{c}'$ if there is a reaction $\alpha = (\boldsymbol{r}, \boldsymbol{p}) \in R$ such that $\boldsymbol{r} \leq \boldsymbol{c}$ and $\boldsymbol{c}' = \boldsymbol{c} - \boldsymbol{r} + \boldsymbol{p}$. The transitive and reflexive closure of $\Rightarrow_\mathcal{N}$ is denoted by $\Rightarrow_\mathcal{N}^*$. If $\mathcal{N}$ is clear from the context, then we simply write $\Rightarrow$ and $\Rightarrow^*$ for $\Rightarrow_\mathcal{N}$ and $\Rightarrow_\mathcal{N}^*$, respectively. If $\boldsymbol{c} \Rightarrow^* \boldsymbol{c}'$, then we say $\boldsymbol{c}'$ is *reachable* from $\boldsymbol{c}$.

For $\boldsymbol{c} \in \mathbb{N}^\Lambda$, we define $\mathsf{pre}_\mathcal{N}(\boldsymbol{c}) = \{\boldsymbol{c}' \in \mathbb{N}^\Lambda \mid \boldsymbol{c}' \Rightarrow_\mathcal{N}^* \boldsymbol{c}\}$ and $\mathsf{post}_\mathcal{N}(\boldsymbol{c}) = \{\boldsymbol{c}' \in \mathbb{N}^\Lambda \mid \boldsymbol{c} \Rightarrow_\mathcal{N}^* \boldsymbol{c}'\}$. Again we omit the subscript $\mathcal{N}$ if the CRN $\mathcal{N}$ is clear from the context. Note that for $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{N}^\Lambda$, we have $\boldsymbol{c} \in \mathsf{pre}(\boldsymbol{c}')$ if and only if $\boldsymbol{c}' \in \mathsf{post}(\boldsymbol{c})$ if and only if $\boldsymbol{c} \Rightarrow^* \boldsymbol{c}'$. We extend $\mathsf{pre}(\boldsymbol{c})$ and $\mathsf{post}(\boldsymbol{c})$ to sets $X \subseteq \mathbb{N}^\Lambda$ in the natural way: $\mathsf{pre}(X) = \bigcup_{\boldsymbol{c} \in X} \mathsf{pre}(\boldsymbol{c})$ and $\mathsf{post}(X) = \bigcup_{\boldsymbol{c} \in X} \mathsf{post}(\boldsymbol{c})$.

Petri net theory is a very well established theory of concurrent computation [20]. We recall here that CRNs are essentially equivalent to Petri nets. In Petri net terminology, molecules are called "tokens", species are called "places", reactions are called "transitions", and configurations are called "markings". Due to this correspondence, we can apply results from Petri net theory to CRNs (which we will do in this paper, cf. Theorem 3.2). Conversely, the results shown in this paper can be reformulated straightforwardly in terms of Petri nets. Vector addition systems [17] form a model nearly equivalent to CRNs and Petri nets, where reactions roughly correspond to vectors with integer entries.[5] In the special case of population protocols [3], each reaction $\alpha = (\boldsymbol{r}, \boldsymbol{p})$ obeys $\|\boldsymbol{r}\| = \|\boldsymbol{p}\| = 2$. As a result, for each configuration $\boldsymbol{c}$ of a population protocol, both $\mathsf{pre}(\boldsymbol{c})$ and $\mathsf{post}(\boldsymbol{c})$ are finite (because there are only a finite number of configurations $\boldsymbol{c}'$ with $\|\boldsymbol{c}'\| = \|\boldsymbol{c}\|$). In that model, molecules are called "agents", species are called "states", and reactions are called "transitions".

2.2 Symmetric output-stable deciders

We now recall how one can compute using CRNs. Say we want to decide whether or not the number n of molecules of species X is even. One way to do this is by introducing the reaction $X + X \to \varnothing$.[6] If n is even, then eventually all molecules are consumed, and if n is odd, then eventually there is exactly one molecule of species X present. Once the CRN has stabilized, the presence of a molecule of species X signals that n is odd (i.e., there were an odd number of molecules

---

[5] The only difference is *catalysts*: reactants that are also products, e.g., $C + X \to C + Y$, are allowed in CRNs and Petri nets but not in vector addition systems. Most results for these models are insensitive to this difference.

[6] Notation $\varnothing$ indicates that this reaction has no products.

of species X present initially). Note that in this example there is no molecule of any species that signals that n is even. One may think of a more elaborate example where the presence of say, a molecule of species V_{even}, signals (once the CRN has stabilized) that n is even. In this way, once the CRN has stabilized, X "votes" that n is odd, while V_{even} "votes" that n is even.

A chemical reaction decider $\mathcal{D}$ (introduced in [8]) is a reformulation in terms of CRNs of the notion of population protocol [3] from the field of distributed computing. We define a set of input configurations $\mathcal{I}$ and two sets of "trap configurations", called *output-stable* configurations, $\mathcal{O}_0$ and $\mathcal{O}_1$. We then say that $\mathcal{D}$ is *output-stable* and *decides* the set $\mathcal{I}_1 \subseteq \mathcal{I}$ (with $\mathcal{I}_0 = \mathcal{I} \setminus \mathcal{I}_1$) if for each $i \in \{0, 1\}$ (1) starting from a configuration in $\mathcal{I}_i$, the CRN remains always within reach of a configuration in $\mathcal{O}_i$ (i.e., $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$), and (2) once a configuration is in $\mathcal{O}_i$, it is stuck in $\mathcal{O}_i$ (i.e., $\mathsf{post}(\mathcal{O}_i) = \mathcal{O}_i$).

The sets $\mathcal{I}$, $\mathcal{O}_0$, and $\mathcal{O}_1$ are all of a specific form. There is a subset of *input species* $\Sigma \subseteq \Lambda$; $\mathcal{I}$ consists of nonzero configurations where the all molecules present are in Σ. The output is based on consensus: all the molecules present in an output configuration must agree on the output. More precisely, there is a partition $\{\Gamma_0, \Gamma_1\}$ of Λ (called *0-voters* and *1-voters*, respectively),[7] such that configuration $\boldsymbol{c}$ has *output* $i \in \{0, 1\}$ if all molecules present in $\boldsymbol{c}$ are from Γ_i (i.e., $\boldsymbol{c}\!\restriction_{\Gamma_{1-i}}=\boldsymbol{0}$) and $\boldsymbol{c} \neq \boldsymbol{0}$). A configuration $\boldsymbol{o}$ is defined to be in $\mathcal{O}_i$ — it is *output-stable* — if all configurations of $\mathsf{post}(\boldsymbol{o})$ also have output i.

