

Majority consensus thresholds in competitive Lotka–Volterra populations

Matthias Függer · mfuegger@lmf.cnrs.fr

Université Paris-Saclay, CNRS, ENS Paris-Saclay, LMF

Thomas Nowak · thomas@thomasnowak.net

Université Paris-Saclay, CNRS, ENS Paris-Saclay, LMF & Institut Universitaire de France

Joel Rybicki · joel.rybicki@hu-berlin.de

Humboldt University of Berlin

Abstract. One of the key challenges in synthetic biology is devising robust signaling primitives for engineered microbial consortia. In such systems, a fundamental signal amplification problem is the majority consensus problem: given a system with two input species with initial difference of Δ in population sizes, what is the probability that the system reaches a state in which only the initial majority species is present?

In this work, we consider a discrete and stochastic version of competitive Lotka–Volterra dynamics, a standard model of microbial community dynamics. We identify new threshold properties for majority consensus under different types of interference competition:

- We show that under so-called self-destructive interference competition between the two input species, majority consensus can be reached with high probability if the initial difference satisfies $\Delta \in \Omega(\log^2 n)$, where n is the initial population size. This gives an *exponential* improvement compared to the previously known bound of $\Omega(\sqrt{n \log n})$ by Cho et al. [Distributed Computing, 2021] given for a special case of the competitive Lotka–Volterra model. In contrast, we show that an initial gap of $\Delta \in \Omega(\sqrt{\log n})$ is necessary.
- On the other hand, we prove that under *non-self-destructive* interference competition, an initial gap of $\Omega(\sqrt{n})$ is necessary to succeed with high probability and that a $\Omega(\sqrt{n \log n})$ gap is sufficient.

This shows a strong qualitative gap between the performance of self-destructive and non-self-destructive interference competition. Moreover, we show that if in addition the populations exhibit interference competition between the individuals of the *same* species, then majority consensus cannot always be solved with high probability, no matter what the difference in the initial population counts.

1 Introduction

Synthetic biology is a discipline focusing on the rational engineering of biological systems [17, 18]. Early results built basic computational modules, such as memory [36], clocks [25, 32], and sensors [53] in the bacterium *Escherichia coli*. Recently, synthetic biologists have started to engineer *synthetic consortia* consisting of multiple interacting microbial species that collectively implement *distributed* biological circuits [14, 56, 66, 70]. This has lead the bioengineering community to face a fundamental challenge of distributed systems: the need for robust *coordination primitives* to coordinate the activities of all different microbial populations comprising the circuit [5, 14, 44, 56].

Synthetic biologists have recognized that (a) such problems are studied in the field of distributed computing and that (b) the existing models of distributed computing poorly capture key aspects of microbial systems [12, 27, 45, 50]. Classical models of distributed computing ignore even the most elementary *ecological processes* that take place in microbial consortia, such as the stochastic reproduction and mortality of individual cells and competition between different species [42, 73].

This limitation holds even for many popular models of distributed computing explicitly motivated by biological computation, such as the widely studied population protocol model [1, 8, 10, 13, 30]. Unfortunately, these fundamental ecological processes are (1) in general unavoidable in synthetic microbial consortia, (2) lead to stochastic fluctuations in the community size and composition, which are key drivers of microbial population dynamics [73], and importantly, (3) dealing with them is one of the main challenges in engineering microbial consortia [14, 44, 50].

While theoretical foundations of molecular and biological computation have gained increasing attention in recent years [3, 20, 29, 34, 35, 63, 65], there is scarcely any work focusing on the theory of distributed computing in *microbial consortia*, despite the fact that distributed computing has become a fundamental paradigm in synthetic biology. Here, we initiate the complexity theoretic study of the *computational power of programmable ecological interactions* [54] in microbial circuits.

1.1 The problem: majority consensus

Majority consensus is a fundamental problem in distributed computing [2, 8, 22, 23, 30, 31, 33, 61]. In this problem, each node in the system is given a local input bit, and the task is for each node to output the input bit given to the majority of the nodes.

Majority consensus has also been identified as a useful signaling primitive for engineered microbial consortia in theory [21] and practice [5]. The goal is to design a genetic circuit, where the (gene expression) state of each cell is determined by the species which is in the majority [5]. This primitive can be used as a robust differential signal amplifier [21] when composing complex circuits implemented by different populations [66]. However, it remains unclear how to efficiently solve this problem in the microbial setting, where individuals replicate, die, and compete with one another.

The performance of majority consensus protocols is typically studied as a function of the initial population size n and the initial difference Δ between the counts of the majority and minority input species. In *exact* majority consensus, the goal is to correctly compute the majority with probability 1 for any $\Delta > 0$, whereas in *approximate* majority consensus, the protocols are allowed to fail with a probability that depends on n and Δ . Typically, the aim is to obtain protocols that succeed with high probability in n for as small as possible Δ . Intuitively, the smaller Δ a protocol can deal with, the better the protocol tolerates noise: in the synthetic biology setting, such protocols are instrumental in amplifying signals produced by noisy (biological) processes, e.g., microbial subcircuits.

For example, in the stochastic population protocol model [7], where a random scheduler picks pairs of nodes to interact uniformly at random in each time step, both exact [1, 4, 30, 31, 61] and approximate [8, 24] versions of majority consensus have been studied. There is a simple (but

challenging to analyze) 3-state protocol [8] that solves approximate majority consensus in $O(n \log n)$ interactions with high probability whenever $\Delta \in \Omega(\sqrt{n} \cdot \log n)$. This protocol can in principle implement a cell cycle switch in cellular populations [19]. In contrast, for *exact* majority consensus, it is known that any $O(1)$ -state protocol requires $\Omega(n^2)$ expected interactions [1], but there is a $O(\log n)$ -state protocol that solves the problem in $O(n \log n)$ interactions in expectation [30].

1.2 Our focus: from molecular to microbial computation

The population protocol model captures uncertainty arising from unpredictable *interaction patterns*, but it does not capture *demographic noise*, i.e., the stochastic fluctuations in the community composition that arise from the chance events of the underlying biomolecular [12, 72] and ecological processes [42, 52, 73] taking place in microbial systems. For small populations, as in the case of synthetic microbial consortia, demographic noise is known to have a significant impact on the realized dynamics.

In this work, we take steps towards developing a theory of microbial computation that investigates the *computational power of ecological interactions*. Motivated by the recent experimental [5, 54] and mathematical modeling work utilizing ecological competitive mechanisms [6, 21, 57] in the design of distributed majority consensus algorithms in synthetic microbial consortia, we study the following question at the interface of theoretical computer science and synthetic ecology:

How does (a) demographic noise and (b) the choice of engineered, competitive mechanisms impact the performance of microbial majority consensus protocols?

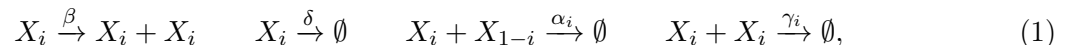
We investigate the majority consensus problem in discrete Lotka–Volterra models of well-mixed, competitive microbial communities. These are standard models used in microbiology that explicitly model the reproductive and competitive dynamics of microbial species over time [26, 42, 73]. The analysis of *stochastic* population models is considered important but highly challenging in the biomathematics [15, 38, 49] and statistical physics communities [28, 48, 60, 67]. Recently, their rigorous analysis has started to gain interest also in theoretical computer science [6, 21, 24].

We develop new techniques to analyze the behavior of majority consensus dynamics in stochastic, competitive Lotka–Volterra models, and show that in principle (1) ecological processes can be *algorithmically exploited* to obtain robust majority consensus protocols, and that (2) different ecological mechanisms can have substantial impact on the performance of microbial protocols.

