

Speed faults in computation by chemical reaction networks*

Ho-Lin Chen[†] Rachel Cummings[‡] David Doty[§] David Soloveichik[¶]

Abstract

Chemical reaction networks (CRNs) formally model chemistry in a well-mixed solution. Assuming a fixed molecular population size and bimolecular reactions, CRNs are formally equivalent to population protocols, a model of distributed computing introduced by Angluin, Aspnes, Diamadi, Fischer, and Peralta (PODC 2004). The challenge of fast computation by CRNs (or population protocols) is to ensure that there is never a bottleneck “slow” reaction that requires two molecules (agent states) to react (communicate), both of which are present in low ($O(1)$) counts. It is known that CRNs can be fast in expectation by avoiding slow reactions with high probability. However, states may be reachable (with low probability) from which the correct answer may only be computed by executing a slow reaction. We deem such an event a *speed fault*. We show that the problems decidable by CRNs guaranteed to avoid speed faults are precisely the *detection problems*: Boolean combinations of questions of the form “is a certain species present or not?”. This implies, for instance, that no speed fault free CRN could decide whether there are at least two molecules of a certain species, although a CRN could decide this in “fast” expected time — i.e. speed fault free CRNs “can’t count.”

Acknowledgements. We thank Damien Woods, Anne Condon, Chris Thachuk, Bonnie Kirkpatrick, Monir Hajiaghayi, and Ján Maňuch for useful discussions.

*The third, and fourth authors were supported by the Molecular Programming Project under NSF grants 0832824 and 1317694, the first author was supported by NSC grant number 101-2221-E-002-122-MY3, the second author was supported by NSF grants CCF-1049899 and CCF-1217770, the third author was supported by a Computing Innovation Fellowship under NSF grant 1019343, NSF grants CCF-1219274 and CCF-1162589, and the fourth author was supported by NIGMS Systems Biology Center grant P50 GM081879.

[†]National Taiwan University, Taipei, Taiwan, holinc@gmail.com

[‡]Northwestern University, Evanston, IL, USA, rachelc@u.northwestern.edu.

[§]California Institute of Technology, Pasadena, CA, USA, ddoty@caltech.edu

[¶]University of California, San Francisco, San Francisco, CA, USA, david.soloveichik@ucsf.edu

1 Introduction

Understanding the principles of molecular computation is essential to making sense of information processing in biological cellular regulatory networks. Further, in engineering of life-like devices (e.g. “wet robots” that can patrol the blood for cancer cells) we are rapidly approaching the point where we are limited by conceptual understanding: How molecular networks can be programmed to process information and carry out computation subject to the natural constraints of aqueous chemistry is still not well-understood.

A foundational model of chemistry commonly used in natural sciences is that of chemical reaction networks (CRNs), i.e., (finite) sets of chemical reactions such as $A + B \rightarrow A + C$. Subject to discrete semantics (integer number of molecules) the model corresponds to a continuous time, discrete state, Markov process [11]. A state of the system is a vector of non-negative integers specifying the molecular counts of the species (e.g., A , B , C), a reaction can occur only when all its reactants are present, and transitions between states correspond to reactions (i.e., when the above reaction occurs the count of B is decreased by 1 and the count of C increased by 1). The transition rate is proportional to the product of the counts of the reactants.¹ CRNs are widely used to describe natural biochemical systems such as the intricate cellular regulatory networks responsible for the information processing within cells. With recent advances in synthetic biology, CRNs are a promising language for the design of artificial biochemical networks. For example, the physical primitive of nucleic-acid strand displacement cascades provides concrete chemical implementations of arbitrary CRNs [4,8,17]. Thus, since in principle any CRN can be built, hypothetical CRNs with interesting behaviors are becoming of more than theoretical interest.

The importance of the CRN model is underscored by the observation that intimately related models repeatedly arise in theoretical computer science under different guises: e.g. vector addition systems [13], Petri nets [15], population protocols [1]. The connection to distributed computing models, in turn, resulted in novel insights regarding natural cellular regulatory networks [5].

Parallelism is a basic attribute of chemistry, and one that is of central importance in understanding molecular information processing. This kind of parallelism is both a blessing and a curse: it can be used to speed up computation, but we must be careful to avoid “race conditions” (reactions happening in an unintended order) which may lead to error.

Consider a very basic task: a chemical system (e.g. cell) responding to molecular signals present in very small quantities. More specifically, say the computation is to produce at least one molecule of Y if and only if there is at least one molecule of species A_1 and at least one molecule of species A_2 . Consider the strategy shown in Fig 1(b). Intuitively, this corresponds to having receptors F that in order to activate need to bind both A_1 and A_2 . By having n receptors F we can increase the rate of the first reaction, but if there is only one molecule of A_1 , there will be at most one molecule of F' and thus the second reaction occurs at a rate independent of the amount of receptor. Thus this scheme is “not parallelizable”.²

A better strategy is to amplify the signal before taking the conjunction: e.g. Fig 1(c). Here the receptors release A_1 back upon interacting with it, and a single A_1 can interact with many receptors (converting them from F to F'). Intuitively, the more receptors F we have, the faster we’ll get a large number of F' ’s, and the faster the Y will get produced via the second reaction. More specifically, observe that starting with $n > 0$ molecules of F , and a molecule of A_1 and A_2 each, the reachable states without Y are: for $0 \leq m \leq n$, $((n - m) F, m F', 1 A_1, 1 A_2)$. From any reachable state without Y , we can reach Y through a sequence of reaction executions where one of the reactants is present in at least $\lfloor n^{1/2} \rfloor$ count,³ and under stochastic chemical kinetics, the expected time to

¹See A.2 for the formal definition of the stochastic model. Since rate constants are inessential for the results in this paper, we assume that all rate constants are 1.

²Bimolecular reaction rates scale inversely with the total volume, and it is impossible to fit arbitrarily many molecules in a fixed volume. While for large enough molecular counts we will run into this finite density constraint, we study the scaling of speed with molecular count before that point is reached. An alternate perspective is that our task is to compute as quickly as possible in volume sufficient to allow molecules of F to fill the volume with constant density [16].

³If $m < \lfloor n^{1/2} \rfloor$, execute the first reaction $\lfloor n^{1/2} \rfloor - m$ times (resulting in $\lfloor n^{1/2} \rfloor$ molecules of F'), and then execute the second reaction. If $m \geq \lfloor n^{1/2} \rfloor$, execute the second reaction.

traverse this path is $O(1/n^{1/2})$ — decreasing with n .⁴ Scheme Fig 1(d) is even faster: it can be shown that from any reachable state, the expected time to produce Y scales as $O(\log n/n)$.

Now consider a slightly different computational task: produce at least one molecule of Y if and only if there are at least 2 molecules of species A . The natural analog of Fig 1(b) fails to be deterministic: the reactions $A + F \rightarrow F'$, $A + F' \rightarrow Y$ suffer from a “race condition” where Y is never produced if both molecules of A happen to react with F . This can be fixed by having the receptor F bind A reversibly⁵ as in Fig. 1(f). However, this scheme is not parallelizeable for the same reason as (b).

The natural analog of the parallelizeable reaction scheme Fig 1(c) will not solve this task correctly at all: With reactions $A + F \rightarrow F' + A$, $A + F' \rightarrow Y$, even a single molecule of A will always lead to a Y .

Also problematic is the scheme shown in Fig 1(g) based on (d). While it is parallelizeable, it also suffers from a race condition that can result in an error. If the two molecules of A happen to react with different receptor types (F_1 and F_2) then Y will be produced. However, if both A 's react with the same receptor type, Y will never be produced.

Informally, our main result is that no CRN is deterministic and parallelizeable at the same time for the “2 A problem” (or any computation that involves counting, rather than simply detecting the presence or absence of input species). Thus deterministic and parallelizeable must be disjoint in Fig. 1(right). Unlike the examples above, we allow a broader range of schemes that could produce and consume Y repeatedly but eventually converge on the presence or absence of Y as the output. In order to define “parallelizeable” formally, we introduce the notion of a “speed fault”. A speed fault occurs if a state is reached such that to stabilize to the correct output from that state requires using a bimolecular reaction with both reactants bounded independently of n . Thus “deterministically parallelizeable” corresponds to speed fault free. Our main result is that the problems decidable by speed fault free CRNs are precisely the *detection problems*: Boolean combinations of questions of the form “is a certain species present or not?”. Thus speed fault free CRNs “can’t count.”

The current work stems from the desire to understand fast deterministic computation in CRNs and population protocols. While sophisticated chemical algorithms and protocols have been developed to compute a large class of functions quickly and without error (see next section), most constructions are not deterministically fast in the same strong sense as they are deterministic. Indeed, deterministic computation is a worst case notion that intuitively ensures correctness no matter what unlucky sequence of reactions occurs. However, fast computation is defined with respect to high probability reaction sequences. Our definition captures the natural worst case notion of speed.⁶

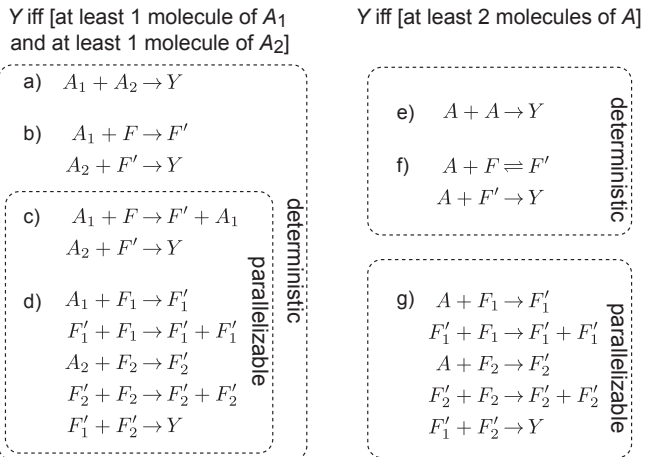


Figure 1: Two molecular computation tasks: predicates “Is there at least 1 molecule of A_1 and at least one molecule of A_2 ?” (left), and “Are there at least 2 molecules of A ?” (right). CRNs (a)-(d) compute the first predicate (left), and CRNs (e)-(g) compute the second (right). Parameter n is the initial amount of F , or F_1 and F_2 species which help in the computation. Informally the parallelizeable CRNs are those that produce the output faster with increasing n . Deterministic CRNs are those that compute correctly no matter what order the reactions happen to occur in. Other strategies (not shown) involve producing Y but consuming it if the predicate is not satisfied.

⁴The rate of a bimolecular reaction is proportional to the product of the counts of the reactants. Thus the expected time from the state with $m < \lfloor n^{1/2} \rfloor$ molecules of F' to reach the state with $\lfloor n^{1/2} \rfloor$ molecules of F' is proportional to $\sum_{i=m}^{\lfloor n^{1/2} \rfloor} 1/(n-i) \leq n^{1/2} \cdot 1/(n-n^{1/2}) = O(1/n^{1/2})$. Finally the rate of the second reaction when there are $\lfloor n^{1/2} \rfloor$ molecules of F' is proportional to $n^{1/2}$ and thus the expected time for it to fire is $O(1/n^{1/2})$ for a total expected time of $O(1/n^{1/2})$. Note that power $n^{1/2}$ was chosen in the analysis to ensure the optimal tradeoff between the rates of individual reaction executions and the total number of reaction executions.

⁵A reversible reaction $A + F \rightleftharpoons F'$ is simply syntactic sugar for two irreversible reactions $A + F \rightarrow F'$ and $F' \rightarrow A + F$.

⁶We observe that in the literature on computation in CRNs and population protocols it is almost never the case that computation is slow because the necessary sequence of reactions is too long — rather, slowdown is dominated by reaction bottlenecks where two low count species must react. Thus in this work we focus on this essential type of delay,

Our positive result shows how any detection problem can be decided by a speed fault free CRN, and further shows that this computation is fast in the standard stochastic chemical kinetics model [11] (see A.2). The largest part of this paper concerns the negative result that only detection problems can be computed by speed fault free CRNs (Section 4.2). The proof of the negative result consists of finding a worst-case reaction sequence that leads to a speed fault, assuming a non-detection problem is computed.

To get a sense of why complex machinery is necessary for the negative result, consider the following naïve approach. Suppose we aim to show that the “2A problem” cannot be computed without speed faults. We note that absent speed-faults, the correct output must be obtained without direct reactions between $O(1)$ -count species; rather the low count species must initiate cascades through intermediary high count species. Thus to obtain the worst-case reaction sequence, we can imagine isolating the low count species into “separate test tubes” and “intercepting communication” between them. We can then imagine using the symmetry between the two A molecules to make the system think that it’s communicating with just one A and fail to detect the second A . This approach fails in at least two ways. First, the distinction between low count and high count species is not static, but rather species present in low count can become high count and vice versa. Second, while a proper choice of reaction sequence can force communication symmetry for arbitrarily long, to prove that a speed fault is unavoidable we need to show that there does not exist a speed fault free continuation that breaks the symmetry. (Intuitively, recall that the example in Fig. 1(g) fails to compute correctly if both A molecules react via the same reaction. For the negative result we need to be able to show that there is no way that we can change the CRN such that it can retry indefinitely.)