Our definition, though equivalent, is phrased differently from the usual one [3], being defined in terms of $\mathcal{I}$, $\mathcal{O}_0$, and $\mathcal{O}_1$ instead of Σ, Γ_0, and Γ_1. This simplifies our generalization of this notion in Section 3.

Definition 2.2. *A symmetric output stable chemical reaction decider (sym-CRD) is a 4-tuple $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$, where $\mathcal{N} = (\Lambda, R)$ is a CRN and there are $\Sigma \subseteq \Lambda$ and a partition $\{\Gamma_0, \Gamma_1\}$ of Λ such that*

1. *$\mathcal{I} = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}\!\restriction_{\Lambda \setminus \Sigma}=\boldsymbol{0}\} \setminus \{\boldsymbol{0}\}$,*
2. *$\mathcal{O}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{L}_i \setminus \mathcal{L}_{1-i}\}$, with $\mathcal{L}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}\!\restriction_{\Gamma_i} \neq \boldsymbol{0}\}$ for $i \in \{0, 1\}$.*
3. *There is a partition $\{\mathcal{I}_0, \mathcal{I}_1\}$ of $\mathcal{I}$ such that $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$ for $i \in \{0, 1\}$.*

Condition 1 states that only species in Σ may be present initially, and at least one must be present. Condition 2 defines $\mathcal{L}_i$ to be configurations with an i-voter, so those in $\mathcal{L}_i \setminus \mathcal{L}_{1-i}$ unanimously vote i, and those in $\mathcal{O}_i$ are stable ("stuck" in the set $\mathcal{L}_i \setminus \mathcal{L}_{1-i}$). Condition 3 states that from every configuration reachable from an initial configuration, a "correct" output stable configuration is reachable from there; this is the usual way of expressing stable computation [6,8]. The relationships between these sets are depicted in Figure 1.

Remark 2.3. A different definition is found in [8] and a number of other papers. That definition relaxes ours in two ways: (1) having both voting and non-voting

---

[7] The definition of [8] allows only a subset of Λ to be voters, i.e., $\Gamma_0 \cup \Gamma_1 \subseteq \Lambda$. This convention is more easily shown to define equivalent computational power than our main results about asymmetric and democratic voting.

species, (2) allowing non-input species in the input configuration (e.g., $\{1N\}$ in the Introduction). It turns out that (1) does not affect the computational power of the model. It is also known [3] that (2) does not alter the computational power (though it may affect the time complexity [5, 12]).
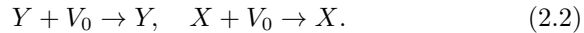
Remark 2.4. We can equivalently define $\mathcal{O}_i = \mathbb{N}^\Lambda \setminus \mathsf{pre}(\mathcal{L}_{1-i} \cup \{\mathbf{0}\})$, a form that will be useful later. To see that this definition is equivalent, observe that $\mathbb{N}^\Lambda \setminus \mathcal{O}_i$ is the set of configurations from which it is possible either to reach $\mathcal{L}_{1-i}$, or to reach *outside* of $\mathcal{L}_i$, and the only point outside *both* is $\mathbf{0}$, so $\mathbb{N}^\Lambda \setminus \mathcal{O}_i = \mathsf{pre}(\mathcal{L}_{1-i} \cup \{\mathbf{0}\})$. Thus $\mathcal{O}_i = \mathbb{N}^\Lambda \setminus \mathsf{pre}(\mathcal{L}_{1-i} \cup \{\mathbf{0}\})$.

Remark 2.5. The $\mathcal{O}_i$ are disjoint and closed under application of reactions: $\mathcal{O}_0 \cap \mathcal{O}_1 = \varnothing$ and $\mathsf{post}(\mathcal{O}_i) = \mathcal{O}_i$.
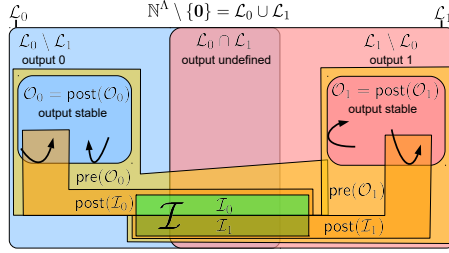
Remark 2.6. Definition 2.2 implies the (weaker) condition that $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$. This can be shown as follows. First, $\mathcal{I}_i \subseteq \mathcal{I}$ and $\mathcal{I}_i \subseteq \mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$, so $\mathcal{I}_i \subseteq \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$. To see the reverse containment, assume $\mathbf{c} \in \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$, but $\mathbf{c} \notin \mathcal{I}_i$, i.e., $\mathbf{c} \in \mathcal{I}_{1-i} \cap \mathsf{pre}(\mathcal{O}_i)$. Let $\mathbf{o} \in \mathsf{post}(\mathbf{c})$ be such that $\mathbf{o} \in \mathcal{O}_i$; such $\mathbf{o}$ exists since $\mathbf{c} \in \mathsf{pre}(\mathcal{O}_i)$. Since $\mathbf{o} \in \mathsf{post}(\mathcal{I}_{1-i}) \subseteq \mathsf{pre}(\mathcal{O}_{1-i})$, we have $\mathbf{o} \in \mathcal{O}_i \cap \mathsf{pre}(\mathcal{O}_{1-i})$. Let $\mathbf{o}' \in \mathsf{post}(\mathbf{o})$ such that $\mathbf{o}' \in \mathcal{O}_{1-i}$. Then $\mathbf{o}' \in \mathsf{post}(\mathcal{O}_i) \cap \mathcal{O}_{1-i}$ — a contradiction because $\mathsf{post}(\mathcal{O}_i) = \mathcal{O}_i$ is disjoint from $\mathcal{O}_{1-i}$.

Since $\mathcal{I}_0 = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_0)$ and $\mathcal{I}_1 = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_1)$ are disjoint, we say that a sym-CRD $\mathcal{D}$ *decides* the set $\mathcal{I}_1$. If a sym-CRD $\mathcal{D}$ decides the set $X \subseteq \mathbb{N}^\Lambda$, then the entries indexed by $\Lambda \setminus \Sigma$ are zero for each $\mathbf{c} \in X$. Therefore, by abuse of notation, we also say that $\mathcal{D}$ *decides* the set $X{\restriction}_\Sigma \subseteq \mathbb{N}^\Sigma$. We will use this convention for all chemical reaction deciders with $\mathcal{I}$ of the given form.