1.3 The model: stochastic Lotka–Volterra dynamics

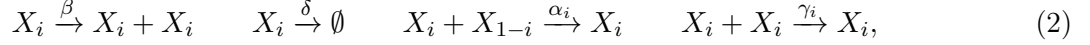
We now introduce the stochastic two-species Lotka–Volterra (LV) models of competing microbial populations. These LV models generalize the models of microbial majority consensus previously studied in the distributed computing community [6, 21]. Lotka–Volterra dynamics can be derived mechanistically from several different biological assumptions. Here, we have derived the models assuming interference competition [43] between the species, as interference competition can be programmed into synthetic microbial populations using readily available genetic modules [54].

The Lotka–Volterra models. As typical for non-spatial biological population models [15, 72], our models are formally represented as chemical reaction networks [16, 40], assuming unit volume. For the sake of brevity, we do not give the full general definition of the chemical reaction network formalism, as we only work with two basic models. Following the standard reaction kinetics notation, the Lotka–Volterra model with *self-destructive* interference competition is given by



where X_i denotes an individual of species $i \in \{0, 1\}$, $\emptyset$ denotes the removal of each reactant species and $\alpha_i, \beta, \delta, \gamma_i \geq 0$ are constant rate parameters. We note that here we treat the reactions with reactants $X_0 + X_1$ and $X_1 + X_0$ of different species formally as different reactions unlike typically in chemical reaction network formalism – this purely for notational convenience.

Our second model is the Lotka–Volterra model with *non-self-destructive* competition given by



for $i \in \{0, 1\}$. We often write $\alpha = \alpha_0 + \alpha_1$ and $\gamma = \gamma_0 + \gamma_1$. We say that the system is *neutral* if both species have identical rate parameters. Unless otherwise specified, we consider neutral communities. In both models, the reactions have at most two reactants. We call reactions with a single reactant *individual reactions* and reactions with two reactants *pairwise interactions* between individuals.

Biological interpretation. We assume a well-mixed setting as encountered in a bioreactor, with a growing microbial community consisting of two populations of different species. We primarily focus on the early stages of the microbial population dynamics, the so-called exponential phase of microbial growth, where the population size is far from the carrying capacity of the environment (e.g., growth is not limited by availability of nutrients or space).

Each individual of input species $i \in \{0, 1\}$ reproduces at per-capita rate $\beta \geq 0$ and dies at per-capita rate $\delta \geq 0$. In addition to these reproductive dynamics, we assume that interactions between individuals happen via interference competition: individuals of species i encounter and individuals of species $1 - i$ at rate $\alpha_i \geq 0$, and individuals of the same species i encounter and kill each other at rate $\gamma_i \geq 0$. Competition between individuals of different species is called *interspecific*, whereas competition between individuals of the same species is *intraspecific*.

When species engage in a competitive interaction, the outcome can be either symmetric (both individuals die) or asymmetric (only one individual dies). These two cases correspond to the two different models (1) and (2), and they have biologically very distinct interpretations. The former scenario corresponds to *self-destructive interference* (e.g., cells release a bacteriocin via lysis), whereas the latter scenario corresponds to *non-self-destructive interference* (e.g., cells secrete a bacteriocin or puncture the membrane of other cells on physical contact). Both types of competition are exhibited by bacterial species [43]. Non-self-destructive interference can also be implemented in engineered populations using, e.g., programmable plasmid conjugation [58], as suggested by Cho et al. [21].

Stochastic kinetics. For both model variants, a *configuration* is a vector $\mathbf{x} = (x_0, x_1) \in \mathbb{N}^2$, where x_i gives the count of species i in the configuration. In a configuration (x_0, x_1) , the individual birth reactions of species i have propensity βx_i and individual death reactions have propensity δx_i . The propensity of an interspecific competition reaction is $\alpha x_0 x_1$ and the propensity of an intraspecific competition reaction is $\gamma x_i(x_i - 1)/2$. The total propensity of the configuration (x_0, x_1) is given by

$$\varphi(x_0, x_1) = \sum_{i \in \{0, 1\}} (\alpha_i x_0 x_1 + \beta x_i + \delta x_i + \gamma_i x_i(x_i - 1)/2).$$

Under continuous-time stochastic kinetics, the time until the next reaction in a configuration $\mathbf{x}$ is exponentially distributed with rate $\varphi(\mathbf{x})$. Given that the current configuration is $\mathbf{x}$, the probability that R is the next reaction is $\varphi_R(\mathbf{x})/\varphi(\mathbf{x})$, where $\varphi_R(\mathbf{x})$ is the propensity of the reaction (as above).

The *stochastic trajectory* under stochastic kinetics is given by a continuous-time Markov process $\mathbf{X} = (\mathbf{X}_t)_{t \geq 0}$ on the state space $\mathbb{N}^2$, where $Q(\mathbf{x}, \mathbf{y})$ gives the propensity of the chain from transitioning from state $\mathbf{x}$ to state $\mathbf{y}$. In this work, we primarily focus on the *jump chain* $\mathbf{S} = (\mathbf{S}_t)_{t \geq 0}$ of $\mathbf{X}$, which

Competition	Self-destructive	Non-self-destructive	Reference
Interspecific only	$\Omega(\sqrt{\log n}) - O(\log^2 n)$	$\Omega(\sqrt{n}) - O(\sqrt{n \log n})$	Sec. 6 and Sec. 7
Both inter- and intraspecific	$\geq n - 1$	$\geq n - 1$	Sec. 8.1
Intraspecific only	∞	∞	Sec. 8.2
Interspecific and $\delta = 0$	$O(\sqrt{n \log n})$	$O(\sqrt{n \log n})$ (*)	[21] and [6]
No competition ($\alpha = \gamma = 0$)	$n - 1$	$n - 1$	[6]

Table 1: Worst-case majority consensus thresholds for different LV models. The first three rows are new results, the last two results are from prior work. (*) The model of [6] is not strictly a special case of the LV model, as it assumes bounded, non-mass-action birth rates.

is the discrete-time Markov chain given by the transition probabilities $P(\mathbf{x}, \mathbf{y}) = Q(\mathbf{x}, \mathbf{y})/\varphi(\mathbf{x})$ for $\varphi(\mathbf{x}) > 0$. If $\varphi(\mathbf{x}) = 0$, then $P(\mathbf{x}, \mathbf{x}) = 1$ and $P(\mathbf{x}, \mathbf{y}) = 0$ for $y \neq x$. Here, $\mathbf{S}_t$ represents the counts of both species after $t \in \mathbb{N}$ reactions have occurred.

Majority consensus. We say that species $i \in \{0, 1\}$ in a state $\mathbf{S}_t = (x_0, x_1)$ is the *majority species* if $x_i > x_{1-i}$ holds. In particular, we say that the majority species in the initial configuration $\mathbf{S}_0$ is the *initial majority species*. The species who is not the initial majority species is the *initial minority species*. We say that a configuration (x_0, x_1) has *reached consensus* if $x_0 = 0$ or $x_1 = 0$. In such a configuration, we say that the species i has won if $x_i > 0$ and $x_{1-i} = 0$. i.e., the species i is the majority species in the configuration (x_0, x_1) . We define the *consensus time* of the chain $\mathbf{S}$ to be

$$T(\mathbf{S}) = \inf\{t : \mathbf{S}_t \text{ has reached consensus}\}$$

the minimum time until (at least) one of the species goes extinct. We say that the chain $\mathbf{S}$ *reaches majority consensus* if $T(\mathbf{S})$ is finite and the initial majority species has positive count at time $T(\mathbf{S})$. We define $\rho(\mathbf{S})$ to be the probability that the chain $\mathbf{S}$ reaches majority consensus.