At the core of our negative result is a pumping technique which formally distinguishes species that can get arbitrarily high with increasing n from species whose counts are bounded by a constant.⁷ We show that all high count species that can be encountered along a trajectory can be pumped to be *simultaneously* high. By looking at such a maximally pumped state we get a sense of not only what did happen but what *can* happen in the future — specifically all “communication” that can occur.

2 Previous work and future directions

Much related work in the distributed computing community is phrased in the language of population protocols rather than CRNs (e.g. [2]). While population protocols are equivalent to CRNs with exactly two reactants and two products, and thus a fixed population size, CRNs can naturally describe reactions that consume or produce net molecules. As a result CRNs can potentially explore an unbounded state space, and certain questions that are not natural for population protocols become germane for CRNs (for example: Turing universality). Because our negative result naturally applies to a changing population size, we phrase this paper in the language of CRNs.

CRNs have a surprisingly rich computational structure. If we allow the number of species and reactions to scale with the size of the input (i.e. we view CRNs as a non-uniform model of computation), then $\log s$ species can deterministically simulate space s -bounded Turing machines [6]. (These results are presented in a model called vector addition systems [13], but easily carry over.) Thus CRNs are a very powerful model of non-uniform computation. On the other hand, we ask what functions can be computed by a fixed CRN (i.e. fixed number of species and reactions, with input encoded in the initial molecular counts, which corresponds to a uniform model). In this setting, CRNs are not Turing universal, unless we allow for some probability of error [3, 16]. In attempting Turing universal computation, there will provably always be “race conditions” that lead to error if certain reactions occur in a (maybe unlikely but possible) malicious order. The fact that even such Turing universal computation is possible, and indeed can be made “fast” is surprising since finite CRNs necessarily must represent binary data strings in a unary encoding, since they lack positional information to tell the difference between two molecules of the same species.

captured in our notion of speed faults.

⁷Note that our pumping lemma is very different from a similarly called “pumping lemma” of ref. [2], which shows that how input can be increased without changing the output (thus pumping “input”).

Deterministic computation of both predicates and functions has been exactly characterized, and corresponds to semilinear sets and functions [2,7]. Angluin, Aspnes, and Eisenstat [2] showed that all semilinear predicates can be deterministically computed in expected $O(n \text{ polylog } n)$ “interactions” (molecules bumping into each other). In a volume of fixed size, with n molecules, there are an expected $\Theta(n^2)$ such interactions per unit time, which yields expected time $O((1/n)\text{polylog } n)$ — decreasing with n . Our results imply that when computing semilinear predicates other than the detection problems, it is always possible to reach a state (speed fault) from which the expected time to finish the computation is $\Omega(1)$ — independent of n . It is easy to reconcile the two results: in the construction of ref. [2], the probability that a speed fault is reached decreases with n , and thus the total expected time decreases with n as well. Our result implies that this is a necessary feature of any such construction, and is not simply due to insufficient cleverness of the researchers to avoid speed faults.

Other work showing the challenges in parallelizing CRNs include the investigation of running multiple copies of networks in parallel [9], and the inability of networks starting with only high count species to delay the production of any species [10].

While in this work we focused on parallelizable predicates, it remains to explore the class of parallelizable functions. For example, if the initial amount of A is the input and the final amount of B is the output, then we can think of the reaction $F + A \rightarrow 2B$ as deterministically computing $f(x) = 2x$. Clearly as the amount of F increases, the computation converges faster. On the other hand, we believe that computing division by 2 should not be possible without speed faults, although that remains to be shown.

Since the occurrence of a speed fault leads to a slow computational bottleneck, speed faults affect the tail bounds on the distribution of the computation time. Indeed, two CRNs may compute with the same fast expected time, but the one susceptible to speed faults will likely have a larger probability of taking significantly longer. It remains to rigorously draw out the connection between tail bounds and speed faults.

3 Preliminaries

3.1 Chemical reaction networks

If Λ is a finite set (in this paper, of chemical species), we write \mathbb{N}^Λ to denote the set of functions $f : \Lambda \rightarrow \mathbb{N}$. Equivalently, we view an element $\mathbf{c} \in \mathbb{N}^\Lambda$ as a vector of $|\Lambda|$ nonnegative integers, with each coordinate “labeled” by an element of Λ . Given $S \in \Lambda$ and $\mathbf{c} \in \mathbb{N}^\Lambda$, we refer to $\mathbf{c}(S)$ as the *count of S in \mathbf{c}* . Let $\|\mathbf{c}\|_\infty = \max_{S \in \Lambda} \mathbf{c}(S)$. We write $\mathbf{c} \leq \mathbf{c}'$ to denote that $\mathbf{c}(S) \leq \mathbf{c}'(S)$ for all $S \in \Lambda$, and $\mathbf{c} < \mathbf{c}'$ if $\mathbf{c} \leq \mathbf{c}'$ and $\mathbf{c} \neq \mathbf{c}'$. Since we view vectors $\mathbf{c} \in \mathbb{N}^\Lambda$ equivalently as multisets of elements from Λ , if $\mathbf{c} \leq \mathbf{c}'$ we say \mathbf{c} is a *subset* of \mathbf{c}' . Given $\mathbf{c}, \mathbf{c}' \in \mathbb{N}^\Lambda$, we define the vector component-wise operations of addition $\mathbf{c} + \mathbf{c}'$, subtraction $\mathbf{c} - \mathbf{c}'$, and scalar multiplication $n\mathbf{c}$ for $n \in \mathbb{N}$. For a set $\Delta \subset \Lambda$, we view a vector $\mathbf{c} \in \mathbb{N}^\Delta$ equivalently as a vector $\mathbf{c} \in \mathbb{N}^\Lambda$ by assuming $\mathbf{c}(S) = 0$ for all $S \in \Lambda \setminus \Delta$. Write $\mathbf{c} \upharpoonright \Delta$ to denote the vector $\mathbf{d} \in \mathbb{N}^\Delta$ such that $\mathbf{c}(S) = \mathbf{d}(S)$ for all $S \in \Delta$. Given $S_1, \dots, S_k \in \Lambda$, $\mathbf{c} \in \mathbb{N}^\Lambda$, and $n_1, \dots, n_k \in \mathbb{Z}$, we write $\mathbf{c} + \{n_1 S_1, \dots, n_k S_k\}$ to denote vector addition of \mathbf{c} with the vector $\mathbf{v} \in \mathbb{Z}^{\{S_1, \dots, S_k\}}$ with $\mathbf{v}(S_i) = n_i$.

Given a finite set of chemical species Λ , a *reaction* over Λ is a triple $\alpha = \langle \mathbf{r}, \mathbf{p}, k \rangle \in \mathbb{N}^\Lambda \times \mathbb{N}^\Lambda \times \mathbb{R}^+$, specifying the stoichiometry (amount consumed/produced) of the reactants and products, respectively, and the *rate constant* k . A reaction is *unimolecular* if it has one reactant and *bimolecular* if it has two reactants. We use no higher-order reactions in this paper. If not specified, assume that $k = 1$ (this is the case for all reactions in this paper), so that the reaction $\alpha = \langle \mathbf{r}, \mathbf{p}, 1 \rangle$ is also represented by the pair $\langle \mathbf{r}, \mathbf{p} \rangle$. For instance, given $\Lambda = \{A, B, C\}$, the reaction $A + 2B \rightarrow A + 3C$ is the pair $\langle (1, 2, 0), (1, 0, 3) \rangle$. A (*finite*) *chemical reaction network (CRN)* is a pair $N = (\Lambda, R)$, where Λ is a finite set of chemical *species*, and R is a finite set of reactions over Λ . A *state* of a CRN $N = (\Lambda, R)$ is a vector $\mathbf{c} \in \mathbb{N}^\Lambda$.

Given a state \mathbf{c} and reaction $\alpha = \langle \mathbf{r}, \mathbf{p} \rangle$, we say that α is *applicable* to \mathbf{c} if $\mathbf{r} \leq \mathbf{c}$ (i.e., \mathbf{c} contains enough of each of the reactants for the reaction to occur). If α is applicable to \mathbf{c} , then write $\alpha(\mathbf{c})$ to denote the state $\mathbf{c} + \mathbf{p} - \mathbf{r}$ (i.e., the state that results from applying reaction α to \mathbf{c}). If

$\mathbf{c}' = \alpha(\mathbf{c})$ for some reaction $\alpha \in R$, we write $\mathbf{c} \xrightarrow{1}_N \mathbf{c}'$, or $\mathbf{c} \rightarrow^1 \mathbf{c}'$ when N is clear from context. An *execution sequence* \mathcal{E} is a finite or infinite sequence of states $\mathcal{E} = (\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \dots)$ such that, for all $i \in \{1, \dots, |\mathcal{E}| - 1\}$, $\mathbf{c}_{i-1} \rightarrow^1 \mathbf{c}_i$. Call the reactions that cause each $\mathbf{c}_{i-1} \rightarrow^1 \mathbf{c}_i$ a *reaction sequence*.⁸ If a finite execution sequence starts with \mathbf{c} and ends with \mathbf{c}' , we write $\mathbf{c} \xrightarrow{N} \mathbf{c}'$, or $\mathbf{c} \rightarrow \mathbf{c}'$ when N is clear from context, and we say that \mathbf{c}' is *reachable* from \mathbf{c} . We write $\mathbf{c} \rightarrow_r \mathbf{c}'$ to denote that the sequence of reactions r , applied to \mathbf{c} , results in \mathbf{c}' .

3.2 Stable decidability of predicates

We now review the definition of stable decidability of predicates introduced by Angluin, Aspnes, and Eisenstat [2]. Intuitively, some species “vote” for a yes/no answer, and a CRN N is a stable decider if N is guaranteed to reach a consensus vote.

A *chemical reaction decider* (CRD) is a tuple $\mathcal{D} = (\Lambda, R, \Sigma, \Upsilon, \phi, \mathbf{s})$, where (Λ, R) is a CRN, $\Sigma \subseteq \Lambda$ is the *set of input species*, $\Upsilon \subseteq \Lambda$ is the set of *voters*, $\phi : \Upsilon \rightarrow \{0, 1\}$ is the (*Boolean*) *output function*, and $\mathbf{s} \in \mathbb{N}^{\Lambda \setminus \Sigma}$ is the *initial context*. An input to \mathcal{D} will be a vector $\mathbf{i}_\Sigma \in \mathbb{N}^\Sigma$. Thus \mathcal{D} together with \mathbf{i}_Σ defines a *valid* initial state $\mathbf{i} \in \mathbb{N}^\Lambda$ defined by $\mathbf{i} \upharpoonright \Sigma = \mathbf{i}_\Sigma$ and $\mathbf{i} \upharpoonright (\Lambda \setminus \Sigma) = \mathbf{s}$.

We extend ϕ to a partial function $\Psi : \mathbb{N}^\Lambda \dashrightarrow \{0, 1\}$ as follows. $\Psi(\mathbf{c})$ is undefined if either $\mathbf{c}(X) = 0$ for all $X \in \Upsilon$, or if there exist $X_0, X_1 \in \Upsilon$ such that $\mathbf{c}(X_0) > 0$, $\mathbf{c}(X_1) > 0$, $\phi(X_0) = 0$ and $\phi(X_1) = 1$. Otherwise, there exists $b \in \{0, 1\}$ such that $(\forall X \in \Upsilon)(\mathbf{c}(X) > 0 \implies \phi(X) = b)$; in this case, the *output* $\Psi(\mathbf{c})$ of state \mathbf{c} is b .

A state \mathbf{o} is *output stable* if $\Psi(\mathbf{o})$ is defined and, for all \mathbf{c} such that $\mathbf{o} \rightarrow \mathbf{c}$, $\Psi(\mathbf{c}) = \Psi(\mathbf{o})$. We say that a CRD \mathcal{D} is *stable* if, for any valid initial state $\mathbf{i} \in \mathbb{N}^\Lambda$, there exists $b \in \{0, 1\}$ such that, for every state \mathbf{x} reachable from \mathbf{i} , there is an output stable state \mathbf{o} reachable from \mathbf{x} such that $\Psi(\mathbf{o}) = b$ (i.e., \mathcal{D} always converges to a defined output b on input \mathbf{i} , where b depends only on \mathbf{i} and not on the path taken). If \mathcal{D} is stable, then some unique subset $S_0 \subseteq \mathbb{N}^\Sigma$ of all possible initial states always converges to output 0 and stays with that output, and the remainder $S_1 = \mathbb{N}^\Sigma \setminus S_0$ always converges to output 1 and stays with that output. We say that \mathcal{D} *stably decides* the set S_1 , or that \mathcal{D} *stably decides* the predicate $\psi : \mathbb{N}^\Sigma \rightarrow \{0, 1\}$ defined by $\psi(\mathbf{i} \upharpoonright \Sigma) = 1$ if $\mathbf{i} \upharpoonright \Sigma \in S_1$ and $\psi(\mathbf{i} \upharpoonright \Sigma) = 0$ if $\mathbf{i} \upharpoonright \Sigma \in S_0$.