Fig. 1. Venn diagram of configurations that define sym-CRD. Subset relationships depicted in their most general form: $\mathcal{I}_i \subseteq \mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$, and $\mathcal{O}_i \subseteq \mathcal{L}_i \setminus \mathcal{L}_{1-i}$. $\mathsf{pre}(\mathcal{O}_0)$ and $\mathsf{pre}(\mathcal{O}_1)$ partition the set $\mathcal{I} = \mathcal{I}_0 \cup \mathcal{I}_1$.

Example 2.7. We construct a sym-CRD $\mathcal{D}$ that decides the set $x \not\equiv y \mod m$ where x and y are non-negative integer variables, not both zero, and $m \geq 2$ is an integer constant. The variables x and y represent initial counts of species X and Y, respectively. Let $\Sigma = \{X, Y\}$, $\Gamma_0 = \{V_0\}$, $\Gamma_1 = \{X, Y\}$, and $\Lambda = \Gamma_0 \cup \Gamma_1$ be as in Definition 2.2, with the following reactions:

$$mX \to V_0, \quad mY \to V_0, \quad X + Y \to V_0, \tag{2.1}$$

$$Y + V_0 \to Y, \quad X + V_0 \to X. \tag{2.2}$$

We argue that $\mathcal{D}$ decides the set $\{\mathbf{c} \in \mathbb{N}^\Sigma \setminus \{\mathbf{0}\} \mid \mathbf{c}(X) \not\equiv \mathbf{c}(Y) \mod m\}$. Indeed, if $x \equiv y \mod m$, then eventually all X and Y molecules are consumed by the reactions of (2.1). The last time one of these reactions occurs introduces

a V_0 molecule (there is a last reaction since x and y are not both zero). So eventually we obtain a configuration $\boldsymbol{c} \in \mathcal{L}_0 \setminus \mathcal{L}_1$ for which no reaction can be applied anymore. Thus $\boldsymbol{c} \in \mathcal{O}_0$. If $x \not\equiv y \mod m$, then eventually we reach a configuration with one of X or Y, but not both, remaining. The remaining X or Y molecules consume all V_0 molecules by the reactions of (2.2), without the possibility of producing any more. So eventually we obtain a configuration $\boldsymbol{c}' \in \mathcal{L}_1 \setminus \mathcal{L}_0$ for which no reaction can be applied anymore. Thus $\boldsymbol{c}' \in \mathcal{O}_1$.

2.3   Semilinear sets

We say that $X \subseteq \mathbb{N}^\Lambda$ is *linear* if there is a finite set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{N}^\Lambda$ and $\boldsymbol{b} \in \mathbb{N}^\Lambda$ such that $X = \{\boldsymbol{b} + \sum_{i=1}^k n_i \boldsymbol{v}_i \mid n_1, \ldots, n_k \in \mathbb{N}\}$. We say that $X \subseteq \mathbb{N}^\Lambda$ is *semilinear* if X is the union of a finite number of linear sets. Semilinear sets are precisely the sets definable in Presburger arithmetic, which is the first-order theory of natural numbers with addition. As a consequence, the class of semilinear sets is closed under union, intersection, complementation, and projection [15]. An useful characterization of semilinear sets is that they are exactly the sets expressible as finite unions, intersections, and complements of sets of one of the following two forms: *threshold sets* of the form $\{\boldsymbol{x} \mid \sum_{i=1}^k a_i \cdot \boldsymbol{x}(i) < b\}$ for some constants $a_1, \ldots, a_k, b \in \mathbb{Z}$ or *mod sets* of the form $\{\boldsymbol{x} \mid \sum_{i=1}^k a_i \cdot \boldsymbol{x}(i) \equiv b \mod c\}$ for some constants $a_1, \ldots, a_k \in \mathbb{Z}$ and $b, c \in \mathbb{N}$.

The following result was shown in [3, 4]. In fact, the result was shown for output-stable population protocols, which form a subclass of the sym-CRDs. However, the proof is sufficiently general to hold for sym-CRDs as well.[8]

Theorem 2.8 ([3, 4]). *Let $X \subseteq \mathbb{N}^\Sigma \setminus \{\boldsymbol{0}\}$. Then X is semilinear if and only if there is a sym-CRD that decides X.*

For a configuration $\boldsymbol{c} \in \mathbb{N}^\Sigma$, $\mathsf{pre}(\boldsymbol{c})$ and $\mathsf{post}(\boldsymbol{c})$ are in general *not* semilinear [16]. Hence the semilinearity of Theorem 2.8 is due to additional "computational structure" of a sym-CRD. We repeatedly use the following notion of upwards closure to prove that certain sets are semilinear. The results below were shown or implicit in earlier papers [4, 10]. We say $X \subseteq \mathbb{N}^\Lambda$ is *closed upwards* if, for all $\boldsymbol{c} \in X$, $\boldsymbol{c}' \geq \boldsymbol{c}$ implies $\boldsymbol{c}' \in X$.

Lemma 2.9. *Every closed upwards set $X \subseteq \mathbb{N}^\Lambda$ is semilinear.*

Lemma 2.10. *If $X \subseteq \mathbb{N}^\Lambda$ is closed upwards, then so are $\mathsf{pre}(X)$ and $\mathsf{post}(X)$.*

Our results require $\mathsf{pre}(\boldsymbol{0})$ to be semilinear. Observe that $\mathsf{pre}(\boldsymbol{0}) = \{\boldsymbol{0}\}$ if and only if for each reaction $\alpha = (\boldsymbol{r}, \boldsymbol{p})$, $\boldsymbol{p} \neq \boldsymbol{0}$. The next lemma shows that we can assume this holds for sym-CRDs without loss of generality.

---

[8] Indeed, the negative result of [4] that sym-CRDs decide only semilinear sets is more general than stated in Theorem 2.8, applying to *any* reachability relation $\Rightarrow^*$ on $\mathbb{N}^\Lambda$ that is reflexive, transitive, and "additive" ($\boldsymbol{x} \Rightarrow^* \boldsymbol{y}$ implies $\boldsymbol{x} + \boldsymbol{c} \Rightarrow^* \boldsymbol{y} + \boldsymbol{c}$). Also, the negative result of [4] implicitly assumes that the zero vector $\boldsymbol{0}$ is not reachable (i.e., $\mathsf{pre}(\boldsymbol{0}) = \{\boldsymbol{0}\}$). This assumption is manifest for population protocols (if the population size is non-zero). For CRNs, this assumption can be readily removed; see Lemma 2.11.
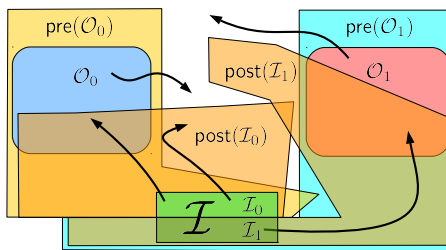
Lemma 2.11. *For every sym-CRD $\mathcal{D}$, there is a sym-CRD $\mathcal{D}'$ deciding the same set such that, for each reaction $\alpha = (\boldsymbol{r}, \boldsymbol{p})$ of $\mathcal{D}'$, $\boldsymbol{p} \neq \boldsymbol{0}$.*

3   Generalized chemical reaction deciders

In this section, we formulate a more generalized definition of CRDs that captures the original "symmetric" definition (sym-CRD) in Section 2.2 and the new "asymmetric" definition (asym-CRD) in Section 4, as well as the "democratic" definition (dem-CRD) in Section 5. In this section we show how to use a result of [13] to re-prove the result of Angluin, Aspnes, and Eisenstat [4] that sym-CRDs decide only semilinear sets. This is a warmup to our main results, shown in Sections 4 and 5, that asym-CRDs and dem-CRDs decide exactly the semilinear sets.