Without loss of generality, we assume throughout that the first species is the initial majority species, i.e., $\mathbf{S}_0 = (a, b)$ for some $a > b > 0$. We use $n = a + b$ to denote the size of the total initial population. For every time step $t \geq 0$, we define $\Delta_t = S_{t,0} - S_{t,1}$ to be the difference between the counts of the initial majority and the initial minority species. In particular, Δ_0 gives the initial gap between the initial majority and minority species.

1.4 Our contributions

In this work, we are interested in how the probability $\rho(\mathbf{S})$ of majority consensus behaves as a function of the initial gap Δ_0 under stochastic kinetics. To this end, we say that $\Psi(n) \geq 0$ is a *majority consensus threshold* for a Lotka–Volterra model if $\rho(\mathbf{S}) \geq 1 - 1/n$ if and only if $\Delta_0 \geq \Psi$. We identify the asymptotic (and sometimes exact) majority consensus thresholds for two-species competitive Lotka–Volterra systems under different modes of inference competition. Our main results are summarized in Table 1 and are as follows.

Interspecific competition. In systems with interspecific competition and no intraspecific competition (i.e., the case $\alpha > 0$ and $\gamma = 0$) we show the following results. First, for self-destructive interference competition, we show that the majority consensus threshold lies in a polylogarithmic range between $\Omega(\sqrt{\log n})$ and $O(\log^2 n)$. This is an *exponential improvement* in the previous upper bound of $O(\sqrt{n \log n})$ for self-destructive competition shown by Cho et al. [21] in the special case

with $\delta = 0$ (i.e., a model with no individual death reactions). Our result applies to a much larger class of models: for any choice of $\beta, \delta, \alpha > 0$, the chain reaches majority consensus with high probability provided that the gap is $\Omega(\log^2 n)$. In contrast, we show that if the gap is $o(\sqrt{\log n})$ the chain fails to reach majority consensus with constant probability.

Second, for non-self-destructive interference competition, we show that the majority threshold lies in a polynomial range between $\Omega(\sqrt{n})$ and $O(\sqrt{n \log n})$. This shows an *exponential separation* between competitive Lotka–Volterra models with self-destructive and non-self-destructive interspecific interference competition in these models.

In comparison, recently Andaur et al. [6] gave an upper bound of $O(\sqrt{n \log n})$ for a slightly different model with resource-consumer dynamics and interference competition. However, their proof is only for models without individual death reactions ($\delta = 0$), assumes bounded non-mass-action growth dynamics, and their analysis only guarantees majority consensus with probability $1 - O(1/\sqrt{n})$ and not with true high probability (i.e., probability $1 - 1/n^c$ for any constant $c > 0$).

Our upper bound holds for any constant $\alpha > 0$ and $\beta, \delta \geq 0$, i.e., our model also allows for individual death reactions. Moreover, we can show that majority consensus is reached with high probability provided that the gap is $\Omega(\sqrt{n \log n})$. Finally, proof technique can also be applied to the model of Andaur et al. [6] in a straightforward manner, yielding a stronger probability guarantee for majority consensus also in their model with bounded, non-mass-action growth rates.

For both self-destructive and non-destructive interspecific competition, we show that consensus is reached within $O(n)$ events both in expectation and with high probability in the absence of intraspecific competition.

Intraspecific competition. Finally, unlike prior work, we also investigate majority consensus in systems with *intraspecific interference competition*, i.e., competition between the individuals of the same species. We show that systems with intraspecific competition can have fundamentally different behavior in terms of majority consensus. Namely, we show that such majority consensus thresholds do not always exist. The results are given in Appendix 8.

First, if intraspecific and interspecific competition are equally strong ($\alpha = \gamma$) for self-destructive competition, then the probability of majority consensus is equal to the initial proportion of the majority species. This implies that majority consensus threshold is at $n - 1$, and that such systems cannot solve majority consensus with high probability (in the true sense). A similar result holds for non-self-destructive competition with $\gamma = 2\alpha$. Second, we also show that systems with only intraspecific interference competition have no majority consensus thresholds: given any gap, the chain fails to reach majority consensus with at least positive constant probability.

1.5 Technical challenges

For the stochastic LV models without intraspecific competition, our analysis uses a new technical approach for bounding the noise arising from individual reactions and asymmetric outcomes of non-self-destructive competition. We introduce a new “asynchronous, pseudo-coupling” technique that can be used to bound the behaviour of the two-species process before it reaches consensus using what we call “nice” single-species birth-death chains.

On a high-level, this approach is similar in spirit to the recent coupling techniques of Andaur et al. [6] and Cho et al. [21], which couple a two-species chain with easier to analyze single-species birth-death chains. However, there are key differences, and our approach is more general.

Limitations of prior coupling approaches. The existing techniques for microbial majority consensus [6, 21] are closely tailored to specific model variants they consider. When trying to use

these techniques to the LV models, both of the previously existing coupling techniques critically break when the process is allowed to have certain stochastic events that may occasionally benefit the minority species (e.g., individual death reactions of the majority species decreasing the discrepancy).

Cho et al. [21] use elaborate coupling arguments between several different chains to first establish a bound on the extinction time, and then couple the two-species chain to parallel birth-only Yule processes to bound the probability of reaching majority consensus using the regularized incomplete beta function. In addition to the actual coupling argument, the proof critically assumes that no individual deaths can occur and that competition is self-destructive. Moreover, the bound obtained with this technique is exponentially far from the real bound, as we show. The submartingale argument of Andaur et al. [6] allows for non-self-destructive competition at the cost of forbidding individual death reactions and losing the guarantee of true with high probability majority consensus.

Overview of new techniques. We resolve the limitations of prior techniques by moving from usual Markovian couplings that update the single-species chain and the two-species chain simultaneously to a new type of an “asynchronous pseudo-coupling”. That is, our construction is not strictly speaking a coupling between the two-species and single-species processes, as the coupled chains are not updated in lock-step. Despite this complication, we can still carefully extract useful information about the distribution of events in the two-species chain using stopping time arguments.

This is achieved in part by also deriving a much more detailed accounting of the (demographic) noise arising from birth reactions of the minority species and death reactions of the majority species. We abstract the properties of the single-species chain required by the pseudo-coupling, and providing a more fine-grained analysis of the distribution of events in the single-species chain.

Furthermore, our technique makes far fewer assumptions about the structure of the two-species system and properties of the dominating chains. Therefore, we suspect that our approach can be further extended to analyze more realistic models beyond Lotka–Volterra dynamics such as models incorporating general non-bounded resource-consumer dynamics (i.e., exploitative competition) in addition to interference competition. Indeed, our new techniques give both a simpler analysis and stronger guarantees of the previously studied microbial majority consensus dynamics [6, 21].

Our approach in a nutshell. Our main conceptual approach is to analyze the probability of reaching majority consensus $\rho(\mathbf{S})$ by considering the following random-length sum defined by

$$F(\mathbf{S}) = \sum_{t=1}^{T(\mathbf{S})} F_t, \quad (3)$$

where $F_t = \Delta_{t-1} - \Delta_t$. Note that $F_t > 0$ if the discrepancy changed in favor of the initial minority species and $F_t < 0$ if the discrepancy changed in favor of the initial majority species at time step t .