A set $A \subseteq \mathbb{N}^k$ is *linear* if $A = \{ \mathbf{b} + \sum_{i=1}^p n_i \mathbf{u}_i \mid n_1, \dots, n_p \in \mathbb{N} \}$ for some constant vectors $\mathbf{b}, \mathbf{u}_1, \dots, \mathbf{u}_p \in \mathbb{N}^k$. A is *semilinear* if it is a finite union of linear sets.

The following theorem is due to Angluin, Aspnes, and Eisenstat [2]:

Theorem 3.1 ([2]). *A set $A \subseteq \mathbb{N}^k$ is stably decidable by a CRD if and only if it is semilinear.*

The following lemma is useful for assuming properties of infinite sequences of states.

Lemma 3.2 (Higman’s Lemma [12]). *The set of states \mathbb{N}^k is well-quasi-ordered. In particular, every infinite sequence $\mathbf{x}_0, \mathbf{x}_1, \dots$ of states has an infinite nondecreasing subsequence $\mathbf{x}_{i_0} \leq \mathbf{x}_{i_1} \leq \dots$, where $i_0 < i_1 < \dots \in \mathbb{N}$, and every set $U \subseteq \mathbb{N}^k$ has a finite number of minimal elements.*

If a YES voter (or any other species, for that matter) cannot be produced by any sequence of reactions from a state \mathbf{y} , then it cannot be produced from any subset $\mathbf{y}' \leq \mathbf{y}$. The following lemma — proven in the appendix — is useful when we want to argue the other way: that for certain species, beyond a certain value, *increasing* their counts cannot affect the ability or inability of the state to produce a YES voter. Say that a state \mathbf{c} is *committed* if, for all states \mathbf{z} such that $\mathbf{c} \rightarrow \mathbf{z}$, $\mathbf{z}(S) = 0$ for all YES-voting species S . In particular, all output-stable NO states are committed.

Lemma 3.3. *For each CRD, there is a constant c such that, for all committed states \mathbf{c} , if $\mathbf{c}(S) > c$ for some $S \in \Lambda$, then for all $n \in \mathbb{Z}$, $\mathbf{c} + \{nS\}$ is also committed.*

⁸When an initial state to which a reaction sequence is applied is clear from context, we may overload terminology and refer to a reaction sequence as if it is an execution sequence, since the states traversed may be derived from the initial state and the reaction sequence.

4 Speed fault free CRDs

In this section we show our main result that speed fault free CRDs decide only “detection problems,” i.e., detecting the presence or absence of a species, but not distinguishing between two different positive counts of it. Below we formalize the concept of “speed fault free” discussed informally in the introduction. Briefly, a CRN experiences a speed fault if it reaches a state from which the only paths of reactions that reach a correct state are those with a bimolecular reaction between two constant-count species (a “slow” reaction). To allow for “parallelization” of the computation, we introduce a “fuel” species F , whose count is allowed to start arbitrarily high.⁹ Increasing the amount of fuel species is analogous to increasing the amount of “receptor” in the introduction.

Let $\mathcal{D} = (\Lambda, R, \Sigma, \Upsilon, \phi, \mathbf{s})$ be a stable CRD, where $\Sigma = \{A_1, \dots, A_k\}$ and $\Lambda \setminus \Sigma$ contains a special “fuel” species F , with variable initial count n (and the initial count of every other species in $\Lambda \setminus (\Sigma \cup \{F\})$ is constant with respect to n). Write the initial state of \mathcal{D} with some number n_i of each A_i and n molecules of F as $\mathbf{i}_n(n_1, \dots, n_k)$.

Let $f \in \mathbb{N}$, let $\alpha \in R$ be a reaction and $\mathbf{x} \in \mathbb{N}^\Lambda$ be a state. We say that α occurring in state \mathbf{x} is f -fast if either α is unimolecular, or if it is bimolecular, then at least one reactant has count at least f in \mathbf{x} . An execution sequence is called f -fast if all reactions in it are f -fast. If $\mathbf{x} \rightarrow \mathbf{y}$ via a f -fast execution sequence, then we write $\mathbf{x} \xrightarrow{f} \mathbf{y}$.¹⁰

Definition 4.1. A stable CRD \mathcal{D} is *speed fault free* if for all n_1, \dots, n_k and all $f \in \mathbb{N}$, for all sufficiently large n ,¹¹ for any state \mathbf{x} such that $\mathbf{i}_n(n_1, \dots, n_k) \rightarrow \mathbf{x}$, there is an output stable state \mathbf{y} (which has the correct answer with respect to n_1, \dots, n_k by the stability of \mathcal{D}) such that $\mathbf{x} \xrightarrow{f} \mathbf{y}$.

In other words, from any reachable state (whether reachable by a fast or slow sequence of reactions), there is always a sequence of fast reactions that reaches the correct answer. Conversely, if a CRN is not speed fault free, then there is some constant f such that, for infinitely many n , getting the correct answer from a reachable state \mathbf{x} requires the slowest sort of reaction there is: a bimolecular reaction between two molecules of count at most the constant f .

Definition 4.2. A set $S \subseteq \mathbb{N}^k$ is a *simple detection set* if there is a $1 \leq i \leq k$ such that $S = \{ (x_1, \dots, x_k) \in \mathbb{N}^k \mid x_i > 0 \}$. A set is a *detection set* if it is expressible as a combination of finite unions, intersections, and complements of simple detection sets.

In other words, the predicate corresponding to S is a finite Boolean combination of questions of the form “is a certain species present?”. The following theorem is the main result of this paper.

Theorem 4.3. *The sets decidable by speed fault free CRDs are precisely the detection sets.*

We show each direction of Theorem 4.3 in two separate lemmas, Lemma 4.4 and Lemma 4.11.

4.1 Detection problems are decidable by speed fault free CRDs

This is the easier direction of Theorem 4.3, so we give only the intuition behind the proof here, and we do not formally define the model of stochastic chemical kinetics used to prove the expected running time. See the appendix for detailed definitions and the proof.

Lemma 4.4. *Every detection set is decidable by a speed fault free CRD. This CRD takes expected time $O(\log n)$ to stabilize under the standard model of stochastic chemical kinetics.*

⁹Allowing multiple fuel species F_1, F_2, \dots does affect our results since one of our reactions can be $F \rightarrow F_1 + F_2 \dots$

¹⁰It is worth noting that fast reaction sequences are not necessarily fast in the standard sense of stochastic kinetics, since although each reaction occurs quickly, it could be that there are a huge number of reactions in the sequence. Since our main result is a lower bound, this does not hurt the argument (and our upper bound result also shows that it is possible to decide detection problems quickly under the standard stochastic model).

¹¹ n may need to be larger for different input values n_1, \dots, n_k and different constants f .

Proof sketch. To detect whether a species A is present or not, we may use “epidemic” reactions $A + F \rightarrow F_a$ and $F_a + F \rightarrow 2F_a$, where F votes NO and F_a votes YES. That is, if A encounters an F , then F changes state to F_a , and this information is “broadcast” throughout the population of F ’s. Since the sum $\mathbf{c}(F_a) + \mathbf{c}(F) = n$ is constant in any reachable state \mathbf{c} , the second bimolecular reaction always has a reactant with count $\geq n/2$ (hence that reaction is always $\frac{n}{2}$ -fast), and the output-stable YES state is reached when all F ’s are converted to F_a . The extension to k input species just means that each F must store k bits, one for each input species. \square

4.2 Speed fault free CRDs decide only detection problems

Throughout this section, let $\mathcal{D} = (\Lambda, R, \Sigma, \Upsilon, \phi, \mathbf{s})$ be an arbitrary speed fault free CRD with $\Sigma = \{A_1, \dots, A_k\}$ and fuel species F as in Definition 4.1.

4.2.1 Pumpable sets of species

This section defines *pumpable* sets of species (species whose counts can be made arbitrarily high) and proves some basic properties about them. For example, the fuel species F is trivially pumpable. If there is a reaction $F + A \rightarrow F' + A$, then F' is pumpable (if there is an A), because F can be arbitrarily high. To get a handle on the notion of speed fault free, we define pumping to enforce a certain kind of self-consistency (Π -friendly): you can pump without requiring any bimolecular reactions where both reactants are not pumpable.

Let $\Pi \subseteq \Lambda$. If a reaction is either unimolecular, or if it is bimolecular it has at least one reactant in Π , say the reaction is Π -friendly. If $\mathbf{x} \rightarrow \mathbf{y}$ via a reaction sequence in which all reactions are Π -friendly, then we write $\mathbf{x} \rightarrow^\Pi \mathbf{y}$. Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \mathbf{z}_3 \dots)$, where each $\mathbf{z}_n \in \mathbb{N}^\Lambda$, be an infinite nondecreasing sequence of states. A set of species $\Pi \subseteq \Lambda$ is *Z-pumpable* if there exists a sequence of states $X = (\mathbf{x}_1, \mathbf{x}_2, \dots)$ such that: (1) for all $S \in \Pi$, $\mathbf{x}_m(S) \geq m$, and (2) for all $m \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that $\mathbf{z}_n \rightarrow^\Pi \mathbf{x}_m$. Call such a sequence (\mathbf{x}_m) a *pumping sequence* for Z . Π is *maximal Z-pumpable* if it is Z -pumpable and no strict superset of Π is Z -pumpable.

Because (\mathbf{z}_n) is a nondecreasing sequence, if $\mathbf{z}_n \rightarrow^\Pi \mathbf{x}_m$ for some $n, m \in \mathbb{N}$, then for all $n' > n$, there is a superset $\mathbf{x}'_m \geq \mathbf{x}_m$ such that $\mathbf{z}_{n'} \rightarrow^\Pi \mathbf{x}'_m$, and $\mathbf{x}'_m(S) \geq m$ for all $S \in \Pi$. Therefore we freely assume that the n in the definition of pumpable grows without bound as $m \rightarrow \infty$. This is a technical condition that will be useful later.

The next proposition states that along any execution path that makes species in a maximal pumpable set high, all other species have their counts bounded by a constant. It is proven in the appendix. Intuitively, it holds because if any other species $S \notin \Pi$ could get high via some reaction sequence r , then we could make the species in Π so high that we are able to hold some in reserve, then execute r , and then we would have S and all of Π high at the same time, contradicting the maximality of Π .

Proposition 4.5. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable, with pumping sequence (\mathbf{x}_m) . Then there is a constant c such that, for all states \mathbf{y} and $m, n \in \mathbb{N}$ such that either 1) $\mathbf{z}_n \rightarrow^\Pi \mathbf{y} \rightarrow^\Pi \mathbf{x}_m$ or 2) $\mathbf{x}_m \rightarrow^\Pi \mathbf{y}$, then for all $S \in \Lambda \setminus \Pi$, $\mathbf{y}(S) < c$.*

We will use Proposition 4.5 repeatedly throughout this section, but its most important consequence is the following corollary (proven formally in the appendix), which says essentially that the only way to get something outside of Π “high” is by executing a “slow” reaction.

Corollary 4.6. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable, with pumping sequence (\mathbf{x}_m) . Let $c \in \mathbb{N}$ be the constant in Proposition 4.5 ensuring that in any state \mathbf{c} such that $\mathbf{x}_m \rightarrow^\Pi \mathbf{c}$, $\mathbf{c}(S) < c$ for all $S \in \Lambda \setminus \Pi$. Then if $\mathbf{x}_m \rightarrow_p \mathbf{c}'$ and $\mathbf{c}'(S) \geq c$, where $S \in \Lambda \setminus \Pi$, some reaction in p is not c -fast.*

The next proposition is proven in the appendix.

Proposition 4.7. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable. For any species $X, Y \in \Pi$ such that R contains a reaction $X + Y \rightarrow \dots$, all products of the reaction are contained in Π .*

4.2.2 Main proof

Supposing for the sake of contradiction that \mathcal{D} decides some non-detection set, then there must exist some species A_i (assume without loss of generality that $i = 1$), and an input value $(n_1, n_2, \dots, n_k) \in \mathbb{N}^k$, where $n_1 \geq 1$, with answer NO (without loss of generality) but input value $(n_1 + 1, n_2, \dots, n_k)$ has answer YES.