In the generalized notion defined below we have dropped the specific structure of $\mathcal{I}$, $\mathcal{O}_0$, and $\mathcal{O}_1$ (they are now arbitrary subsets of $\mathbb{N}^\Lambda$) and we have replaced the requirement that $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$ by the weaker condition that $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$ (recall Remark 2.6). Also, we do not use the term "stable" in reference to this generalized notion, since there is no requirement that the output configurations $\mathcal{O}_i$ be closed under application of reactions (i.e., we allow $\mathcal{O}_i \subsetneq \mathsf{post}(\mathcal{O}_i)$).

Fig. 2. Venn diagram of configurations that define generalized chemical reaction decider (gen-CRD). Like sym-CRD, $\mathsf{pre}(\mathcal{O}_0)$ and $\mathsf{pre}(\mathcal{O}_1)$ partition the input set $\mathcal{I} = \mathcal{I}_0 \cup \mathcal{I}_1$. Differences with sym-CRD: 1) Possibly $\mathcal{O}_i \subsetneq \mathsf{post}(\mathcal{O}_i)$ (output is not necessarily "stable"). 2) Although $\mathcal{I}_i \subseteq \mathsf{pre}(\mathcal{O}_i)$ (correct output reachable initially), yet possibly $\mathsf{post}(\mathcal{I}_i) \not\subseteq \mathsf{pre}(\mathcal{O}_i)$ (correct output could become unreachable).

The relationships among the sets relevant to the definition below are depicted in Figure 2.

Definition 3.1. *A* generalized chemical reaction decider (gen-CRD) *is a 4-tuple $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$, where $\mathcal{N} = (\Lambda, R)$ is a CRN, $\mathcal{I}, \mathcal{O}_0, \mathcal{O}_1 \subseteq \mathbb{N}^\Lambda$, and there is a partition $\{\mathcal{I}_0, \mathcal{I}_1\}$ of $\mathcal{I}$ such that $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$ for $i \in \{0, 1\}$.*

Observe that every sym-CRD is a gen-CRD. However, the requirements to be a gen-CRD are weaker than for sym-CRDs: (1) the condition $\mathsf{post}(\mathcal{O}_i) = \mathcal{O}_i$ need not hold for gen-CRDs, so it may be possible to "escape" from $\mathcal{O}_i$, and (2) since $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$ need not hold for gen-CRDs, it is possible to take a "wrong" route starting from $\mathcal{I}_i$ such that $\mathcal{O}_i$ becomes unreachable.[9]

Despite these relaxations, observe that the following property of sym-CRDs is retained in gen-CRDs: $\mathcal{I}$ is the disjoint union of $\mathcal{I}_0 = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_0)$ and $\mathcal{I}_1 = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_1)$, *i.e.*, from each input configuration, *exactly one* of the two output sets $\mathcal{O}_0$ or $\mathcal{O}_1$ is reachable. We say that a gen-CRD $\mathcal{D}$ *decides* the set $\mathcal{I}_1$.

---

[9] While Definition 3.1 appears almost too general to be useful, Corollary 3.3 says that if $\mathcal{I}, \mathcal{O}_0, \mathcal{O}_1$ are semilinear, then so are $\mathcal{I}_0, \mathcal{I}_1$, which implies that any CRD definition that can be framed as such a gen-CRD must decide only semilinear sets.

Definition 3.1 is inspired by the following key Petri net result from [13, Theorem 10] (formulated here in terms of CRNs).

Theorem 3.2 ([13]). *Let $\mathcal{N}$ be a CRN and $\mathcal{O}_0, \mathcal{O}_1, \mathcal{I} \subseteq \mathbb{N}^\Lambda$ be semilinear. If $\{\mathcal{I}_0, \mathcal{I}_1\}$ is a partition of $\mathcal{I}$ with $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$ for $i \in \{0,1\}$, then $\mathcal{I}_0$ and $\mathcal{I}_1$ are semilinear.*

We say that a gen-CRD $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ is *semilinear* if $\mathcal{I}$, $\mathcal{O}_0$, and $\mathcal{O}_1$ are all semilinear. We immediately have the following corollary to Theorem 3.2.

Corollary 3.3. *If a semilinear gen-CRD decides $X \subseteq \mathbb{N}^\Lambda$, then X is semilinear.*

As a by-product of the results shown in [13], the reverse direction of Theorem 2.8 (which is the most difficult implication) was reproven in [13] for the case of population protocols. That proof however essentially uses the fact that, for population protocols, $\mathsf{post}(\boldsymbol{c})$ is finite for all configurations $\boldsymbol{c}$, which is not true for CRNs in general. Fortunately, one may still obtain the full reverse direction of Theorem 2.8 by showing that every sym-CRD is semilinear (cf. the proof of Theorem 3.4 below) and then invoking Corollary 3.3.

We now use this machinery to re-prove the result, due originally to Angluin, Aspnes, and Eisenstat [4], that sym-CRDs decide only semilinear sets.

Theorem 3.4. *Every sym-CRD decides a semilinear set.*

Proof. Let $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ be a sym-CRD. Let $\mathcal{I}' = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}{\restriction}_{\Lambda \setminus \Sigma} = \boldsymbol{0}\}$. The complement of $\mathcal{I}'$ is closed upwards, thus $\mathcal{I}'$ is semilinear, as is $\mathcal{I} = \mathcal{I}' \setminus \{\boldsymbol{0}\}$.