The random variable $F(\mathbf{S})$ counts how much the initial gap changes in favor of the initial *minority* species. This captures the effects of *demographic noise* on the difference between the initial majority and minority species before consensus is reached. Provided that $T(\mathbf{S})$ is finite with probability 1, we have that $\rho(\mathbf{S}) = \Pr[F < \Delta_0] = 1 - \Pr[F \geq \Delta_0]$, that is, we can connect the probability of reaching majority consensus with an given initial gap to the cumulative density function of F . Since the random variables $T(\mathbf{S})$ and $F_1, \dots, F_{T(\mathbf{S})}$ are dependent, getting a handle on the distribution of F can be non-trivial. Moreover, in general $(F_t)_{t \geq 0}$ does not give a (sub)martingale, so it is not clear if one can easily employ readily available martingale concentration bounds to analyze the process.

To bound F , we observe that the demographic noise $F = F_{\text{ind}} + F_{\text{comp}}$ can be divided to two components, where F_{ind} is the noise arising out of reproductive dynamics (i.e., individual birth and

death reactions) and F_{comp} is the noise arising from competition dynamics. This allows us to obtain a much more refined bound on total noise in the system by studying both components separately, which previous techniques were unable to do.

Intuitively, under self-destructive competition, there is only noise from the individual events, which are fairly rare under the mass action dynamics. This noise turns out to be “polylogarithmic”. On the other hand, under non-self-destructive competition, there is additional noise coming from the chance outcomes of the competition events. We show there are $O(n)$ competition events before extinction. The outcomes of the competition events are (intuitively) similar to a random walk on the line. The key challenge is that the events causing different types of noise are interleaved.

1.6 Open problems and future directions

We suspect that in general the bound of $O(\log^2 n)$ for self-destructive competition is not tight for all parameter ranges (e.g., it clearly is not tight in the case $\beta = \delta = 0$). The natural step is to identify the tight asymptotic majority consensus threshold in the case $\alpha, \beta, \delta > 0$. Second, while we focus on Lotka–Volterra models, we believe our techniques are applicable to a wider variety of stochastic population models beyond the competitive, two-species Lotka–Volterra model.

We show that in the worst-case, intraspecific competition can badly hinder the probability of reaching majority consensus. For example, with self-destructive competition, we identify that the majority consensus threshold is $O(\log^2 n)$ with $\alpha > 0$ and $\gamma = 0$ and $n - 1$ when $\alpha = \gamma > 0$. An interesting open problem is to identify at which point does the majority consensus threshold enter a sublinear or polylogarithmic regime when $\alpha > 0$ is a fixed constant and $\gamma \rightarrow 0$.

On a conceptual level, our results suggest interesting *computational trade-offs* in the design of microbial circuits: majority consensus protocols utilizing self-destructive competition seem to be less sensitive to demographic noise than protocols based on non-self-destructive competition. However, the trade-off is that the former is much more costly at the individual level. From a bioengineering perspective, it would be interesting to investigate if circuits utilizing non-self-destructive interference are evolutionarily more stable than circuits utilizing self-destructive competition.

Finally, one may surmise that the computational trade-offs implied by this work are solely theoretical. For example, idealized well-mixed, mass action Lotka–Volterra models do not capture the full range of microbial dynamics. However, such models have been experimentally observed to provide reasonable approximations in many situations [26, 42, 47] also in the context of synthetic microbial consortia [5, 54]. Regardless, future work should further investigate how relaxing the model assumptions influence the predicted computational trade-offs and experimentally test this. This requires developing new proof techniques for dealing with non-mass action models and/or explicit spatial dynamics.

1.7 Limitations and biological assumptions

This paper focuses on new mathematical techniques for the analysis of distributed consensus dynamics in stochastic interaction models. Like in all biological models, there are several biological assumptions and restrictions made in the stochastic Lotka–Volterra models we study. Indeed, our work shows how subtle differences in model assumptions can fundamentally impact the solvability and complexity of majority consensus in microbial population models.

Markovian, mass action dynamics. As typical for analytical models, we assume memoryless species and mass action dynamics. We do not model the internal biophysical or metabolic state of the individual bacteria. In reality, the reproductive dynamics of microbes can depend on their

internal state and changing environmental conditions [42]. While such models may very well be used in numerical simulations, their analysis seems to be far beyond the reach of current proof techniques. Regardless, our technique can still be used to analyze also models with non-mass-action reactions, such as those in the biological reaction network model of Andaur et al. [6].

Physical aspects and asymptotics. We assume unit volume. This differs from how molecular systems are analyzed at the so-called thermodynamic limit, where the ratio of volume and initial population size remains constant as $n \rightarrow \infty$. We do not assume such a constant initial population density.

While the limit of an infinitely dense cell population is clearly non-physical, the asymptotic analysis here illuminates the impact of different biological mechanisms on the performance of microbial protocols with increasing initial cell populations (i.e., larger inputs). In a wetlab setting, the initial population size can typically vary over a large range, from 1 mL^{-1} to about 10^9 mL^{-1} for *E. coli*. Nevertheless, it remains open to demonstrate that within this range, asymptotic effects are dominating, as suggested by numerical simulations done in prior work (see, e.g., [21]).

Finally, it is important to note that while our LV models allow for increasing initial density and there is no explicitly assumption bounding the maximum size of a population, the population size is regulated by a carrying capacity induced by density-dependent competition. In particular, when including also intraspecific competition ($\gamma > 0$), the stochastic LV models exhibits the full logistic growth regime usually observed for microbial populations even after competitive exclusion.

1.8 Structure of the paper

We have delegated some of the proofs to the Appendix. In addition, Section 2 further discusses the background on previous theoretical and empirical work regarding Lotka–Volterra models and related work on majority consensus in other models of computing. We start with preliminaries in Section 3. In Section 4, we establish some useful results regarding the behavior of certain single-species chains, which we then use to bound the behavior of competitive two-species Lotka–Volterra chains in Section 5. Section 6 and Section 7 give bounds for majority consensus thresholds under self-destructive and non-self-destructive interference competition. The lower bounds for majority consensus thresholds in systems with intraspecific competition appear in Section 8.

2 Related work

2.1 Lotka–Volterra models

Named after the pioneering work of Lotka [55] and Volterra [71], Lotka–Volterra (LV) models have been the cornerstone of theoretical biology and microbial ecology for over one hundred years. Gause [37] reported the first experimental validations of the two-species Lotka–Volterra equation using yeast and protozoa microcosms in 1934. Since then, LV models have become widely applied in microbial ecology [42]. They have been found to predict well even many aspects of complex microbial communities [26, 47, 69]. Given the abundance of work on LV models, we only give a brief overview, focusing on competitive LV models in well-mixed populations.

In biology, the competitive Lotka–Volterra ordinary differential equation model has become the baseline model for ecological community dynamics [46, 59, 62]. The simplicity of numerical simulation of ODE models has rendered them a popular choice for modelling synthetic microbial consortia dynamics [54, 57, 74]. These deterministic models are derived under the assumption of (infinitely) large continuous populations, ignoring the important stochastic effects driving the dynamics of real

(finite and discrete) populations [15, 52, 73]. Despite their shortcomings, the key reason for the popularity of the deterministic models is simple: even complex ODE models are easy to simulate numerically, while even simple stochastic models tend to be difficult to analyse [15, 38, 49].

This challenge has motivated a significant line of research devoted to the analysis of stochastic, discrete Lotka–Volterra models in well-mixed populations, typically using the tools of statistical mechanics [28, 48, 60, 67]. These models closely resemble to the models we use, and many papers in this area also investigate coexistence and extinction probabilities (as a function of time). However, they do not investigate the probability of *majority consensus* as a function of n and Δ , nor compare how different choices of competitive mechanisms impact these probabilities. Furthermore, these models usually assume only non-self-destructive competition and finite maximum population size. This subtly differs from our model, where the carrying capacity arises from interference competition. In the theoretical computer science literature, convergence and thresholds of discrete LV dynamics have been studied in the population protocol model [24] in bounded populations.