For each $n \in \mathbb{N}$, write the initial state of \mathcal{D} with answer NO and n fuels as $\mathbf{i}_n (= \{n_1 A_1, n_2 A_2, \dots, n_k A_k, nF, \mathbf{s} \upharpoonright (\Lambda \setminus (\Sigma \cup \{F\}))\})$. Our argument will focus on the initial state \mathbf{i}_n with answer NO, for sufficiently large n , and \mathbf{i}_n with one extra copy of A_1 , which has answer YES. We will show that for sufficiently large n , $\mathbf{i}_n + \{A_1\}$ is able to reach a state with no YES-voting species, from which the only way to produce a YES voter is to execute a slow bimolecular reaction.

Let $I = (\mathbf{i}_1, \mathbf{i}_2, \dots)$ be the infinite increasing sequence of all such initial states with answer NO.

We first observe that for a maximal I -pumpable set Π there is a very controlled manner in which the species in Π can be pumped.

Lemma 4.8. *Let Π be maximal I -pumpable. There exists $\mathbf{d} \in \mathbb{N}^\Lambda$, with $\mathbf{d}(S) > 0 \iff S \in \Pi$, and an infinite sequence of states $\mathbf{x}_1, \mathbf{x}_2, \dots$ such that, for all $m \in \mathbb{Z}^+$, $\mathbf{x}_m + \mathbf{d} = \mathbf{x}_{m+1}$, and there exists $n \in \mathbb{N}$ such that $\mathbf{i}_n \xrightarrow{\Pi} \mathbf{x}_m$.*

A proof is given in the appendix. Intuitively, it follows from Leroux’s recent theorem [14] showing that the set of all reachable states of a CRN, while not semilinear, is “close” to being closed under addition (it is a finite union of translations of sets closed under addition); hence if we can find a state with counts in Π high, then we can add (something close to) that state to itself arbitrarily many times to get higher counts, and that state is reachable as well.

Define the sequence of output-stable NO states (\mathbf{y}_m) inductively as follows. For the base case, let \mathbf{y}_1 be any output-stable NO state such that $\mathbf{x}_1 \xrightarrow{r_1} \mathbf{y}_1$; such a path r_1 must exist because \mathcal{D} is stable. Inductively assume that $\mathbf{x}_{m-1} \xrightarrow{r_{m-1}} \mathbf{y}_{m-1}$. Then $\mathbf{x}_m = \mathbf{x}_{m-1} + \mathbf{d} \xrightarrow{r_{m-1}} \mathbf{y}_{m-1} + \mathbf{d}$. Let $f_m \in \mathbb{N}$ be the largest number such that there is a f_m -fast path p_m from $\mathbf{y}_{m-1} + \mathbf{d}$ to an output-stable NO state \mathbf{y}_m , i.e., $\mathbf{y}_{m-1} + \mathbf{d} \xrightarrow{f_m}_{p_m} \mathbf{y}_m$.¹² Thus the sequence of reactions r_m is r_{m-1} followed by p_m . In other words $r_m = \bigoplus_{i=1}^m p_i$, where \bigoplus denotes concatenation. Note that by the definition of speed fault free, $\lim_{m \rightarrow \infty} f_m = \infty$, because f -fast paths to output-stable states exist from any state reachable from any sufficiently large initial state.

By Corollary 4.6, once m is sufficiently large, any f_m -fast reaction sequence from \mathbf{x}_m to \mathbf{y}_m must be Π -friendly. We therefore assume that we in fact start counting from this value of m , i.e., \mathbf{x}_1 is defined to be a state with counts of species in Π at least m large enough that f_m is larger than the constant c in Proposition 4.5. Therefore, we have that for all m , $\mathbf{x}_m \xrightarrow{\Pi} \mathbf{y}_m$.

By Higman’s Lemma there is an infinite nondecreasing subsequence $Y = (\mathbf{y}_{s_1}, \mathbf{y}_{s_2}, \dots)$, where $\mathbf{y}_{s_n} \leq \mathbf{y}_{s_{n+1}}$ for all n . Let $\Gamma = \left\{ S \in \Lambda \mid \lim_{n \rightarrow \infty} \mathbf{y}_{s_n}(S) = \infty \right\}$. By Proposition 4.5, $\Gamma \subseteq \Pi$ since $\mathbf{x}_{s_n} \xrightarrow{\Pi} \mathbf{y}_{s_n}$. Let $\Delta = \Pi \setminus \Gamma$. These are the species that are “high” in (\mathbf{x}_m) but are bounded in Y .¹³ By the definition of Γ , there is a constant c (as in Proposition 4.5) such that, for all $n \in \mathbb{N}$ and $S \in \Delta$, $\mathbf{y}_{s_n}(S) < c$. Thus an infinite subsequence of \mathbf{y}_{s_n} ’s has equal counts of all species $S \in \Lambda \setminus \Gamma$. Since counts of species in Γ grow without bound as $n \rightarrow \infty$, there is an infinite subsequence of *that* on which *every* element of Γ has its count strictly increase on each subsequent \mathbf{y}_{s_n} state. We therefore assume that $Y = (\mathbf{y}_{s_1}, \mathbf{y}_{s_2}, \dots)$ is exactly this infinite subsequence (to avoid introducing many new variable names), where each $\mathbf{y}_{s_n}(S) = \mathbf{y}_{s_{n+1}}(S)$ if $S \in \Lambda \setminus \Gamma$ and $\mathbf{y}_{s_n}(S) < \mathbf{y}_{s_{n+1}}(S)$ if $S \in \Gamma$.

Recall that a state is *committed* if it cannot produce a YES voter; hence all output stable NO states are committed. The next lemma shows that changing counts of pumpable species by a “small” amount in \mathbf{x}_m , so long as m is sufficiently large, cannot change the ability of \mathbf{x}_m to reach a committed state (not necessarily output-stable NO, as there may be an absence of NO voters in the new reached state). Intuitively, \mathbf{e} will represent a change in counts due to “processing” the extra copy of A_1 (the

¹²Let $f_m = f_{m-1} + 1$ if there are f -fast paths from $\mathbf{y}_{m-1} + \mathbf{d}$ to \mathbf{y}_m for all $f \in \mathbb{N}$; this may occur if there are reactions such as $X \rightarrow X + F$ that can generate arbitrarily large counts of some molecules without consuming others.

¹³Because species in Δ are bounded, Proposition 4.5 does not apply to them, so their addition to output-stable NO states may cause the production of YES voters.

one that changes the correct answer in state $\mathbf{i}_n(n_1, \dots, n_k)$ from NO to YES), and the following lemma will help us to derive a contradiction because the extra copy of A_1 should enable the production of a YES voter.

Lemma 4.9. *Let sequences (\mathbf{x}_m) and (\mathbf{y}_m) be as defined above. For all $\epsilon \in \mathbb{N}$, there exists $\epsilon' \in \mathbb{N}$ such that the following holds. For all $\mathbf{e} \in \mathbb{Z}^\Pi$ with $|\mathbf{e}| \leq \epsilon$, for infinitely many m , there exists $\mathbf{e}_m \in \mathbb{Z}^\Gamma$ with $|\mathbf{e}_m| \leq \epsilon'$, and $m_2 < m$ such that $\mathbf{x}_m + \mathbf{e} \rightarrow^\Pi \mathbf{y}_{m_2} + \mathbf{e}_m$ and $\mathbf{y}_{m_2} + \mathbf{e}_m$ is committed.*

In other words, if we perturb \mathbf{x}_m by a small \mathbf{e} , then it retains the ability to reach a state that is committed (if not output-stable NO), and that state will be “close” to one of the output-stable NO states \mathbf{y}_{m_2} in the sequence Y . Although the exact distance \mathbf{e}_m from \mathbf{y}_{m_2} depends on m , we can place a bound ϵ' on that distance that depends only on ϵ , the size of the perturbation of \mathbf{x}_m . Lemma 4.9 is formally proved in the appendix. Here we outline the main ideas of the proof.

Proof sketch. We know that $\mathbf{x}_m \rightarrow_{r_m} \mathbf{y}_m$. Consider applying r_m to $\mathbf{x}_m + \mathbf{e}$ to get $\mathbf{y}_m + \mathbf{e}$. This may not work because it could drive some species negative, and the final state may not be committed. The key idea is that we can write an ordering $\Delta = \{N_1, \dots, N_l\}$ such that we can add or remove from r_m reactions of the form $\alpha_i : R + N_i \rightarrow P_1 + \dots + P_k$ where R, P_1, \dots, P_k are in $\Gamma \cup \{N_{i+1}, \dots, N_l\}$. This gives a way to “fix” the count of N_i to make its count equal to its count in $\mathbf{y}_{m_2} + k\mathbf{d}$ by either removing α_i (to increase) or adding it (to decrease), while affecting only species in Π or “after” N_i (hence their counts will be fixed later). The counts of N_1, \dots, N_{i-1} , which have already been fixed, are unaffected by the surgery to fix N_i , because they do not appear in α_i . When we are done, we have increased the “error” in species in Γ to \mathbf{e}' , but by Lemma 3.3 they can have any count in \mathbf{y}_m and it is still committed.

How do we know such an ordering exists on Δ ? It helps to think in reverse. Consider the potential function $\Phi(\mathbf{c}) = \sum_{N \in \Delta} \mathbf{c}(N)$; then $\Phi(\mathbf{x}_m)$ is large (at least $|\Delta|m$) and $\Phi(\mathbf{y}_m)$ is constant, say c . On the path from \mathbf{x}_m to \mathbf{y}_m , when Φ is “small” (say, \hat{c}) but still bigger than c it cannot get smaller by reactions of the form $N_i + N_j \rightarrow \dots$, since $N_i, N_j \in \Delta$, or that reaction would be slow. Therefore to get Φ from \hat{c} down to c requires reactions with at most one reactant in Δ . Furthermore, if any product were in Δ , this would not decrease the value of Φ , hence some reaction must be of the desired form: consuming exactly one element of Δ and producing only elements of Π . This element is N_l , the last in the ordering. Inductively defining an ordering on $\Delta \setminus \{N_l\}$ gives the entire ordering. \square

The next lemma uses Lemma 4.9 to show that, from state $\mathbf{x}_m + \mathbf{e}$, with \mathbf{e} “small,” we can reach a committed state in which every species that can be “high”, is actually high.

Lemma 4.10. *Let sequence (\mathbf{x}_m) be as defined above. For all $\epsilon \in \mathbb{N}$, there exists $c \in \mathbb{N}$ and $\Omega \subseteq \Lambda$ such that the following holds. For all $\mathbf{e} \in \mathbb{Z}^\Pi$ such that $|\mathbf{e}| \leq \epsilon$, there exists an infinite sequence $W_{\mathbf{e}} = (\mathbf{w}_n)$ of states such that, for all $n \in \mathbb{N}$, there exists $m_n \in \mathbb{N}$, such that the following is true: (1) $\mathbf{x}_{m_n} + \mathbf{e} \rightarrow^\Pi \mathbf{w}_n$, (2) \mathbf{w}_n is committed, (3) for all $S \in \Omega$, $\mathbf{w}_n(S) \geq n$, and (4) for all $S \in \Lambda \setminus \Omega$ and all \mathbf{u} such that $\mathbf{w}_n \rightarrow^\Pi \mathbf{u}$, $\mathbf{u}(S) \leq c$.*

Proof. Choose ϵ' for ϵ as in Lemma 4.9. From Lemma 4.9, recall the infinite sequence $(\mathbf{y}_{m_2} + \mathbf{e}_m)$, where m ranges over an infinite number of values, and for each m , $m_2 < m$ is a function of m . Define the constant $\mathbf{h} \in \mathbb{N}^\Gamma$ by $\mathbf{h}(S) = \epsilon'$ for all $S \in \Gamma$. From the infinite sequence of m 's, define the infinite sequence $(\mathbf{y}_{m_2} - \mathbf{h})$, letting the k 'th element of this sequence be denoted \mathbf{z}_k , and let $Z = (\mathbf{z}_k)$ denote this sequence. Note that $\mathbf{y}_{m_2} - \mathbf{h} \leq \mathbf{y}_{m_2} + \mathbf{e}_m$, since ϵ' bounds every component of \mathbf{e}_m .

Let Ω be a maximal Z -pumpable set. By the fact that each \mathbf{z}_k is \mathbf{y}_{m_2} minus a constant \mathbf{h} for arbitrarily large m_2 , for all $S \in \Gamma$, $\lim_{k \rightarrow \infty} \mathbf{z}_k(S) = \infty$. Therefore by Proposition 4.5, $\Gamma \subseteq \Omega \subseteq \Pi$. Intuitively, Γ contains only species that happened to grow without bound in the states \mathbf{y}_m , whereas Ω contains all those species that, starting from (a state close to) \mathbf{y}_m , could grow without bound. Let (\mathbf{w}'_n) be a pumping sequence for Ω , i.e., for all $n \in \mathbb{N}$, there exists $k \in \mathbb{N}$ such that $\mathbf{z}_k \rightarrow^\Omega \mathbf{w}'_n$ and $\mathbf{w}'_n(S) \geq n$ for all $S \in \Omega$. Finally, for all n , define $\mathbf{w}_n = \mathbf{w}'_n + \mathbf{h} - \mathbf{e}_{m_n}$, where m_n is the value of m corresponding to k (recall to each \mathbf{z}_k there corresponds an m). Note that because $\Omega \subseteq \Pi$, any Ω -friendly reaction is also Π -friendly, hence $\mathbf{z}_k \rightarrow^\Omega \mathbf{w}'_n$ implies that $\mathbf{z}_k \rightarrow^\Pi \mathbf{w}'_n$.