We now show that each $\mathcal{O}_i$ is semilinear. Let $\mathcal{L}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}{\restriction}_{\Gamma_i} \neq \boldsymbol{0}\}$ as in Definition 2.2. By Remark 2.4, $\mathcal{O}_i = \mathbb{N}^\Lambda \setminus \mathsf{pre}(\mathcal{L}_{1-i} \cup \{\boldsymbol{0}\}) = \mathbb{N}^\Lambda \setminus (\mathsf{pre}(\mathcal{L}_{1-i}) \cup \mathsf{pre}(\boldsymbol{0}))$. By Lemma 2.11 we may assume that each reaction $\alpha = (\boldsymbol{r}, \boldsymbol{p})$ of $\mathcal{D}$ has $\boldsymbol{p} \neq \boldsymbol{0}$, so $\mathsf{pre}(\boldsymbol{0}) = \{\boldsymbol{0}\}$, which is semilinear. Since $\mathcal{L}_{1-i}$ is closed upwards, by Lemma 2.10, $\mathsf{pre}(\mathcal{L}_{1-i})$ is also closed upwards, so semilinear by Lemma 2.9. Since semilinear sets are closed under union and complement, $\mathcal{O}_i$ is also semilinear, so $\mathcal{D}$ is a semilinear gen-CRD. The theorem follows by Corollary 3.3.   □

Remark 3.5. From the hypothesis $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$ in Definition 2.2, we used only the weaker conclusion $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$. In other words, we need merely that $\mathcal{O}_i$ is *initially* reachable from $\mathcal{I}_i$ itself (*and* that $\mathcal{O}_{1-i}$ is unreachable from $\mathcal{I}_i$, since $\mathsf{pre}(\mathcal{O}_0)$ and $\mathsf{pre}(\mathcal{O}_1)$ partition $\mathcal{I}$). We do not require that $\mathcal{O}_i$ *remains* reachable from every configuration reachable from $\mathcal{I}_i$ (i.e., $\mathsf{post}(\mathcal{I}_i)$). Hence one could weaken part 3 of Definition 2.2 to use the condition $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$, and Theorem 3.4 still holds.[10]

Despite Remark 3.5, if a gen-CRD *does* obey the stronger condition $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$, then a convenient property holds: each $\mathcal{O}_i$ may be enlarged without altering the set $\mathcal{I}_1$ decided by the gen-CRD, so long as $\mathcal{O}_{1-i}$ remains unreachable from $\mathcal{O}_i$. The following lemma formalizes this.

---

[10] In contrast, the proof of [4] crucially requires the hypothesis $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$.

Lemma 3.6. *Let $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ be a gen-CRD that decides $\mathcal{I}_1$ and let $\mathcal{I}_0 = \mathcal{I} \setminus \mathcal{I}_1$. For $i \in \{0,1\}$, assume that $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$, and let $\mathcal{O}'_i \supseteq \mathcal{O}_i$ with $\mathsf{post}(\mathcal{O}'_i) \cap \mathcal{O}_{1-i} = \varnothing$. Then $\mathcal{D}' = (\mathcal{N}, \mathcal{I}, \mathcal{O}'_0, \mathcal{O}'_1)$ is a gen-CRD deciding $\mathcal{I}_1$.*

Proof. We have $\mathcal{I}_i = \mathsf{pre}(\mathcal{O}_i) \cap \mathcal{I} \subseteq \mathsf{pre}(\mathcal{O}'_i) \cap \mathcal{I}$ for $i \in \{0,1\}$. To show that this inclusion is an equality, it suffices to show that $\mathsf{pre}(\mathcal{O}'_0) \cap \mathcal{I}$ and $\mathsf{pre}(\mathcal{O}'_1) \cap \mathcal{I}$ are disjoint.

Let $\boldsymbol{i} \in \mathcal{I}_i$. Then $\boldsymbol{i} \in \mathsf{pre}(\mathcal{O}_i) \subseteq \mathsf{pre}(\mathcal{O}'_i)$. Assume to the contrary $\boldsymbol{i} \in \mathsf{pre}(\mathcal{O}'_{1-i})$. Let $\boldsymbol{o} \in \mathcal{O}'_{1-i} \cap \mathsf{post}(\boldsymbol{i})$, so $\boldsymbol{o} \in \mathsf{post}(\boldsymbol{i}) \subseteq \mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$. Thus $\mathcal{O}'_{1-i} \cap \mathsf{pre}(\mathcal{O}_i) \neq \varnothing$. In other words, $\mathsf{post}(\mathcal{O}'_{1-i}) \cap \mathcal{O}_i \neq \varnothing$ — a contradiction. Hence $\mathsf{pre}(\mathcal{O}'_0) \cap \mathcal{I}$ and $\mathsf{pre}(\mathcal{O}'_1) \cap \mathcal{I}$ are disjoint. $\qquad\square$

4   Asymmetric output-stability

We now give a natural alternative output convention for CRDs, which we call an asymmetric output-stable CRD (asym-CRD). Whereas the output i of a sym-CRD is based on both the presence of species of one type Γ_i and the absence of a species of a different type Γ_{1-i}, the output of an asym-CRD is based solely on the presence or absence of a single species type Γ_1.

For each $\boldsymbol{i} \in \mathcal{I}$ the CRD can either (1) reach a configuration $\boldsymbol{o}$ so that for each configuration $\boldsymbol{o}'$ reachable from $\boldsymbol{o}$ (including $\boldsymbol{o}$ itself) we have $\boldsymbol{o}'\!\restriction_{\Gamma_1} \neq \boldsymbol{0}$ or (2) reach a configuration $\boldsymbol{o}$ so that for each configuration $\boldsymbol{o}'$ reachable from $\boldsymbol{o}$ we have $\boldsymbol{o}'\!\restriction_{\Gamma_1} = \boldsymbol{0}$. Similarly to gen-CRDs, and unlike sym-CRDs,[11] it is not required that such a configuration $\boldsymbol{o}$ is reachable from *any* configuration $\boldsymbol{c}$ reachable from the initial $\boldsymbol{i}$, merely that such a $\boldsymbol{o}$ is reachable from $\boldsymbol{i}$ itself. Even this more liberal assumption does not allow the CRD to decide a non-semilinear set.

Definition 4.1. *An* asymmetric output-stable chemical reaction decider (asym-CRD) *is a gen-CRD $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$, where there are $\Sigma \subseteq \Lambda$ and voting species $\Gamma_1 \subseteq \Lambda$ such that*

1. *$\mathcal{I} = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}\!\restriction_{\Lambda \setminus \Sigma} = \boldsymbol{0}\} \setminus \{\boldsymbol{0}\}$, and*
2. *$\mathcal{O}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{V}_i\}$ for $i \in \{0,1\}$, with $\mathcal{V}_1 = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}\!\restriction_{\Gamma_1} \neq \boldsymbol{0}\}$ and $\mathcal{V}_0 = \mathbb{N}^\Lambda \setminus \mathcal{V}_1$.[12]*

Condition 1 states that only species in Σ may be present initially, and at least one must be present. Condition 2 defines $\mathcal{V}_1$ and $\mathcal{V}_0$ to be configurations with and without Γ_1 voters, and $\mathcal{O}_i$ to be the stable subsets of $\mathcal{V}_i$.