Comparison with deterministic kinetics. We note that for the two stochastic Lotka–Volterra models we study, the continuous-valued approximations under deterministic mass action kinetics correspond to the usual deterministic competitive Lotka–Volterra equations [42, 46]. For two-species in the neutral case, this is given by the non-linear differential equations

$$\frac{dx_i}{dt} = x_i(r - \alpha'x_{1-i} - \gamma'x_i), \quad (4)$$

where x_i is the density of species $i \in \{0, 1\}$, $r = \beta - \delta$ is the intrinsic growth rate, $\gamma' = \gamma_0 = \gamma_1 \geq 0$ is the rate of intraspecific competition, and $\alpha' \geq 0$ is the rate of interspecific competition so that for model (1), we have $\alpha' = \alpha = \alpha_0 + \alpha_1$, and for model (2) we have $\alpha' = \alpha_0 = \alpha_1$. It is easy to see that in this model, if $\alpha' > \gamma'$, then the species with the higher initial density will deterministically always win. Thus, this model fails to capture the stochastic effects occurring in finite populations.

2.2 Majority consensus in stochastic interaction models

The majority consensus problem has been studied in many asynchronous, stochastic interaction models in fully-connected networks (i.e., well-mixed systems). The problem and its extension, plurality consensus, have also been studied in synchronous, gossip models with static population size [9, 11, 23, 33, 39]. However, here we primarily focus on stochastic asynchronous models that most closely resemble our setting.

Majority consensus in stochastic biological population models. The closest to our work is the recent work by Cho et al. [21] who considered a special case of our discrete, stochastic Lotka–Volterra model in a two-species chemical reaction network model. They showed that an initial gap of $\Omega(\sqrt{n \log n})$ is sufficient under the assumption $\delta = 0$ and self-destructive interspecific interference competition. In contrast to their results, our results hold also for systems with $\delta > 0$ and show that an exponentially smaller gap of only $\Omega(\log^2 n)$ suffices for any constant rate constants $\alpha_0, \alpha_1, \beta, \delta > 0$.

More recently, Andaur et al. [6] considered a resource-consumer model of biological population dynamics with non-self-destructive interference competition. Their model is not strictly speaking a chemical reaction network model, as they allow for non-mass-action reactions. They showed that in their model $\Omega(\sqrt{n \log n})$ gap suffices with probability $1 - O(1/\sqrt{n})$. However, their model assumes bounded, non-mass-action growth and no individual death reactions (i.e., $\delta = 0$). With some work, our new technique can also applied to their model to obtain a bound of $\Omega(\sqrt{n \log n})$ with high

probability, i.e., with probability at least $1 - 1/n^k$, for any constant $k > 0$. This is because their dominating chain is also a “nice chain” in the sense we define in Section 4.

Chemical reaction networks. Condon et al. [22] consider the majority consensus problem and multi-valued consensus in general chemical reaction systems that also allow *trimolecular* reactions (i.e., a single reaction can involve up to three reactants and products). They gave bi- and trimolecular reaction networks for which an initial gap of $\Omega(\sqrt{n \log n})$ is sufficient with high probability.

While this work operates in the same formal model, the protocols are not directly comparable with protocols in the two-species Lotka–Volterra model. However, interestingly, their “heavy-B” protocol, in which reactions have two reactants and three products, and bimolecular “double-B” protocol with two reactants and two products for each rule, resemble our protocols with self-destructive competition. On the other hand, the rules in the protocol “single-B” resemble non-self-destructive competition. However, these protocols employ three species. They also give a two-species trimolecular protocol that uses three reactants and products.

Population protocols. The population protocol model [7] has become a popular model to study computation in well-mixed chemical solutions. The model is a special case of the chemical reaction network model, where each rule has exactly two reactants and products, and each reaction has unit rate. In particular, this implies that the total population size n remains static throughout the execution of the protocol.

For majority consensus, Angluin et al. [8] considered approximate majority agreement in the population protocol model. They give a simple 3-state algorithm that succeeds with high probability when the initial gap is $\Omega(\sqrt{n} \cdot \log n)$ and the number of steps to reach consensus is $O(n \log n)$ with high probability. The same cancellation-idea used by this protocol also appears in protocols in a variety of other models, including our Lotka–Volterra protocols and other protocols relying on competition [6, 21], the “single-B” protocol of Condon et al. [22] and the ternary signalling protocol of Perron et al. [64]. Perron et al. [64] analysed such a cancellation protocol and showed that if the gap is linear gap the protocol succeeds with very high probability, i.e., fails to converge to the initial majority value with exponentially small probability.

Draief and Vojnović [31] showed that there exists a 4-state protocol that always reaches majority consensus with *any* positive gap in $O(n^2)$ expected interactions. The same protocol was also described and analysed by Mertzios et al. [61]. This variant of majority consensus, where the protocols are required to succeed with probability 1, is often referred to *exact majority* in the population protocol literature. Alistarh et al. [4] showed that using $O(\log^3 n)$ states, exact majority can be solved in only n polylog n interactions in expectation and with high probability.

Subsequently, Alistarh et al. [1] showed that any $o(\log \log n)$ -state protocol for exact majority that succeeds requires $n^2 / \text{polylog } n$ interactions. They also showed that using $O(\log^2 n)$ states majority can be solved in $O(n \log^2 n)$ interactions. The upper bound result was subsequently improved to $O(\log^2 n)$ states and the lower bound strengthened to $\Omega(\log n)$ for a large class of protocols by Alistarh et al. [2]. Recently, Doty et al. [30] gave a $O(\log n)$ -state protocol that stabilises in $\Theta(n \log n)$ expected interactions, matching the lower bounds.

Czyżowicz et al. [24] considered population protocols with discrete Lotka–Volterra-like stochastic dynamics. Their Lotka–Volterra dynamics are different, as they operate in the population protocol model, where the total population size remains static. In particular, there are no individual birth and death reactions. In this setting, they gave a 4-state protocol that solves majority consensus with high probability provided that the ratio of the initial input counts a and b is $a/b = 1 + \varepsilon/(1 - \varepsilon)$ for a constant $\varepsilon > 0$, that is, the initial gap is linear. For their analysis, they also use a coupling

technique with delays, but their technique is different, as they operate in the population protocol model with a static maximum size for the system.

While the population protocol model is in spirit similar to the stochastic biological population models we consider, the full range of algorithmic techniques available in the population protocol model are not easily (or if at all) realizable in synthetic biology setting. For example, one cannot completely enforce which reactions are active (e.g., by assuming that no birth/death events occur or that they occur only during certain phases of the protocol), a technique used by cancellation-doubling protocols for fast exact majority protocols. In particular a key challenge in the synthetic biology setting is that reproductive and ecological dynamics are interleaved with the engineered biological circuits, which in combination give rise to the microbial protocol [5, 54].

Finally, we note that Goldwasser et al. [41] studied the impact of changing population composition in a synchronous variant of the population protocol model, where an adversary is allowed to make a bounded number of *arbitrary* insertions and deletions of individuals per round. They showed that even under such adversarial changes to population structure, a certain population balancing problem can be solved efficiently with only small number of states. In their setting, the individuals can also trigger duplication (birth) and self-destruction (death) events. In contrast, in our setting, the birth and death events are dictated by the underlying stochastic Lotka–Volterra dynamics.