We now show that (\mathbf{w}_n) has the properties claimed in the statement of the lemma. As noted above, there is a reaction sequence — call it p — that takes $\mathbf{z}_k = \mathbf{y}_{m_2} - \mathbf{h}$ to \mathbf{w}'_n , i.e., $\mathbf{y}_{m_2} - \mathbf{h} \rightarrow_p^\Pi \mathbf{w}'_n$.

Since $\mathbf{x}_{m_n} + \mathbf{e} \rightarrow \mathbf{y}_{m_2} + \mathbf{e}_{m_n} \geq \mathbf{y}_{m_2} - \mathbf{h}$, p is applicable to $\mathbf{y}_{m_2} + \mathbf{e}_{m_n}$. Therefore $\mathbf{x}_{m_n} + \mathbf{e} \rightarrow^\Pi \mathbf{y}_{m_2} + \mathbf{e}_{m_n} \rightarrow_p^\Pi \mathbf{w}'_n + \mathbf{h} - \mathbf{e}_{m_n} = \mathbf{w}_n$, so (1) holds. Since all states reachable from a committed state are committed, and $\mathbf{y}_{m_2} + \mathbf{e}'$ is committed, it follows that \mathbf{w}_n is committed, so (2) holds. Since (\mathbf{w}'_n) is a pumping sequence for Ω , $\mathbf{w}'_n(S) \geq n$ for all $S \in \Omega$, so (3) follows from the fact that $\mathbf{w}_n \geq \mathbf{w}'_n$. Finally, since Ω is maximal, and $\mathbf{h} - \mathbf{e}_{m_n}$ is over $\Gamma \subseteq \Omega$ (and thus can't make anything not in Ω higher), (4) follows by Proposition 4.5. \square

For a reaction sequence q applied to a state \mathbf{x} to give $\mathbf{x} \rightarrow_q \mathbf{y}$, where \mathbf{x} is written as a sum of two states $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}$, we say that q has a *parallel decomposition from* $(\mathbf{x}_1, \mathbf{x}_2)$ if there exists a partition of q into two disjoint subsequences of reactions (q_1, q_2) such that $\mathbf{x}_1 \rightarrow_{q_1} \mathbf{y}_1$, $\mathbf{x}_2 \rightarrow_{q_2} \mathbf{y}_2$, and $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$. In other words, if we imagine splitting \mathbf{x} into two “tubes” \mathbf{x}_1 and \mathbf{x}_2 , then the evolution determined by the reaction sequence q can be interpreted as entirely within the tubes.

Suppose a reaction sequence p is applicable to $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, but p does not have a parallel decomposition from $(\mathbf{x}_1, \mathbf{x}_2)$. Then there is a longest prefix q of p (possibly q is empty) such that q has a parallel decomposition (q_1, q_2) from $(\mathbf{x}_1, \mathbf{x}_2)$, then we call q the *join* of p from $(\mathbf{x}_1, \mathbf{x}_2)$. Write $(\mathbf{x}'_1, \mathbf{x}'_2)$ to be such that $\mathbf{x}_1 \rightarrow_{q_1} \mathbf{x}'_1$ and $\mathbf{x}_2 \rightarrow_{q_2} \mathbf{x}'_2$. In other words, q is the furthest that \mathbf{x}_1 and \mathbf{x}_2 can evolve on their own before the next reaction in p requires a molecule from \mathbf{x}'_1 and a molecule from the other \mathbf{x}'_2 . Therefore the next reaction must be bimolecular $X_1 + X_2 \rightarrow \dots$, and it must be the case that $\mathbf{x}'_1(X_2) = 0$ and $\mathbf{x}'_2(X_1) = 0$, otherwise one of the reaction sequences q_1 or q_2 could be extended by that reaction while remaining a parallel decomposition, and q would not be the longest prefix of p with a parallel decomposition.

The next lemma shows that speed fault free CRDs decide only detection problems. Lemma 4.11 is formally proved in the appendix. Here we outline the main ideas of the proof.

Lemma 4.11. *\mathcal{D} is not speed fault free.*

Proof sketch. \mathbf{x}_m is a state reachable from $\mathbf{i}_{n'}(n_1, \dots, n_k)$, i.e., a state with answer NO. Consider also tracking the evolution of one copy of A_1 (recall $n_1 \geq 1$), and keeping track of the “non-pumpable descendants” of these reactions. Since non-pumpable species have bounded counts (Proposition 4.5), reactions consuming them must be either unimolecular, or the other reactant must be high, hence in Π . Therefore the notion of a “tree” of nonpumpable descendants of A_1 is well-defined: any reaction changing them has precisely one nonpumpable reactant (and some number of nonpumpable products). Let \mathbf{b} be the vector representing these descendants in \mathbf{w}_n .

Consider adding an additional copy of A_1 to the initial state, so the answer should be YES. Withhold it from reacting until the state $\mathbf{x}_m + \{A_1\}$ has been reached. Then since counts of non-pumpable species are bounded, a constant number of reactions are necessary to alter this A_1 so that its nonpumpable base descendants become \mathbf{b} . Furthermore, to make this happen requires consuming and producing pumpable species, but the number is a constant depending only on the CRN; it is independent of m . This creates the state $\mathbf{x}_m + \mathbf{e} + \mathbf{b}$, where $\mathbf{e} \in \mathbb{Z}^\Pi$. Lemma 4.10 shows that (now again withholding \mathbf{b} from reacting) from $\mathbf{x}_m + \mathbf{e} + \mathbf{b}$ we can reach a state $\mathbf{w}_n + \mathbf{b}$ in which all species that can be “high”, are high, and \mathbf{w}_n is committed; i.e., cannot produce a YES voter. However, $\mathbf{w}_n + \mathbf{b}$ is not committed; we started from a state in which the answer is YES.

Split the state $\mathbf{w}_n + \mathbf{b}$ into two states \mathbf{l}_n and \mathbf{r}_n (a “left” and “right” tube), in which \mathbf{r}_n is simply \mathbf{b} plus $n/2$ copies of every species in Ω . This is a subset of a committed state, hence is committed. \mathbf{l}_n is a subset of \mathbf{w}_n , hence is also committed. However, there is some path p from $\mathbf{l}_n + \mathbf{r}_n$ that produces a YES voter. Our goal is to show this necessitates a slow reaction. Consider the longest prefix q of p that represents a “parallel decomposition”: i.e. q can be partitioned into disjoint subsequences q_1 and q_2 , each applicable to \mathbf{l}_n and \mathbf{r}_n , respectively. Fix p and consider increasing n to n' and apply p to $\mathbf{l}_{n'} + \mathbf{r}_{n'}$. This may increase the longest prefix q of p that has a parallel decomposition, but q cannot be all of p or else we could produce a YES voter in one of the tubes, contradicting the fact that they are each committed.

The reaction after q must be bimolecular $L + R \rightarrow \dots$, where L is absent in the right tube, and R is absent in the left tube. However, since the counts of elements of Ω start at least n' and we used only $n < n'$ along path p , for sufficiently large n' there will be some left over in each tube. Therefore neither L nor R is in Ω , so by Proposition 4.5 their counts are constant and this reaction is slow. \square

References

- [1] Dana Angluin, James Aspnes, Zoë Diamadi, Michael Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18:235–253, 2006. Preliminary version appeared in PODC 2004.
- [2] 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.
- [3] Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(3):183–199, September 2008. Preliminary version appeared in DISC 2006.
- [4] Luca Cardelli. Strand algebras for DNA computing. *Natural Computing*, 10(1):407–428, 2011.
- [5] Luca Cardelli and Attila Csikász-Nagy. The cell cycle switch computes approximate majority. *Scientific Reports*, 2, 2012.
- [6] E. Cardoza, Richard J. Lipton, and Albert R. Meyer. Exponential space complete problems for Petri nets and commutative semigroups (preliminary report). In *STOC 1976: Proceedings of the 8th annual ACM Symposium on Theory of Computing*, pages 50–54. ACM, 1976.
- [7] Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. *Natural Computing*, 2013. to appear. Preliminary version appeared in DNA 2012.
- [8] Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10):755–762, 2013.
- [9] Anne Condon, Alan Hu, Ján Maňuch, and Chris Thachuk. Less haste, less waste: On recycling and its limits in strand displacement systems. *Journal of the Royal Society Interface*, 2:512–521, 2012. Preliminary version appeared in DNA 2011.
- [10] David Doty. Timing in chemical reaction networks. In *SODA 2014: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, January 2014. to appear.
- [11] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [12] Graham Higman. Ordering by divisibility in abstract algebras. *Proceedings of the London Mathematical Society*, 3, 1952.
- [13] Richard M Karp and Raymond E Miller. Parallel program schemata. *Journal of Computer and system Sciences*, 3(2):147–195, 1969.
- [14] Jérôme Leroux. Vector addition systems reachability problem (a simpler solution). In *The Alan Turing Centenary Conference*, pages 214–228, June 2012.
- [15] Carl A Petri. Communication with automata. Technical report, DTIC Document, 1966.
- [16] David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.
- [17] 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, 2010. Preliminary version appeared in DNA 2008.

A Appendix

This appendix gives proofs that were not included in the main text.

A.1 Committed states unaffected by change in high-count species

Lemma 3.3. *For each CRD, there is a constant c such that, for all committed states \mathbf{c} , if $\mathbf{c}(S) > c$ for some $S \in \Lambda$, then for all $n \in \mathbb{Z}$, $\mathbf{c} + \{nS\}$ is also committed.*

Proof. The negation of a state \mathbf{u} being committed is that there exists a (possibly empty) sequence of reactions r such that $\mathbf{u} \rightarrow_r \mathbf{z}$ and there is a YES-voting species Y with $\mathbf{z}(Y) > 0$. Call such a state *uncommitted*, and let U be the set of all uncommitted states. Note that U is “closed under superset”: if $\mathbf{u} \in U$, then all $\mathbf{v} \geq \mathbf{u}$ are also uncommitted since r can be applied to \mathbf{v} as well. Hence it is clear that if $n \leq 0$, $\mathbf{c} + \{nS\}$ is committed, since $\mathbf{c} + \{nS\} \leq \mathbf{c}$.

By Higman’s Lemma, every set $U \subseteq \mathbb{N}^\Lambda$ has a finite number of minimal elements, i.e. vectors $\mathbf{m} \in U$ such that, for all $\mathbf{u} \in U$, $\mathbf{u} \leq \mathbf{m} \implies \mathbf{u} = \mathbf{m}$. Call this set of minimal elements M . Then for every $\mathbf{u} \in U$, there exists $\mathbf{m} \in M$ such that $\mathbf{m} \leq \mathbf{u}$. Thus, by our observation above that U contains all $\mathbf{u} \geq \mathbf{m}$ for any $\mathbf{m} \in U$, we conclude that $U = \bigcup_{\mathbf{m} \in M} \{ \mathbf{u} \mid \mathbf{u} \geq \mathbf{m} \}$, i.e., U is a finite union of “cones”. The lemma follows by choosing $c = \max_{\mathbf{m} \in M, S \in \Lambda} \mathbf{m}(S)$. \square

A.2 Detection problems are decidable by speed fault free CRDs

Since we argue that the CRD is fast under the standard stochastic time model [11] in addition to being speed fault free, we must first define this model. Since all rate constants in this paper are 1, we define time assuming this to be true.

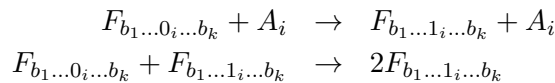
Given a fixed volume v and current state \mathbf{c} , the *propensity* of a unimolecular reaction $\alpha : X \rightarrow \dots$ in state \mathbf{c} is $\rho(\mathbf{c}, \alpha) = \mathbf{c}(X)$. The propensity of a bimolecular reaction $\alpha : X + Y \rightarrow \dots$, where $X \neq Y$, is $\rho(\mathbf{c}, \alpha) = \frac{\mathbf{c}(X)\mathbf{c}(Y)}{v}$. The propensity of a bimolecular reaction $\alpha : X + X \rightarrow \dots$ is $\rho(\mathbf{c}, \alpha) = \frac{1}{2} \frac{\mathbf{c}(X)(\mathbf{c}(X)-1)}{v}$. The propensity function determines the kinetics of the system as follows. The time until the next reaction occurs is an exponential random variable with rate $\rho(\mathbf{c}) = \sum_{\alpha \in R} \rho(\mathbf{c}, \alpha)$. The probability that next reaction will be a particular α_{next} is $\frac{\rho(\mathbf{c}, \alpha_{\text{next}})}{\rho(\mathbf{c})}$.