Example 4.2. Consider the following asym-CRD $\mathcal{D}'$, where $\Sigma = \{X, Y\}$ and $\Gamma_1 = \{X, Y\}$, which decides the same set as in Example 2.7 (i.e., $x \not\equiv y \mod m$).

$$mX \to \varnothing, \quad mY \to \varnothing, \quad X + Y \to \varnothing. \tag{4.1}$$

---
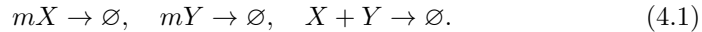
[11] As noted, sym-CRDs could be defined by replacing the requirement $\mathsf{post}(\mathcal{I}_i) \subseteq \mathsf{pre}(\mathcal{O}_i)$ with $\mathcal{I}_i = \mathcal{I} \cap \mathsf{pre}(\mathcal{O}_i)$ and retain the same power, but for clarity we retain the original definition.

[12] Just as for sym-CRDs, $\mathsf{post}(\mathcal{O}_i) = \mathcal{O}_i$. Note that $\mathcal{V}_1$ above is the same as $\mathcal{L}_1$ in Definition 2.2, but $\mathcal{L}_0 \neq \mathcal{V}_0$, since $\mathcal{L}_1$ and $\mathcal{L}_0$ can have nonempty intersection if there are conflicting voters present in some configuration.

If $x \equiv y \mod m$, then eventually all X and Y molecules are consumed and we obtain configuration $\boldsymbol{c} = \boldsymbol{0} \in \mathcal{O}_0$. Otherwise, all X and Y cannot be consumed, and we are in $\mathcal{O}_1$. This example illustrates that the asym-CRD computing convention may permit a simpler implementation in some cases. Indeed, compared with Example 2.7, (4.1) has 2 fewer reactions and 1 fewer species (and is also "faster" since fewer reactions need to occur).

We first observe that asym-CRDs have at least the computational power of sym-CRDs.

Observation 4.3. *Let $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ be a sym-CRD deciding X, with voter partition $\{\Gamma_0, \Gamma_1\}$. Then $\mathcal{D}' = (\mathcal{N}, \mathcal{I}, \mathcal{O}'_0, \mathcal{O}'_1)$, where, for $i \in \{0, 1\}$, $\mathcal{O}'_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{V}_i\}$, with $\mathcal{V}_i$ as in Definition 4.1 (with respect to Γ_1), is an asym-CRD deciding X.*

Proof. This follows from Lemma 3.6 since (1) $\mathcal{O}_i \subseteq \mathcal{O}'_i$ and (2) $\mathsf{post}(\mathcal{O}'_i) = \mathcal{O}'_i$ is disjoint from $\mathcal{O}_{1-i}$ for $i \in \{0, 1\}$. $\qquad\square$

We now show that asym-CRDs have *no greater* computational power than sym-CRDs. This is not as immediate as the other direction. First, observe that an asym-CRD may not be a sym-CRD; if we interpret species $V_0 \in \Lambda \setminus \Gamma_1$ as voting "0", then a sym-CRD is required to *eliminate* them to output "1", but not an asym-CRD. Moreover, a direct transformation of an asym-CRD into a sym-CRD appears difficult. Intuitively, the problem is that the absence of molecules in Γ_1 is not detectable by a CRN, so there is no obvious way to ensure that a species $V_0 \in \Lambda \setminus \Gamma_1$ is produced only if all $V_1 \in \Gamma_1$ are absent. The next obvious proof strategy would be to show, as in the proof of Theorem 3.4, that every asym-CRD is a semilinear gen-CRD. However, it is not clear whether $\mathcal{O}_1$ is semilinear. Nonetheless, due to the generality of Definition 3.1 and Theorem 3.2, we can define a semilinear gen-CRD that decides the same set, by taking a subset of $\mathcal{O}_1$ that is provably semilinear and still satisfies the necessary reachability constraints, even though the gen-CRD we define is *not* in fact an asym-CRD (in particular, its "output" set $\mathcal{O}_1$ is not closed under application of reactions).

Recall that a *homomorphism* $f : \mathbb{N}^\Lambda \to \mathbb{Z}$ obeys $f(\boldsymbol{c} + \boldsymbol{c}') = f(\boldsymbol{c}) + f(\boldsymbol{c}')$ for all $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{N}^\Lambda$. Some examples include $f(\boldsymbol{c}) = \boldsymbol{c}(S)$ for some $S \in \Lambda$, $f(\boldsymbol{c}) = \|\boldsymbol{c}{\restriction}_\Delta\|$ for some $\Delta \subseteq \Lambda$, or $f(\boldsymbol{c}) = \boldsymbol{c}(S_1) - \boldsymbol{c}(S_2)$ for some $S_1, S_2 \in \Lambda$.

For a CRN $\mathcal{N}$ and a function $f : \mathbb{N}^\Lambda \to \mathbb{Z}$, we define $\mathsf{nondec}_f = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \forall \boldsymbol{c}' \in \mathsf{post}(\boldsymbol{c}), f(\boldsymbol{c}') \geq f(\boldsymbol{c})\}$ as the set of configurations $\boldsymbol{c}$ in which f is minimal among all the configurations reachable from $\boldsymbol{c}$.

We now prove a key lemma, which will be used for characterizing both asym-CRDs in this section and dem-CRDs in Section 5.

Lemma 4.4. *Let $\mathcal{N}$ be a CRN and $f : \mathbb{N}^\Lambda \to \mathbb{Z}$ a homomorphism. Let $\mathcal{O} = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{V}\}$ with $\mathcal{V} = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid f(\boldsymbol{c}) > 0\}$. Then $\mathcal{O} \cap W$ is semilinear and $\mathsf{pre}(\mathcal{O} \cap W) = \mathsf{pre}(\mathcal{O})$, where $W = \mathsf{nondec}_f$.*

Proof. We first prove $\mathsf{pre}(\mathcal{O} \cap W) = \mathsf{pre}(\mathcal{O})$. Obviously, $\mathsf{pre}(\mathcal{O} \cap W) \subseteq \mathsf{pre}(\mathcal{O})$. To prove the reverse containment, let $\boldsymbol{c} \in \mathsf{pre}(\mathcal{O})$. Hence $\boldsymbol{c} \in \mathsf{pre}(\boldsymbol{o})$ for some $\boldsymbol{o} \in \mathcal{O}$.