3 Preliminaries

We use $\mathbb{N} = \{0, 1, \dots\}$ to denote the set of nonnegative integers. For any $n > 0$, we use $\log n$ for the base-2 logarithm of n and $\ln n$ for the natural logarithm of n . For any integer $n > 0$, we write H_n to denote the n^{th} Harmonic number given by the sum $\sum_{i=1}^n 1/i$, which is lower-bounded by $\ln n$.

Concentration bounds. We say that a random variable X is a *Bernoulli* random variable if X only takes values in $\{0, 1\}$. A random variable X is said to be $O(f(n))$ *with high probability* if for any fixed constant $k \geq 0$ there exists a constant $C(k)$ such that $\Pr[X \geq C(k)f(n)] \leq 1/n^k$. To establish with high probability bounds, we use two standard results on the concentration of random sums.

Lemma 1 (Chernoff bounds). *Let $X = X_1 + \dots + X_n$ be the sum of n independent Bernoulli random variables. Then*

- (1) $\Pr[X \geq (1 + \varepsilon) \cdot \mathbf{E}[X]] \leq \exp(-\mathbf{E}[X] \cdot \varepsilon^2 / (2 + \varepsilon))$ for any $\varepsilon > 0$, and
- (2) $\Pr[X \leq (1 - \varepsilon) \cdot \mathbf{E}[X]] \leq \exp(-\mathbf{E}[X] \cdot \varepsilon^2 / 2)$ for any $0 < \varepsilon < 1$.

We will also make use of a special case of Hoeffding’s inequality for random variables restricted to the range $[-1, 1]$.

Lemma 2 (Hoeffding’s inequality). *Let $X = X_1 + \dots + X_n$ be the sum of n independent random variables, where $X_i \in [-1, 1]$. Then for any $t \geq 0$*

$$\Pr[|X - \mathbf{E}[X]| \geq t] \leq 2 \cdot \exp\left(-\frac{2t^2}{n}\right).$$

For our lower bounds, we also make use of the following *anti-concentration bound* implied by the Central Limit Theorem.

Lemma 3. *Let $\varepsilon \in (0, 1)$ be a constant and X be the sum of n i.i.d. random variables $X_1, \dots, X_n$ with mean 0 and variance 1. Then there is a constant $\theta = \theta(\varepsilon) > 0$ such that for any sufficiently large n , we have $\Pr[X > \theta\sqrt{n}] \geq \varepsilon$.*

Stochastic domination and couplings. Let X_1 and X_2 be two non-negative random variables. If $\Pr[X_2 \geq x] \geq \Pr[X_1 \geq x]$ for all $x \geq 0$, then we write $X_1 \preceq X_2$ and say that X_2 *stochastically dominates* X_1 . The random variable $(\widehat{X}_1, \widehat{X}_2)$ is said to be a *coupling* of X_1 and X_2 if the distribution of $\widehat{X}_i$ is the same as the distribution X_i , that is, $\Pr[X_i \geq x] = \Pr[\widehat{X}_i \geq x]$ for any $x \geq 0$. We make use of the following simple lemma.

Lemma 4. *Let $X = X_1 + \dots + X_n$ be a sum of (not necessarily independent) Bernoulli random variables and $Y = Y_1 + \dots + Y_n$ be a sum of independent Bernoulli random variables.*

- (1) *If $\Pr[X_i = 1 \mid X_1, \dots, X_{i-1}] \leq \Pr[Y_i = 1]$ for each $1 \leq i \leq n$, then $X \preceq Y$.*
- (2) *If $\Pr[X_i = 1 \mid X_1, \dots, X_{i-1}] \geq \Pr[Y_i = 1]$ for each $1 \leq i \leq n$, then $Y \preceq X$.*

The proofs of Lemma 3 and Lemma 4 are given in Appendix 9.

4 Bounds for nice single-species chains

Let $p, q: \mathbb{N} \rightarrow [0, 1]$ such that $p(n) + q(n) \leq 1$ for all $n \geq 0$. The birth-death chain defined by p and q is the discrete-time Markov chain $N = (N_t)_{t \in \mathbb{N}}$ on the state space $\mathbb{N}$, where in each step the chain goes from state n to $n + 1$ with *birth probability* $p(n)$, to state $n - 1$ with *death probability* $q(n)$. The probability $h(n) = 1 - p(n) - q(n)$ is the *holding probability* of the chain in state n . A state x is *absorbing* if $p(x) = q(x) = 0$.

We assume that $p(n) > 0$ and $q(n) > 0$ for all $n > 0$ and $p(0) = q(0) = 0$ so that 0 is the unique absorbing state. The *absorption time* or *extinction time* of a chain N is $E(N) = \min\{t : N_t = 0\}$, that is, the minimum time until the chain reaches the unique absorbing state. We say the chain is *nice* if there exist constants $C, D > 0$ such that $p(n) \leq C/n$ and $q(n) \geq D$ for all $n > 0$.

Bounds on number of births for nice chains. For any nice chain started, we will bound the extinction time $E(N)$ of a chain and the number $B(N)$ of birth events that occur before extinction. We assume throughout that $N_0 = n$, and with a slight abuse of notation, we write $E(N) = E(n)$ and $B(N) = B(n)$ for such a chain N . We first asymptotically bound the extinction time. The lower bound follows immediately from the fact that the chain needs to decrement at least n times to reach state 0. For the upper bound, the result follows from a known result for the absorption time of discrete-time birth-death processes; see [68, Theorem 3.1] and [6, Lemma 3].

Lemma 5. *For any nice chain started in state n , its expected extinction time is $\mathbf{E}[E(n)] = \Theta(n)$.*

Equipped with the above lemma, we can bound the number of birth events before extinction.

Lemma 6. *For any nice chain, the expected number of births satisfies $\mathbf{E}[B(n)] \in O(\log n)$.*

Proof. By Lemma 5 and the law of total expectation, for any non-negative integer-valued random variable X we have the bound

$$\mathbf{E}[E(X)] = \sum_{k=0}^{\infty} \Pr[X = k] \cdot \mathbf{E}[E(k)] \leq C' \sum_{k=0}^{\infty} \Pr[X = k] \cdot k = C' \mathbf{E}[X],$$

where C' is the constant of the upper bound from Lemma 5.

We say that a time step t is a *holding step* if the chain stays in the same state (i.e., neither a birth or death happen). Let $R = R(n)$ be the minimum time until exactly n non-holding steps have

occurred after starting the chain in state n . Clearly, $n \leq R(n) \leq E(n)$, since the chain needs to decrement at least n times before going extinct. Moreover, $R(n)$ is finite with probability 1, as $E(n)$ is finite with probability 1. Let B_t denote the number of birth events and D_t denote the number of death events that have occurred by step $t \geq 0$. Observe that $D_R + B_R = n$, as at time step R , there have been exactly n non-holding steps. Thus, at time R the chain is in state

$$N_R = n - D_R + B_R = n - (n - B_R) + B_R = 2B_R.$$

Moreover, note that

$$B(n) \leq B_R + E(N_R) = B_R + E(2B_R),$$

since the number of birth events after step R is upper bounded by the extinction time of the chain started in state N_R . Thus, by linearity of expectation and the above, we get that

$$\mathbf{E}[B(n)] \leq \mathbf{E}[B_R] + \mathbf{E}[E(2B_R)] = (2C' + 1) \cdot \mathbf{E}[B_R].$$

We next show that $\mathbf{E}[B_R] \in O(\log n)$, which implies the claim of the lemma.