The kinetic model is based on the physical assumption of well-mixedness valid in a dilute solution. Thus, we assume the *finite density constraint*, which stipulates that the volume to execute a CRN must be proportional to the maximum molecular count obtained during execution [16].

Lemma 4.4. *Every detection set is decidable by a speed fault free CRD. This CRD takes expected time $O(\log n)$ to stabilize under the standard model of stochastic chemical kinetics.*

Proof. We describe the CRD using the language of “agent states” from population protocols. By this we mean that for each “agent” F with l different states, there are species F_1, \dots, F_l , and an agent “changing state” from i to j upon reacting with a molecule X means that we have a reaction of the form $X + F_i \rightarrow F_j + \dots$.

Let A_1, \dots, A_k the input species of the detection problem. F has 2^k states to keep track of the k bits “is A_i present?” for each $1 \leq i \leq k$, so that each fuel species is written $F_{b_1 \dots b_k}$ where each $b_i \in \{0, 1\}$, and the initial state of all n F ’s is $F_{0_1 \dots 0_k}$. For each $F_{b_1 \dots 0_i \dots b_k}$, we have the reactions



First we argue that the expected time to converge to an answer is $O(\log n)$. Each F species votes in accordance with the detection problem, according to whether its state indicates the answer is

yes or no. We start with n copies of F and at most $O(1)$ copies of each A_i .¹⁴ In expected time $\log k = O(1)$ with respect to n (since k is a constant depending on the detection problem but not on n), each A_i species with positive count has encountered at least one F molecule. It takes $O(\log n)$ expected time for a single bit to propagate to the entire population of F 's. Therefore it takes at most $kO(\log n) = O(\log n)$ (again, since k is a constant independent of n) for all bits to propagate. At this point all fuel agents have the same state, so they all vote unanimously and correctly.

Now we argue that the CRD is speed fault free. Let \mathbf{x} be any reachable state, and let A_i be a species with $\mathbf{x}(A_i) > 0$ that has not yet reacted with any fuel. Then all n fuel molecules have bit $b_i = 0_i$, so for at least one assignment to the other $k - 1$ bits, some fuel species has count at least $\frac{n}{2^{k-1}}$, so the first reaction above with that fuel species is $\frac{n}{2^{k-1}}$ -fast. Suppose that $\mathbf{x}(A_i) > 0$ but has reacted with at least one fuel. Then the second reaction, executed at most n times, will stabilize the value of bit b_i to be 1_i . Since the number of fuel molecules with bit 0_i plus the number with bit 1_i is exactly n , by a similar argument, there must be some assignment to all k bits such that a fuel species with that assignment has count at least $\frac{n}{2^k}$, whence the second reaction with that species is $\frac{n}{2^k}$ -fast. Therefore there is always a $\frac{n}{2^k}$ -fast path to an output-stable state with the correct answer, which occurs when all bits b_i corresponding to positive-count species A_i become 1_i , implying that the CRD is speed fault free. \square

A.3 Speed fault free CRDs decide only detection problems

A.3.1 Pumpable sets of species

Proposition 4.5. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable, with pumping sequence (\mathbf{x}_m) . Then there is a constant c such that, for all states \mathbf{y} and $m, n \in \mathbb{N}$ such that either 1) $\mathbf{z}_n \xrightarrow{\Pi} \mathbf{y} \xrightarrow{\Pi} \mathbf{x}_m$ or 2) $\mathbf{x}_m \xrightarrow{\Pi} \mathbf{y}$, then for all $S \in \Lambda \setminus \Pi$, $\mathbf{y}(S) < c$.*

Proof. Suppose otherwise. First we handle case (1). Then there exists $S \in \Lambda \setminus \Pi$ and an infinite number of $m, n \in \mathbb{N}$ and strictly increasing $k_m \in \mathbb{N}$ and states \mathbf{y}_m such that $\mathbf{z}_n \xrightarrow{\Pi} \mathbf{y}_m \xrightarrow{r_m} \mathbf{x}_m$, and $\mathbf{y}_m(S) \geq k_m$. Furthermore, by Higman's Lemma we can assume each $\mathbf{y}_m < \mathbf{y}_{m+1}$. Let $\mathbf{d}_{m,m'} = \mathbf{y}_{m'} - \mathbf{y}_m$, where m' is chosen sufficiently large that $k_{m'} \geq m + \mathbf{y}_m(S)$. Then $\mathbf{y}_{m'} - \{mS\} \geq \mathbf{y}_m$, so $\mathbf{y}_{m'} - \{mS\} \xrightarrow{r_m} \mathbf{x}_m + \mathbf{d}_{m,m'} - \{mS\}$, whence from $\mathbf{y}_{m'}$ we can reach a state (via r_m) with at least m copies of every species in Π (since it contains \mathbf{x}_m) and at least m copies of S (since those copies were not needed to execute r_m), contradicting the maximality of Π .

Case (2) follows by a similar argument. There exists $S \in \Lambda \setminus \Pi$ and an infinite number of $m \in \mathbb{N}$ such that $\mathbf{x}_m \xrightarrow{r_m} \mathbf{y}_m$ and $\mathbf{y}_m(S) \geq k_m$, where k_m increases without bound as $m \rightarrow \infty$. By Higman's Lemma we can assume each $\mathbf{x}_m < \mathbf{x}_{m+1}$, so that r_m is applicable to $\mathbf{x}_{m'}$ for all $m' > m$. Then by choosing m' sufficiently large, $\mathbf{x}_{m'} \xrightarrow{r_m} \mathbf{y}_{m'}$, where $\mathbf{y}_{m'}(S) \geq m$ and for all $X \in \Pi$, $\mathbf{y}_{m'}(X) \geq m'$, showing that $\Pi \cup \{S\}$ is Z -pumpable and contradicting the maximality of Π . \square

Corollary 4.6. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable, with pumping sequence (\mathbf{x}_m) . Let $c \in \mathbb{N}$ be the constant in Proposition 4.5 ensuring that in any state \mathbf{c} such that $\mathbf{x}_m \xrightarrow{\Pi} \mathbf{c}$, $\mathbf{c}(S) < c$ for all $S \in \Lambda \setminus \Pi$. Then if $\mathbf{x}_m \xrightarrow{p} \mathbf{c}'$ and $\mathbf{c}'(S) \geq c$, where $S \in \Lambda \setminus \Pi$, some reaction in p is not c -fast.*

Proof. Proposition 4.5 ensures that if p is Π -friendly, then no such $S \in \Lambda \setminus \Pi$ exists. Therefore p is not Π -friendly. Let $\alpha : X + Y \rightarrow \dots$ be the first reaction along p that is not Π -friendly. Since the state immediately preceding this reaction is reachable by a Π -friendly path, Proposition 4.5 tells us that all species $S \in \Lambda \setminus \Pi$ have count less than c . Therefore α is a bimolecular reaction between two molecules with count less than c , hence it is not c -fast. \square

¹⁴So that the total volume required is $O(n)$, although the initial count of A_i maybe be larger than n , we measure time asymptotically with respect to increasing n , so we consider the initial count of A_i to be constant with respect to n . With some care we could handle cases when the count of the A 's is much larger than n , simply by having all A molecules act like fuels as well, but since this complicates the exposition without changing the basic proof idea, we keep the presentation simply by assuming $n = \Omega(\#_0 A_i)$ for each i .

Proposition 4.7. *Let $Z = (\mathbf{z}_1 \leq \mathbf{z}_2 \leq \dots)$ be a infinite nondecreasing sequence of states, and let $\Pi \subseteq \Lambda$ be maximal Z -pumpable. For any species $X, Y \in \Pi$ such that R contains a reaction $X + Y \rightarrow \dots$, all products of the reaction are contained in Π .*

Proof. For each $m \in \mathbb{N}$, let \mathbf{x}_m be reachable from some \mathbf{z}_n such that $\mathbf{x}_m(S) \geq m$ for all $S \in \Pi$. If $X \neq Y$, then from state \mathbf{x}_m execute the reaction $X + Y \rightarrow \dots$ $m/2$ times. This results in a state in which all products of the reaction, as well as X and Y , have count at least $m/2$, and all other species in Π besides X and Y have count at least $m > m/2$. Since $m/2$ grows unboundedly, this establishes that the products are maximal Z -pumpable as well. If $X = Y$, then execute the reaction $m/3$ times, which ensures that the counts of X, Y , and all products are at least $m/3$, similarly establishing that the products are pumpable. \square

A.3.2 Main proof (of Lemma 4.11)

We say that $M \subseteq \mathbb{N}^\Lambda$ is a *monoid* if $\mathbf{0} \in M$ and $M + M \subseteq M$, i.e., M is closed under addition. Say that C is a *monoid coset* (a.k.a. *monoid offset*) if $C = \mathbf{b} + M$ for some constant vector $\mathbf{b} \in \mathbb{N}^\Lambda$ and monoid $M \subseteq \mathbb{N}^\Lambda$. We say that a set is *almost semilinear* if it is a finite union of monoid cosets.

For $X \subseteq \mathbb{N}^\Lambda$, let $\text{post}^\Pi(X) = \{ \mathbf{y} \in \mathbb{N}^\Lambda \mid (\exists \mathbf{x} \in X) \mathbf{x} \xrightarrow{\Pi} \mathbf{y} \}$ be the set of states reachable from some state in X , via Π -friendly reactions.

We require the following theorem, due to Leroux [14].

Theorem A.1 ([14]). *If $\Pi \subseteq \Lambda$ and $X \subseteq \mathbb{N}^\Lambda$ is semilinear, then $\text{post}^\Pi(X)$ is almost semilinear.*

Lemma 4.8. *Let Π be maximal I -pumpable. There exists $\mathbf{d} \in \mathbb{N}^\Lambda$, with $\mathbf{d}(S) > 0 \iff S \in \Pi$, and an infinite sequence of states $\mathbf{x}_1, \mathbf{x}_2, \dots$ such that, for all $m \in \mathbb{Z}^+$, $\mathbf{x}_m + \mathbf{d} = \mathbf{x}_{m+1}$, and there exists $n \in \mathbb{N}$ such that $\mathbf{i}_n \xrightarrow{\Pi} \mathbf{x}_m$.*

Proof. For each $m \in \mathbb{N}$, let \mathbf{x}'_m be reachable via Π -friendly reactions from some \mathbf{i}_n such that $\mathbf{x}'_m(S) \geq m$ for all $S \in \Pi$. Note that I is semilinear (in fact linear). By Theorem A.1, there exists $\mathbf{b}_1, \dots, \mathbf{b}_l \in \mathbb{N}^\Lambda$ and monoids $M_1, \dots, M_l \subseteq \mathbb{N}^\Lambda$ such that $\text{post}^\Pi(I) = \bigcup_{j=1}^l (\mathbf{b}_j + M_j)$. Let $n > \max_{1 \leq j \leq l, S \in \Lambda} \mathbf{b}_j(S)$.

Since $\mathbf{x}'_n \in \text{post}^\Pi(I)$, there is a j such that $\mathbf{x}'_n \in \mathbf{b}_j + M_j$.

Let $\mathbf{d} = \mathbf{x}'_n - \mathbf{b}_j$. Define $\mathbf{x}_1 = \mathbf{x}'_n$, and for all $m \in \mathbb{Z}^+$, define $\mathbf{x}_{m+1} = \mathbf{x}_m + \mathbf{d}$. Note that each $\mathbf{x}_m \in \mathbf{b}_j + M_j$ since M_j is closed under addition, so \mathbf{x}_m is reachable from sufficiently large \mathbf{i}_n . Since $\mathbf{x}'_n(S) > \mathbf{b}_j(S)$ for all $S \in \Pi$, $\mathbf{d}(S) = \mathbf{x}'_n(S) - \mathbf{b}_j(S) > 0$. The fact that $\mathbf{d}(S) > 0 \implies S \in \Pi$ follows from the maximality of Π : any species S for which $\mathbf{d}(S) > 0$ satisfies $\mathbf{x}_m(S) \geq m$, which implies that $S \in \Pi$. \square

The next two lemmas reference the sequences of states (\mathbf{x}_m) and (\mathbf{y}_m) as defined in Section 4.2.2.

Lemma 4.9. *Let sequences (\mathbf{x}_m) and (\mathbf{y}_m) be as defined above. For all $\epsilon \in \mathbb{N}$, there exists $\epsilon' \in \mathbb{N}$ such that the following holds. For all $\mathbf{e} \in \mathbb{Z}^\Pi$ with $|\mathbf{e}| \leq \epsilon$, for infinitely many m , there exists $\mathbf{e}_m \in \mathbb{Z}^\Gamma$ with $|\mathbf{e}_m| \leq \epsilon'$, and $m_2 < m$ such that $\mathbf{x}_m + \mathbf{e} \xrightarrow{\Pi} \mathbf{y}_{m_2} + \mathbf{e}_m$ and $\mathbf{y}_{m_2} + \mathbf{e}_m$ is committed.*

Proof. Note that the difficulty in the proof lies in handling species $S \in \Delta$ such that $\mathbf{e}(S) \neq 0$. (If \mathbf{e} is 0 on all of Δ then the lemma follows by letting $\mathbf{e}_m = \mathbf{e}$ and applying Lemma 3.3.)