Since every $o' \in \mathsf{post}(o)$ satisfies $f(o') > 0$, there is $o' \in \mathsf{post}(o)$ such that $f(o')$ is minimal among all configurations in $\mathsf{post}(o)$. Thus $o' \in W$. Since $\mathsf{post}(\mathcal{O}) = \mathcal{O}$, we have $o' \in \mathcal{O}$. Hence, $o' \in \mathcal{O} \cap W$. Now, $o \in \mathsf{pre}(o')$ and $c \in \mathsf{pre}(o)$, and so $c \in \mathsf{pre}(o')$. Therefore, $c \in \mathsf{pre}(\mathcal{O} \cap W)$, so $\mathsf{pre}(\mathcal{O}) \subseteq \mathsf{pre}(\mathcal{O} \cap W)$.

We now show that $\mathcal{O} \cap W$ is semilinear. Observe that the set $\mathbb{N}^\Lambda \setminus W = \{c \in \mathbb{N}^\Lambda \mid \exists c' \in \mathsf{post}(c), f(c') < f(c)\}$ is closed upwards. Indeed, if $c \in \mathbb{N}^\Lambda \setminus W$ and $c' \in \mathsf{post}(c)$ with $f(c') < f(c)$, then for all $d \in \mathbb{N}^\Lambda$, $c' + d \in \mathsf{post}(c + d)$ and $f(c' + d) = f(c') + f(d) < f(c) + f(d) = f(c + d)$. Thus $\mathbb{N}^\Lambda \setminus W$ is semilinear by Lemma 2.9, and hence also W. Since $\mathcal{O} \subseteq \mathcal{V}$, we have $\mathcal{O} \cap W \subseteq \mathcal{V} \cap W$. Conversely, if $c \in \mathcal{V} \cap W$, then $f(c) > 0$ since $c \in \mathcal{V}$, and for all $c' \in \mathsf{post}(c)$, $f(c') \geq f(c) > 0$ since $c \in W$. Thus $c \in \mathcal{O} \cap W$, showing $\mathcal{O} \cap W = \mathcal{V} \cap W$, which is semilinear since $\mathcal{V}$ and W are. $\square$

Using Lemma 4.4 we show that every asym-CRD can be changed into a semilinear gen-CRD by choosing $\mathcal{O}_1 \cap W$, rather than $\mathcal{O}_1$, as its "output 1" set of configurations. Note that unlike in the definition of sym-CRD and asym-CRD, $\mathcal{O}_1 \cap W$ is *not* in general closed under application of reactions.

Lemma 4.5. *Let $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ be an asym-CRD deciding X and Γ_1 be as in Definition 4.1. Let $W = \mathsf{nondec}_f(\Gamma_1)$ with $f : \mathbb{N}^\Lambda \to \mathbb{Z}$ defined as $f(c) = \|c{\restriction}_{\Gamma_1}\|$ for all $c \in \mathbb{N}^\Lambda$. Then $\mathcal{D}' = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1 \cap W)$ is a semilinear gen-CRD deciding X.*

Proof. Observe that f is a homomorphism. Now, Lemma 4.4 tells us that $\mathsf{pre}(\mathcal{O}_1 \cap W) = \mathsf{pre}(\mathcal{O}_1)$; thus $\mathcal{D}'$ decides X.

To complete the proof, it suffices to show that $\mathcal{D}'$ is semilinear. $\mathcal{I}$ is obtained from the closed-upwards set $\mathbb{N}^\Sigma \setminus \{0\}$ by padding zeros for the species of $\Lambda \setminus \Sigma$, so $\mathcal{I}$ is semilinear. $\mathcal{O}_1 \cap W$ is semilinear by Lemma 4.4. To see that $\mathcal{O}_0$ is semilinear, let $\mathcal{V}_0$ and $\mathcal{V}_1$ be as in Definition 4.1. Clearly $\mathcal{V}_1$ is closed upwards, so semilinear. So, (1) $\mathsf{pre}(\mathcal{V}_1)$ is also closed upwards and therefore semilinear (by Lemma 2.10 and Lemma 2.9) and (2) $\mathcal{V}_0 = \mathbb{N}^\Lambda \setminus \mathcal{V}_1$ is semilinear. Thus, $\mathcal{O}_0 = \mathcal{V}_0 \setminus \mathsf{pre}(\mathcal{V}_1)$ is semilinear since the class of semilinear sets is closed under set difference. $\square$

The following is the first of two main results of this paper. It says that the computational power of sym-CRDs equals that of asym-CRDs; they both decide exactly the semilinear sets.

Theorem 4.6. *Let $X \subseteq \mathbb{N}^\Sigma \setminus \{0\}$. Then X is semilinear if and only if there is an asym-CRD that decides X.*

Proof. The forward direction follows from Observation 4.3 and Theorem 2.8. For the reverse direction, let $\mathcal{D}$ be an asym-CRD deciding X. By Lemma 4.5, there is a semilinear gen-CRD $\mathcal{D}'$ deciding X, which is semilinear by Corollary 3.3. $\square$

5   Democratic output-stability

Another reasonable alternative output convention is the one most naturally associated with the term "voting": a *democratic* output convention in which, rather

than requiring a consensus, we define output by majority vote. In this case, for sets of voting species Γ_0 and Γ_1, the only undefined outputs occur in "tie" configurations $\boldsymbol{c}$ where $\|\boldsymbol{c}{\restriction}_{\Gamma_0}\| = \|\boldsymbol{c}{\restriction}_{\Gamma_1}\|$. In this section we show that such CRDs have equivalent computing power to sym-CRDs.

Definition 5.1. *A* democratic output-stable chemical reaction decider (dem-CRD) *is a gen-CRD* $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$, *where there are* $\Sigma \subseteq \Lambda$ *and a partition* $\{\Gamma_0, \Gamma_1\}$ *of* Λ *such that*

1. $\mathcal{I} = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \boldsymbol{c}{\restriction}_{\Lambda \setminus \Sigma} = \boldsymbol{0}\} \setminus \{\boldsymbol{0}\}$,
2. $\mathcal{O}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{M}_i\}$, *with* $\mathcal{M}_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \|\boldsymbol{c}{\restriction}_{\Gamma_i}\| > \|\boldsymbol{c}{\restriction}_{\Gamma_{1-i}}\|\}$ *for* $i \in \{0, 1\}$.