Let us consider the value of B_R . Since $R \geq n$ is the minimum time until exactly n non-holding steps have happened, we have

$$B_R = \sum_{i=1}^n X_i,$$

where X_i is the indicator variable whether the i th non-holding step is a birth event or not. Note that $\mathbf{E}[X_i] \leq C/(n - i + 1)$, as the chain is in state at least $n - i + 1$ after i non-holding steps. Therefore, the expectation of B_R satisfies

$$\mathbf{E}[B_R] = \mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^n X_i\right] \leq C \sum_{i=1}^n \frac{1}{n - i + 1} = C \sum_{i=1}^n \frac{1}{i} = CH_n,$$

where H_n is the n th Harmonic number. This proves the claim. $\square$

With the bound on expected number of births, some calculations and the application of Markov's inequality and Chernoff bounds yield the following claims for all sufficiently large n .

Lemma 7. *Let $k > 0$. For any nice chain, there exists a $C(k)$ such that $\Pr[B(n) \geq C(k) \cdot \log^2 n] \leq 1/n^k$.*

Proof. Let $c > 0$ such that $\mathbf{E}[B(n)] \leq c \log n$ for any $n > 0$; by Lemma 6 such a constant exists. Set $L = \lceil 2c \log n \rceil$, $K = \lceil k \log n \rceil$, and $M = n + LK$. We will assume that n is sufficiently large so that $M \leq n^2$ holds. We will make K trials, where in each trial $1 \leq i \leq K$, we run the chain until either L birth events occur or the chain hits state 0. After the i th trial, the chain is in some state $n_i \leq n + iL \leq n + kL \leq M \leq n^2$. We say that the i th trial is successful if the chain hits the state 0, and otherwise, it fails. Note that $B(n_i) \leq B(n^2)$ for each $1 \leq i \leq K$ and that $\mathbf{E}[B(n^2)] \leq c \log n^2 = 2c \log n \leq L$. Since the chain is Markovian and by Markov's inequality, the probability that the i th experiment fails is at most

$$\Pr[B(n_k) \geq L] \leq \Pr[B(n^2) \geq L] \leq \Pr[B(n^2) \geq 2 \cdot \mathbf{E}[B(n^2)]] \leq \frac{1}{2}.$$

The probability that all K experiments fail is thus at most $1/2^K \leq 1/n^k$. After the K experiments, there have been at most $KL \in O(\log^2 n)$ birth events. Hence the claim follows. $\square$

Lemma 8. *Let $k > 0$. For any nice chain, there exists a $\theta(k)$ such that $\Pr[E(n) \geq \theta(k) \cdot n] \leq 1/n^k$.*

Proof. Since the chain is nice, there exists a constant $D \in (0, 1)$ such that $q(m) \geq D$ for all $m > 0$. Set $R = \lceil 6n/D \rceil$. We first bound the number X of holding steps in non-absorbing states (i.e., states other than 0) that occur during the first R steps. Let X_t be the indicator variable denoting whether a holding step in a non-absorbing state occurs at step $t \geq 1$. Clearly, $N_t = 0 \implies X_t = 0$. For any $n > 0$, the probability of such a holding step in state n is at most $1 - p(n) - q(n) \leq 1 - D$, i.e., $\Pr[X_t = 1 \mid X_1, \dots, X_{t-1}] \leq 1 - D$. Let $Y_1, \dots, Y_R$ be a sequence of independent Bernoulli random variables with $\Pr[Y_t = 1] = 1 - D$. By Lemma 4, we have that

$$X = \sum_{t=1}^R X_t \preceq \sum_{t=1}^R Y_t = Y.$$

Note that $\mathbf{E}[Y] = (1 - D)R \in \Theta(n)$ and set $\varepsilon = \sqrt{3(k+1) \ln n / \mathbf{E}[Y]}$. Since $\mathbf{E}[Y] \in \Theta(n)$, we have $\varepsilon \in o(1)$. Applying the Chernoff bound from Lemma 1 with ε yields

$$\Pr[X \geq (1 + \varepsilon) \mathbf{E}[Y]] \leq \Pr[Y \geq (1 + \varepsilon) \mathbf{E}[Y]] \leq \exp\left(-\frac{\varepsilon^2}{3} \cdot \mathbf{E}[Y]\right) \leq 1/n^{k+1}.$$

Moreover, by Lemma 7 we get that with probability at most $1/n^{k+1}$ there are more than $C(k+1) \log^2 n$ birth events before the chain goes extinct. Thus for sufficiently large n , with probability $1 - 2/n^k$, by time R there are at most

$$K = (1 + \varepsilon) \mathbf{E}[Y] + C(k) \log^2 n$$

birth events or holding steps in non-absorbing states. For large enough n , we have that $\varepsilon < D/2$ and $C(k) \log^2 n \leq n$. Now

$$\begin{aligned} R - K &= R - (1 + \varepsilon) \mathbf{E}[Y] - C(k) \log^2 n \\ &\geq R - \left(1 + \frac{D}{2}\right) (1 - D) R - n \geq R - \left(1 - \frac{D}{2} - \frac{D^2}{2}\right) R - n \\ &= R \left(\frac{D}{2} + \frac{D^2}{2}\right) - n > \frac{DR}{2} - n \geq \frac{6n}{2} - n = 2n. \end{aligned}$$

Therefore, out of the R steps, there are at least $R - K > 2n$ steps that are either holding steps in state 0 or death events. As the chain never reaches a state higher than $n + O(\log^2 n)$ with high probability, we get that the chain hits the state 0 in R steps with probability at least $1 - 2/n^k$ for all sufficiently large n , as $n + C(k+1) \log^2 n < 2n$. $\square$

5 Dominating chains for Lotka–Volterra systems

We now introduce a “pseudo-coupling” of single-species birth death chains and two-species Lotka–Volterra chains that can be used to *over-approximate* the consensus time and the number of “bad” individual events that can decrease the gap. Unlike previous “dominating chain” approaches [6, 21], which consider restricted choices of rate parameters, our approach works with any choice of $\beta, \delta, \alpha_0, \alpha_1 > 0$. In particular, we allow individual death reactions with $\delta > 0$ and do not assume symmetric interference competition, i.e., $\alpha_0 \neq \alpha_1$ is possible. The condition for the dominating chain is also simpler; this comes with the cost that we do not get a coupling in the strict sense.

5.1 The chain domination lemma

Let $\mathbf{S} = (\mathbf{S}_t)_{t \geq 0}$ be a two-species chain. We say that an event at time step t is *bad* if it decreases the gap between the minimum and maximum species, i.e., if $\Delta_{t+1} = \Delta_t - 1$ holds *conditioned on* that the minimum species has positive count. Let $P(a, b)$ be the probability that the chain $\mathbf{S}$ in state (a, b) has a bad *non-competitive* reaction. We say that an event at time step t is *good* if the species with the smaller count decreases in count. Let $Q(a, b)$ be the probability that the chain $\mathbf{S}$ in state (a, b) has a *good* reaction. Note that the probability of bad competitive events is $1 - P(a, b) - Q(a, b)$.

Let $N = (N_t)_{t \geq 0}$ be a single-species birth-death chain defined by birth function p and death function q . We say that N is a *dominating chain* for $\mathbf{S}$ if for any $a, b \geq 0$ we have

- (D1) $P(a, b) \leq p(\min\{a, b\})$, and
- (D2) $Q(a, b) \geq q(\min\{a, b\})$.

Lemma 9 (Chain domination lemma). *Suppose the single-species chain $N = (N_t)_{t \geq 0}$ is a dominating chain for the two-species chain $\mathbf{S} = (\mathbf{S}_t)_{t \geq 0}$. If $N_0 \geq \min \mathbf{S}_0$, then*

- (a) $T(\mathbf{S}) \preceq E(N)$, and
- (b) $J(\mathbf{S}) \preceq B(N)$,

where $T(\mathbf{S})$ is the consensus time of $\mathbf{S}$, $J(\mathbf{S})$ the number of bad non-competitive reactions in the chain $\mathbf{S}$, $E(N)$ is the extinction time of N , and $B(N)$ the total number of births in the chain N .