We will prove the lemma by using reaction sequences p_i as defined before the lemma, which are Π -friendly, and we will modify them only by inserting or removing reactions already present in the p_i 's; hence they will remain Π -friendly. Therefore for the rest of the proof we do not explicitly annotate reachability relations between states to denote that they are Π -friendly.

Let $m_1, m_2 \in \mathbb{N}$ such that $\mathbf{y}_{m_1}, \mathbf{y}_{m_2} \in Y$. The proof will establish that if we will choose m_1 to be sufficiently large compared to \mathbf{e} and m_2 to be sufficiently large compared to m_1 and \mathbf{e} , then the lemma will hold for $m = 2m_2 - m_1$. In particular, choose m_1 sufficiently large that $f_{m_1} \geq \hat{c} + |\Lambda| \cdot c$, where f_{m_1} and c are defined before the statement of the lemma, and \hat{c} is defined below (but depends only on the CRN and not on c or m_1).

Let $k = m_2 - m_1$ and let $m = m_2 + k = 2m_2 - m_1$. Then defining the reaction sequence $q = \bigoplus_{i=m_1}^{m_2} p_i$, we have that $\mathbf{y}_{m_1} + k\mathbf{d} \xrightarrow{f_{m_1}^q} \mathbf{y}_{m_2}$. Recalling the definition of the reaction sequence $r_{m_1} = \bigoplus_{i=1}^{m_1} p_i$, we have that $\mathbf{x}_m = \mathbf{x}_{m_1} + 2k\mathbf{d} \xrightarrow{r_{m_1}} \mathbf{y}_{m_1} + 2k\mathbf{d} \xrightarrow{q} \mathbf{y}_{m_2} + k\mathbf{d}$. Let $\mathbf{d}_y = \mathbf{y}_{m_2} - \mathbf{y}_{m_1}$, noting that $\mathbf{d}_y(S) \geq 0$ for all $S \in \Lambda$ since $\mathbf{y}_{m_1} \leq \mathbf{y}_{m_2}$. Then $\mathbf{y}_{m_2} + k\mathbf{d} = \mathbf{y}_{m_1} + \mathbf{d}_y + k\mathbf{d} \xrightarrow{q} \mathbf{y}_{m_2} + \mathbf{d}_y$. Recall that all states in Y are committed. Further, recall that by the definition of Y , $\mathbf{d}_y(S) > 0$ if and only if $S \in \Gamma$, since all states in Y have equal counts on species in $\Lambda \setminus \Gamma$. Therefore by Lemma 3.3 (choosing m_2 sufficiently large that $\mathbf{y}_{m_2}(S) > c$ for all $S \in \Gamma$, where c is the constant in Lemma 3.3), $\mathbf{y}_{m_2} + \mathbf{d}_y$ is committed.

Now consider the state $\mathbf{x}_m + \mathbf{e}$. Choosing k large enough compared to ϵ , we have $\mathbf{x}_m + \mathbf{e} \xrightarrow{r_{m_1}} \mathbf{y}_{m_1} + 2k\mathbf{d} + \mathbf{e}$. (We can think of r_{m_1} processing the subset $\mathbf{x}_{m_1} < \mathbf{x}_m$ first, leaving the molecules represented by $2k\mathbf{d}$ alone until \mathbf{y}_{m_1} is reached from \mathbf{x}_{m_1} .)

Then consider applying q to $\mathbf{y}_{m_1} + 2k\mathbf{d} + \mathbf{e}$. We have $\mathbf{y}_{m_1} + 2k\mathbf{d} + \mathbf{e} \xrightarrow{q} \mathbf{y}_{m_2} + k\mathbf{d} + \mathbf{e}$. Since k was chosen larger than the absolute value of any element of \mathbf{e} , this is a valid reaction pathway. The key idea is that we now will do surgery on q to create a new reaction sequence q' , such that $\mathbf{y}_{m_1} + 2k\mathbf{d} + \mathbf{e} \xrightarrow{q'} \mathbf{y}_{m_2} + k\mathbf{d} + \mathbf{e}'$, where $\mathbf{e}' \in \mathbb{Z}^\Gamma$. That is, we show how to alter q through the addition or removal of reactions so that elements of $\Lambda \setminus \Gamma$ end up with exactly the counts they had in $\mathbf{y}_{m_2} + k\mathbf{d}$.

Since $\mathbf{d}_y \in \mathbb{N}^\Gamma$ and $\mathbf{e}' \in \mathbb{Z}^\Gamma$, and each component of \mathbf{d}_y increases unboundedly with k , we can ensure that $\mathbf{d}_y + \mathbf{e}' \geq 0$ on all elements by increasing k . (As we'll see below, \mathbf{e}' can be globally bounded as a function of \mathbf{e} and the CRN, and thus independently of m_1 and m_2 . Thus, $|\mathbf{e}'|$ is not unboundedly increasing with k and we avoid circularity.) Since $\mathbf{d}_y + \mathbf{e}' \geq 0$, we can apply q to $\mathbf{y}_{m_2} + k\mathbf{d} + \mathbf{e}'$, i.e., $\mathbf{y}_{m_2} + k\mathbf{d} + \mathbf{e}' = \mathbf{y}_{m_1} + \mathbf{d}_y + k\mathbf{d} + \mathbf{e}' \xrightarrow{q} \mathbf{y}_{m_2} + \mathbf{d}_y + \mathbf{e}'$. Then Lemma 3.3 implies that the resulting state $\mathbf{y}'_m = \mathbf{y}_{m_2} + \mathbf{d}_y + \mathbf{e}'$ is committed. Taking $\mathbf{e}_m = \mathbf{d}_y + \mathbf{e}'$ satisfies the statement of the lemma. Note that \mathbf{e}_m depends on m because \mathbf{d}_y does.

We now describe how we alter the reaction sequence q applied on $\mathbf{y}_{m_1} + 2k\mathbf{d} + \mathbf{e}$ to achieve these properties. First we need the following claim.

Claim (existence of the ordering \prec of Δ): There is an ordering \prec of Δ , so that we can write $\Delta = \{N_1, \dots, N_l\}$ to indicate $N_i \prec N_{i+1}$ for $1 \leq i < l$, such that for each $1 \leq i \leq l$, there is a reaction α_i with the following properties: 1) α_i has N_i as a reactant, 2) all products, and the possible other reactant (if α_i is bimolecular) are either elements of Γ , or are N_j for $j > i$, and 3) α_i occurs at least $\hat{c}/|R|$ times in q , where \hat{c} is a constant defined later.

Before proving this claim, we observe that it gives a way to “fix” q to ensure that the counts of species in $\Lambda \setminus \Gamma$ will be zero in \mathbf{e}' . We iteratively fix the counts of species in Δ one by one, in the ordering given, i.e. we first adjust q to fix N_1 , then we fix N_2 (while showing that the fixing of N_2 cannot affect the count of N_1 in any state, so it remains fixed), etc. We can define a sequence of $\mathbf{e}'_i \in \mathbb{Z}^{\Gamma \cup \{N_{i+1}, \dots, N_l\}}$ to represent the difference in counts compared to $\mathbf{y}_{m_2} + k\mathbf{d}$ resulting from fixing only the first i species N_1, \dots, N_i (with $\mathbf{e}'_0 = \mathbf{e}$, and $\mathbf{e}'_l = \mathbf{e}'$). The strategy for each i is this: If N_i is negative in \mathbf{e}'_{i-1} , then we remove enough instances of α_i from q ; property (3) ensures that q contains enough instances of α_i . If on the other hand N_i 's count in \mathbf{e}'_{i-1} is positive, then we add instances of α_i at the start of q . By property (2), adding or removing instances of α_i affects only the counts of species in Γ and N_{i+1}, \dots, N_l . Since we fix these counts in the prescribed order, when we are done, the counts of each N_i is equal to its count in $\mathbf{y}_{m_2} + k\mathbf{d}$, while counts of elements of Γ have been altered. Now, the altered reaction sequence q' might take species in $\Gamma \cup \Delta$ to counts lower than in the unaltered execution sequence traversed via q . However, as before, this does not cause a problem because we can arbitrarily increase k . Although q will change with k , circularity is avoided because there is a bound on the number of reaction instances added and removed as a function of \mathbf{e} and the CRN, independent of k .

Bound on the amount of fixing: Let $c_0 = \max_{S \in \Pi} \mathbf{e}(S)$ and let c_s be the maximum stoichiometry coefficient of any product (which bounds the amount that species can change each time the reaction is added or removed). Then we add or remove at most $\epsilon_1 = c_0$ instances of α_1 , which affects the count of N_2, \dots, N_l and species in Γ by at most $c_0 c_s$. Thus, $|\mathbf{e}'_1| \leq (1 + c_s)c_0$ (the original c_0 error plus the additional $c_0 c_s$ error from altering the number of α_1 reactions). At this point, in the worst case, the value of N_2 in \mathbf{e}'_1 is $(1 + c_s)c_0$, and we have to add or remove $\epsilon_2 = (1 + c_s)c_0$ instances of

α_2 , resulting in $|\mathbf{e}'_2| \leq (1 + c_s)^2 c_0$. In general, $\epsilon_i = (1 + c_s)^{i-1} c_0$, and $\mathbf{e}'_i \leq (1 + c_s)^i c_0$. This implies that $\sum_{i=1}^l \epsilon_i = ((1 + c_s)^l - 1) c_0 / c_s$ is an upper bound on the total number of reaction instances we add or remove, and $|\mathbf{e}'_l| = |\mathbf{e}'| \leq (c_s + 1)^l c_0$ is an upper bound on the species in Γ we generate or consume when fixing species in Δ .

Finally, we show that the ordering \prec of Δ of the above claim exists. The ordering follows from the fact that the original execution sequence where q was defined is fast: i.e. the execution sequence $\mathbf{y}_{m_1} + k\mathbf{d} \rightarrow_q \mathbf{y}_{m_2}$ is fast. We show that there exists increasing sets $\emptyset = \Delta_0 \subset \Delta_1 \subset \Delta_2 \subset \dots \Delta_{l-1} \subset \Delta_l = \Delta$, where for each $1 \leq i \leq l$, $\Delta_i \setminus \Delta_{i-1} = \{N_i\}$.

We define the ordering inductively “in reverse.” For all $1 \leq i \leq l$, define $\Phi_i : \mathbb{N}^\Lambda \rightarrow \mathbb{N}$ for all states \mathbf{c} by $\Phi_i(\mathbf{c}) = \sum_{N \in \Delta_i} \mathbf{c}(N)$. Φ_l is well-defined since $\Delta_l = \Delta$, and Φ_i is well-defined once we have defined N_{i+1}, \dots, N_l , because $\Delta_i = \Delta \setminus \{N_{i+1}, \dots, N_l\}$. Because $\mathbf{y}_{m_2} \in Y$, there is a constant c such that $\mathbf{y}_{m_2}(S) \leq c$ for all $S \in \Delta$, so it follows that $\Phi_i(\mathbf{y}_{m_2}) \leq i \cdot c$.

Define $\hat{c} = |R| \cdot (c_s + 1)^l c_0$, where $|R|$ is the number of reactions in the CRN. Because $\mathbf{d}(S) > 0$ for all $S \in \Pi$, $\Phi_i(\mathbf{y}_{m_1} + k\mathbf{d}) \geq k \cdot i$. Therefore, for $k \geq (\hat{c} + i \cdot c) / i$, the sequence q of reactions leading from $\mathbf{y}_{m_1} + k\mathbf{d}$ to \mathbf{y}_{m_2} must contain a subsequence of reactions decreasing Φ_i from at least $\hat{c} + i \cdot c$ to at most $i \cdot c$. Let r be the part of q after the last time Φ_i is above $\hat{c} + i \cdot c$.