Note that $\mathcal{M}_0 \cap \mathcal{M}_1 = \varnothing$, and that $\mathcal{O}_i$ is stable, i.e., $\mathcal{O}_i = \mathsf{post}(\mathcal{O}_i)$. A sym-CRD reaches a consensus, the strongest kind of majority, leading to the following observation implying that dem-CRDs are at least as powerful as sym-CRDs.

Observation 5.2. *Let* $\mathcal{D} = (\mathcal{N}, \mathcal{I}, \mathcal{O}_0, \mathcal{O}_1)$ *be a sym-CRD deciding* X, *with voter partition* $\{\Gamma_0, \Gamma_1\}$. *Then* $\mathcal{D}' = (\mathcal{N}, \mathcal{I}, \mathcal{O}'_0, \mathcal{O}'_1)$, *where* $\mathcal{O}'_i = \{\boldsymbol{c} \in \mathbb{N}^\Lambda \mid \mathsf{post}(\boldsymbol{c}) \subseteq \mathcal{M}_i\}$ *for* $i \in \{0, 1\}$, *with* $\mathcal{M}_i$ *as in Definition 5.1, is a dem-CRD deciding* X.

Proof. This follows from Lemma 3.6 since (1) $\mathcal{O}_i \subseteq \mathcal{O}'_i$ and (2) $\mathsf{post}(\mathcal{O}'_i) = \mathcal{O}'_i$ is disjoint from $\mathcal{O}'_{1-i}$ for $i \in \{0, 1\}$. $\square$

The converse result, that dem-CRDs are no more powerful than sym-CRDs, implies the second main result of this paper. The proof of the following theorem is found in the full version of this paper, and relies on the gen-CRD framework of Section 3 and Lemma 4.4 (choosing f that is the difference between 0 and 1 voters).

Theorem 5.3. *Let* $X \subseteq \mathbb{N}^\Sigma \setminus \{\boldsymbol{0}\}$. *Then* X *is semilinear if and only if there is a dem-CRD that decides* X.

6   Discussion

Using a recent result about Petri nets [13] (cf. Theorem 3.2) we have presented a framework able to capture different output conventions for computational CRNs. The original symmetric consensus-based definition [3] can be fitted in this framework, giving a new proof that such CRNs are limited to computing only semilinear sets. Two additional definitions, an asymmetric existence-based convention, and a symmetric majority-vote convention, can be fitted in this framework, and thus have the same expressive power as the original.

We show that asym-CRDs and dem-CRDs are no more powerful than sym-CRDs by showing that they are limited to deciding semilinear sets, which is known also to apply to sym-CRDs. It would be informative, however, to find a proof that uses a direct simulation argument, showing how to transform an

arbitrary asym-CRD or dem-CRD into a sym-CRD deciding the same set. Along a similar line of thinking, we have defined the computational ability of CRDs without regard to time complexity, which is potentially sensitive to definitional choices, even if the class of decidable sets remains the same [1, 2, 5, 11, 12]. It would be interesting to find cases in which asym-CRDs or dem-CRDs are be able to compute faster than any equivalent sym-CRD.

An open problem is to consider other output conventions, where we possibly step out of semilinearity. For example, consider a designated species V_1 such that for each input configuration $\boldsymbol{d} \in \mathcal{I}$, (1) $\boldsymbol{d} \in \mathcal{I}_1$ if we always eventually reach a configuration $\boldsymbol{c}$ such that all configurations reachable from $\boldsymbol{c}$ has a V_1 molecule, and (2) $\boldsymbol{d} \in \mathcal{I}_0$ if we can never reach such a configuration $\boldsymbol{c}$. Hence the output of a configuration is then based on a behavioral property of the system (whether it is stable) instead of a syntactic property of the configuration (whether it contains a particular molecule). It is not clear how to apply Theorem 3.2, which requires that $\mathcal{I}_0 = \mathcal{I} \cap \mathsf{pre}(S)$ for some semilinear set S.

References

1. Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili, and Ronald L. Rivest. Time-space trade-offs in population protocols. Technical Report 1602.08032, arXiv, 2016.
2. Dan Alistarh and Rati Gelashvili. Polylogarithmic-time leader election in population protocols. In *ICALP 2015: Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming, Kyoto, Japan*, 2015.
3. Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18(4):235–253, 2006.
4. Dana Angluin, James Aspnes, and David Eisenstat. Stably computable predicates are semilinear. In *PODC 2006: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 292–299, New York, NY, USA, 2006. ACM Press.
5. Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(3):183–199, 2008.
6. Dana Angluin, James Aspnes, David Eisenstat, and Eric Ruppert. The computational power of population protocols. *Distributed Computing*, 20(4):279–304, 2007.
7. Robert Brijder. Output stability and semilinear sets in chemical reaction networks and deciders. In *Proceedings of the 20th International Meeting on DNA Computing and Molecular Programming*, pages 100–113, 2014.
8. Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. *Natural Computing*, 13(4):517–534, 2014.
9. Rachel Cummings, David Doty, and David Soloveichik. Probability 1 computation with chemical reaction networks. *Natural Computing*, pages 1–17, 2015.
10. Leonard E. Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American Journal of Mathematics*, 35:413–422, 1913.
11. David Doty and Monir Hajiaghayi. Leaderless deterministic chemical reaction networks. *Natural Computing*, 14(2):213–223, 2015.
12. David Doty and David Soloveichik. Stable leader election in population protocols requires linear time. In *DISC 2015: Proceedings of the 29th International Symposium on Distributed Computing*, Lecture Notes in Computer Science, pages 602–616. Springer Berlin Heidelberg, 2015.
13. Javier Esparza, Pierre Ganty, Jérôme Leroux, and Rupak Majumdar. Verification of Population Protocols. In *26th International Conference on Concurrency Theory (CONCUR 2015)*, volume 42, pages 470–482, 2015.
14. Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

15. Seymour Ginsburg and Edwin H. Spanier. Semigroups, Presburger formulas, and languages. *Pacific Journal of Mathematics*, 16(2):285–296, 1966.
16. John E. Hopcroft and Jean-Jacques Pansiot. On the reachability problem for 5-dimensional vector addition systems. *Theoretical Computer Science*, 8:135–159, 1979.
17. Richard M. Karp and Raymond E. Miller. Parallel program schemata. *Journal of Computer and System Sciences*, 3(2):147–195, 1969.
18. George H Mealy. A method for synthesizing sequential circuits. *Bell System Technical Journal*, 34(5):1045–1079, 1955.
19. Edward F Moore. Gedanken-experiments on sequential machines. *Automata studies*, 34:129–153, 1956.
20. James L. Peterson. Petri nets. *ACM Computing Surveys*, 9(3):223–252, 1977.
21. David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.
22. David Soloveichik, Georg Seelig, and Erik Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, 2010.