Recall we chose m_1 sufficiently large such that $f_{m_1} \geq \hat{c} + |\Lambda| \cdot c \geq \hat{c} + i \cdot c$. Since the largest value of Φ_i on r is $\hat{c} + i \cdot c$, any bimolecular reaction $N_i + N \rightarrow \dots$ in r , where $N \in \Delta_i$, would not be f_{m_1} -fast, since $\Phi_i(\mathbf{c}) \leq \hat{c} + i \cdot c$ implies that for all $N \in \Delta_i$, $\mathbf{c}(N) \leq \hat{c} + i \cdot c \leq f_{m_1}$. Hence every reaction in r is either unimolecular, or if it is bimolecular it is of the form $N_i + S \rightarrow \dots$ where $S \in \Gamma \cup \{N_{i+1}, \dots, N_l\}$, i.e., the other reactant is either in Γ or follows N_i in the ordering \prec . This shows that no such reaction can alter the count of a preceding $N \prec N_i$ by *consuming* it; however, we must handle the case that N could potentially be a product. Since exactly one reactant (N_i) is in Δ_i , if any product is in this set, then that reaction cannot strictly decrease Φ_i . However, r has at least \hat{c} reactions that decrease Φ_i , since it takes Φ_i from $\hat{c} + i \cdot c$ to $i \cdot c$. Therefore, there is a reaction consuming N_i occurring at least $\hat{c} / |R|$ times along r that strictly decreases Φ_i (it decreases Φ_i by exactly one, since the other reactant cannot be in Δ_i). By our choice of \hat{c} , this is sufficiently many instances of α_i to adjust the count of N_i upwards (by removing α_i from r) by up to $(c_s + 1)^l c_0$, which as we observed is an upper bound on the maximum amount that we may need to adjust the count of N_i . This reaction α_i is our desired reaction, which can be added to q at the start to decrease the count of N_i as needed, or removed from r to increase the count of N_i as needed. \square

DS: Clarify that this is where N_i is defined.

Lemma 4.11. \mathcal{D} is not speed fault free.

Proof. For the sake of contradiction suppose otherwise. Let $\Theta = \Lambda \setminus \Pi$. For any state \mathbf{c} , we call $\mathbf{c} \upharpoonright \Theta$ the *base* of \mathbf{c} . By Proposition 4.5, any state reachable by Π -friendly reactions from any of the \mathbf{x}_m 's, has its counts of elements of Θ bounded by a constant c . Therefore there are only a finite number of bases of such states (at most $c^{|\Theta|}$); call this number c_b . Let $c_s = \max_{(\mathbf{r}, \mathbf{p}) \in R} |\mathbf{p}|$ be the maximum stoichiometry coefficient of any product of any reaction. Define $\epsilon = c_s \cdot c_b$. (We explain this choice of ϵ later in the proof.)

For all \mathbf{e} such that $|\mathbf{e}| \leq \epsilon$, let $W_{\mathbf{e}} = (\mathbf{w}_n)$ be the sequence given by Lemma 4.10. By Higman's Lemma there is an infinite nondecreasing subsequence of $W_{\mathbf{e}} = (\mathbf{w}_n)$. For each \mathbf{e} , we consider only values of n such that \mathbf{w}_n is in this subsequence.

Let $B = \{ \mathbf{b} \in \mathbb{N}^\Theta \mid (\exists n, m \in \mathbb{N}, \exists \mathbf{e} \in \mathbb{Z}^\Pi, |\mathbf{e}| \leq \epsilon) \mathbf{x}_m + \mathbf{e} \rightarrow^\Pi \mathbf{w}_n, \mathbf{b} = \mathbf{w}_n \upharpoonright \Theta, \mathbf{w}_n \in W_{\mathbf{e}} \}$ be the set of all bases of states in the sequence $W_{\mathbf{e}}$. We choose a “large” value of n and focus on the state \mathbf{w}_n in Lemma 4.10. Then, because $\lim_{n \rightarrow \infty} \mathbf{w}_n(S) = \infty$ (regardless of which $W_{\mathbf{e}}$ sequence we use) for all $S \in \Omega$, it follows that for any $\mathbf{b} \in B$, \mathbf{b} together with any amount of species in Ω is a subset of some \mathbf{w}_n , hence is committed because all \mathbf{w}_n are committed.

In an initial state $\mathbf{i} = \mathbf{i}'(n_1, \dots, n_k)$ for some amount n' of fuel F and some values n_1, \dots, n_k of A_1, \dots, A_k , consider the base $\mathbf{i} \upharpoonright \Theta$ of \mathbf{i} . We restrict attention to a subset of $\mathbf{i} \upharpoonright \Theta$ in the following way. Recall that we assume $n_1 \geq 1$. Consider the following process. To get from \mathbf{i} to \mathbf{x}_m for some m , a series of reactions are applied. We imagine taking one copy of A_1 and “tracking its base descendants” along the reaction sequence that leads from \mathbf{i} to \mathbf{x}_m , and subsequently from $\mathbf{x}_m + \mathbf{e}$. In other words, the first reaction with our chosen copy of A_1 as a reactant, we look at the products in

Θ and consider them (first generation) “base descendants” of A_1 . Each of those in turn undergoes more reactions, and we track their descendants. More formally, define the multiset $\mathbf{b}_1 = \{A_1\}$. The first reaction with this copy of A_1 , if it produces $C_1, C_2, C_3, \dots, C_l \in \Theta$ (in addition to other products that may be in Π) as products, defines the set $\mathbf{b}_2 = \{C_1, C_2, C_3, \dots, C_l\}$. The first of *those* to react next, for example, C_1 , if it produces products $D_1, D_2, \dots, D_p \in \Theta$, defines the multiset $\mathbf{b}_3 = \{D_1, D_2, \dots, D_p, C_2, C_3, \dots, C_l\}$. Because each reaction changing this multiset is Π -friendly, at most one reactant is in Θ . That is to say, because we consider only Π -friendly reactions, it is well-defined to consider the base descendants of a single copy of A_1 , since they define a tree in which every child (product of a reaction in Θ) has exactly one parent (reactant in Θ).

Now, the state \mathbf{w}_n is defined to be reachable from $\mathbf{x}_m + \mathbf{e}$, not \mathbf{x}_m itself; however $\mathbf{e} \in \mathbb{Z}^\Pi$ implies that the base of \mathbf{x}_m is the same as the base of $\mathbf{x}_m + \mathbf{e}$. Hence it is well-defined to continue tracking the base descendants from the state $\mathbf{x}_m + \mathbf{e}$, all the way to \mathbf{w}_n . Therefore, each $\mathbf{b} \in B$ can be interpreted as a union of base descendants of molecules in the initial state $\mathbf{i}_{n'}(n_1, \dots, n_k)$. We choose any such $\mathbf{b} \in B$; then the base descendants of one copy of A_1 are a subset \mathbf{b}' of \mathbf{b} . Our goal is to evolve the extra copy of A_1 to \mathbf{b}' . Since the number of base states depends only on the CRN and the choice of Π , at most c_b (recall c_b is the total number of base states) reaction applications are required to evolve $\{A_1\}$ to \mathbf{b}' . Recall c_s is the maximum stoichiometry coefficient of any product. Then each reaction application that takes $\{A_1\}$ to \mathbf{b}' can affect the count of species in Π by at most c_s . Therefore, the total change in count of species in Π is bounded by $c_s \cdot c_b$. This is why we take $\epsilon = c_s \cdot c_b$. Applying these reactions to the extra A_1 in state \mathbf{x}_m results in the state $\mathbf{x}_m + \mathbf{e} + \mathbf{b}'$, where $\mathbf{b}' \in \mathbb{Z}^\Theta$, $\mathbf{e} \in \mathbb{Z}^\Pi$, and $|\mathbf{e}| \leq \epsilon$. In other words, once we are in a state in which every species in Π is “high”, we execute just enough reactions to transform $\{A_1\}$ into a base state “consistent with some \mathbf{w}_n state,” and we separate out the change that results in species in Π , calling it \mathbf{e} , and the change that results in the base, calling it \mathbf{b}' .

Therefore, since $\mathbf{x}_m + \mathbf{e} \xrightarrow{\Pi} \mathbf{w}_n$, it follows that $\mathbf{x}_m + \mathbf{e} + \mathbf{b}' \xrightarrow{\Pi} \mathbf{w}_n + \mathbf{b}'$. Furthermore, \mathbf{w}_n is committed. Since $\mathbf{b}' \leq \mathbf{b}$, where $\mathbf{b} \in B$ is by definition a subset of a committed state (\mathbf{w}'_n for some other \mathbf{w}'_n with “large” counts of species in Ω , possibly reachable from $\mathbf{x}_m + \mathbf{e}'$ for a different choice of \mathbf{e}' such that $|\mathbf{e}'| \leq \epsilon$), it follows that $\mathbf{b}' + \mathbf{d}$, where $\mathbf{d} \in \mathbb{N}^\Omega$ is an arbitrary state of species in Ω , is also committed by Lemma 3.3.

We take the state $\mathbf{w}_n + \mathbf{b}'$, which is not committed (it is reachable from $\mathbf{i}_{n'}(n_1 + 1, \dots, n_k)$, which has answer YES, so any state reachable from it must be able to produce a YES voter), and split it into two states, \mathbf{l}_n and \mathbf{r}_n , where \mathbf{r}_n is simply \mathbf{b}' plus $n/2$ copies of each species in Ω , and \mathbf{l}_n is what is left, i.e., \mathbf{w}_n minus $n/2$ copies of each species in Ω . Because $\mathbf{w}_n(S) \geq n$ for each $S \in \Omega$, \mathbf{l}_n and \mathbf{r}_n each have at least $n/2$ copies of each species in Ω . Because $\mathbf{l}_n < \mathbf{w}_n$, and \mathbf{w}_n is committed, it follows that \mathbf{l}_n is committed. The preceding paragraph establishes that \mathbf{r}_n is committed, since it is \mathbf{b}' plus some amount of species in Ω .

Let c be the constant c in Proposition 4.5 (taking Ω to be the maximal Z -pumpable set, where Z is as defined in the proof of Lemma 4.10), such that no species in $\Lambda \setminus \Omega$ can have count c in any state reachable by Ω -friendly reactions from any \mathbf{w}_n . Since $\mathbf{l}_n + \mathbf{r}_n$ is reachable from $\mathbf{i}_{n'}(n_1 + 1, \dots, n_k)$ and \mathcal{D} is speed fault free, for sufficiently large n there is a c -fast path p from $\mathbf{l}_n + \mathbf{r}_n$ that produces a YES voter, i.e., $\mathbf{l}_n + \mathbf{r}_n \xrightarrow{c}_p \mathbf{y}$, where \mathbf{y} contains a YES voter. By Corollary 4.6, p is Ω -friendly; hence every bimolecular reaction has at least one reactant in Ω . Let q be the join of p from $(\mathbf{l}_n, \mathbf{r}_n)$, and let (q_1, q_2) be the parallel decomposition of q from $(\mathbf{l}_n, \mathbf{r}_n)$. Since \mathbf{l}_n and \mathbf{r}_n are committed, they cannot produce a YES voter on their own. Hence $q \neq p$, otherwise one of q_1 or q_2 would be applicable to \mathbf{l}_n or \mathbf{r}_n to produce a YES voter, a contradiction.

Let \mathbf{l}'_n and \mathbf{r}'_n be such that $\mathbf{l}_n \rightarrow_{q_1} \mathbf{l}'_n$ and $\mathbf{r}_n \rightarrow_{q_2} \mathbf{r}'_n$. The next reaction in p after q therefore is bimolecular $\alpha : L + R \rightarrow \dots$, where $\mathbf{l}'_n(R) = 0$ and $\mathbf{r}'_n(L) = 0$. Since α is bimolecular and p is Ω -friendly, at least one reactant L or R must be an element of Ω (perhaps both are in Ω).

Choose a larger value $n' > n$ and consider applying the reaction sequence p to $\mathbf{l}_{n'} + \mathbf{r}_{n'}$, where n' is chosen such that $\mathbf{l}_{n'} + \mathbf{r}_{n'} \geq \mathbf{l}_n + \mathbf{r}_n$ so p is applicable, and such that $\mathbf{l}_{n'}$ and $\mathbf{r}_{n'}$ both have strictly larger counts of L if $L \in \Omega$ and strictly larger counts of R if $R \in \Omega$. Let $\mathbf{l}'_{n'}$ and $\mathbf{r}'_{n'}$ be such that $\mathbf{l}_{n'} \rightarrow_{q_1} \mathbf{l}'_{n'}$ and $\mathbf{r}_{n'} \rightarrow_{q_2} \mathbf{r}'_{n'}$. If $R \in \Omega$, then $\mathbf{l}'_{n'}(R) > \mathbf{l}_n(R)$ and q consumed the same amount of R , implying $\mathbf{l}'_{n'}(R) > 0$. Similarly, $\mathbf{r}'_{n'}(L) > 0$ if $L \in \Omega$. This implies that the instance of reaction α after q is in fact applicable to one of $\mathbf{l}'_{n'}$ or $\mathbf{r}'_{n'}$. Therefore the join can be extended by at least one

reaction to include all of q and the reaction after it.

However, extending the join in this way required only choosing a larger n' and keeping the reaction sequence p fixed. Since there is an infinite sequence of values of n that allow the join to be extended, this implies that the join can be extended to be all of p . This means that for sufficiently large n'' , one of $\mathbf{l}_{n''}$ or $\mathbf{r}_{n''}$ can produce a YES voter on its own, contradicting the fact that $\mathbf{l}_{n''}$ and $\mathbf{r}_{n''}$ are committed for all values of n'' . \square