To prove the lemma, we construct a Markov chain $(\hat{\mathbf{S}}, \hat{N})$ on the state space $\mathbb{N}^2 \times \mathbb{N}$. Strictly speaking, this will *not* be a coupling of $\mathbf{S}$ and N , as only the marginal distribution of $\hat{N}$ will equal the distribution of N . However, for each $t \geq 0$, we show how to extract random variables from $\hat{\mathbf{S}}$ whose marginal distribution equals that of $\mathbf{S}_t$ almost surely (i.e., with probability 1) for every $t \geq 0$.

The pseudo-coupling. We construct the Markov chain $(\hat{\mathbf{S}}, \hat{N})$ as follows. We set $\hat{\mathbf{S}}_0 = \mathbf{S}_0$ and $\hat{N}_0 = N_0 \geq \min \mathbf{S}_0$. Let $(\xi_t)_{t \geq 0}$ be a sequence of i.i.d. random variables distributed uniformly at random in the unit interval $[0, 1)$. We determine the state of the chain for step $t + 1$ inductively:

- (1) Let $\hat{N}_t = m$. We set $\hat{N}_{t+1}$ as follows:
 - (a) If $\xi_t \in [0, p(m))$, then set $\hat{N}_{t+1} = \hat{N}_t + 1 = m + 1$.
 - (b) If $\xi_t \in [1 - q(m), 1)$, then set $\hat{N}_{t+1} = \hat{N}_t - 1 = m - 1$.
 - (c) Otherwise, set $\hat{N}_{t+1} = \hat{N}_t$. (A holding step occurs.)
- (2) Let $\mathbf{S}_t = (a, b)$. If $\min \hat{\mathbf{S}}_t \neq \hat{N}_t$, we set $\hat{\mathbf{S}}_{t+1} = \hat{\mathbf{S}}_t$. Otherwise, we set $\hat{\mathbf{S}}_{t+1}$ as follows:
 - (a) If $\xi_t \in [0, P(a, b))$, then sample $\hat{\mathbf{S}}_{t+1}$ conditioned on the event that $\hat{\mathbf{S}}_t = (a, b)$ and the t^{th} event is a bad non-competitive event.
 - (b) If $\xi_t \in [1 - Q(a, b), 1)$, then sample $\mathbf{S}_{t+1}$ conditioned on the event that $\mathbf{S}_t = (a, b)$ and that the t^{th} event is a good competitive interaction.
 - (c) Otherwise, if $\xi_t \in [P(a, b), 1 - Q(a, b))$, then sample $\mathbf{S}_{t+1}$ conditioned on the event that $\mathbf{S}_t = (a, b)$ and that the t^{th} event is not a good competitive interaction or a bad non-competitive event.

One easily checks that, by construction, the marginal distribution of $\widehat{N}_t$ is equal to the distribution N_t . Moreover, $\widehat{\mathbf{S}}_{t+1} \neq \widehat{\mathbf{S}}_t$ can hold only for steps, where $\widehat{\mathbf{S}}_t = \widehat{N}_t$. So $\widehat{\mathbf{S}}_t$ does not necessarily have the same marginal distribution as $\mathbf{S}_t$. For any $t \geq 0$, define $J_t(\widehat{\mathbf{S}})$ to be the number of bad non-competitive events that have occurred in $\widehat{\mathbf{S}}$ by time t and $B_t(\widehat{N})$ to be the number of birth events that have occurred in $\widehat{N}$ by time t .

We say that an event occurs *almost surely* if it happens with probability 1. For example, we say that N goes extinct almost surely if the extinction time of N is finite with probability 1.

Lemma 10. *If $\min \widehat{\mathbf{S}}_0 = \widehat{N}_0$, then $\min \widehat{\mathbf{S}}_t \leq \widehat{N}_t$ and $J_t(\widehat{\mathbf{S}}) \leq B_t(\widehat{N})$ almost surely for all $t \geq 0$.*

Proof. We proceed by induction on $t \geq 0$. The base case $t = 0$ is trivial. For the induction step, suppose $\min \widehat{\mathbf{S}}_t \leq \widehat{N}_t$ and $J_t(\widehat{\mathbf{S}}) \leq B_t(\widehat{N})$ almost surely. By the law of total probability, we have

$$\begin{aligned} \Pr[\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1}] = \\ \Pr[\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1} \mid \min \widehat{\mathbf{S}}_t = \widehat{N}_t] \cdot \Pr[\min \widehat{\mathbf{S}}_t = \widehat{N}_t] + \\ \Pr[\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1} \mid \min \widehat{\mathbf{S}}_t < \widehat{N}_t] \cdot \Pr[\min \widehat{\mathbf{S}}_t < \widehat{N}_t]. \end{aligned}$$

The condition $\min \widehat{\mathbf{S}}_t < \widehat{N}_t$ implies $\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1}$ since $\widehat{\mathbf{S}}_{t+1} = \widehat{\mathbf{S}}_t$ and $\widehat{N}_{t+1} \leq \widehat{N}_t - 1$ by the transition probabilities of $(\widehat{\mathbf{S}}, \widehat{N})$, as given by rules (1) and (2). For $\min \widehat{\mathbf{S}}_t = \widehat{N}_t$, writing $\widehat{N}_t = m$ and $\widehat{\mathbf{S}}_t = (a, b)$, we distinguish three cases via the law of total probability:

- (1) $\xi_t \in [0, P(a, b))$
- (2) $\xi_t \in [1 - q(m), 1)$
- (3) $\xi_t \in [P(a, b), 1 - q(m))$.

In case (1), by property (D1) of dominating chains, we have $\xi_t \leq P(a, b) \leq p(m)$. Therefore, by rule (1a) we have $\widehat{N}_{t+1} = \widehat{N}_t + 1$ and $\min \widehat{\mathbf{S}}_{t+1} \leq \min \widehat{\mathbf{S}}_t + 1$ by rule (2a). Combining these inequalities implies $\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1}$. In case (2), since $\xi_t \geq 1 - q(m) \geq 1 - Q(a, b)$ by (D2), we have $\widehat{N}_{t+1} = \widehat{N}_t - 1$ by rule (1b) and $\min \widehat{\mathbf{S}}_{t+1} = \min \widehat{\mathbf{S}}_t - 1$ by rule (2b). In case (3), we have $\xi_t \in [P(a, b), 1 - q(m)) \subseteq [p(m), 1 - Q(a, b))$. Thus $\widehat{N}_{t+1} \geq \widehat{N}_t$ and $\min \widehat{\mathbf{S}}_{t+1} \leq \min \widehat{\mathbf{S}}_t$. Therefore, in all cases, the probability of $\min \widehat{\mathbf{S}}_{t+1} \leq \widehat{N}_{t+1}$ is equal to 1. This shows the first inequality.

For the second inequality, we proceed in a similar manner and note that a bad non-competitive event can occur in $\widehat{\mathbf{S}}$ only in the case (1). But then by rule (1a) and (2a) we get that $J_{t+1}(\widehat{\mathbf{S}}) = J_t(\widehat{\mathbf{S}}) + 1 \leq B_t(\widehat{N}) + 1 = B_{t+1}(\widehat{N})$. This completes the induction. $\square$

For every positive integer k , define the random time $\tau(k)$ as the k^{th} smallest time t such that $\min \widehat{\mathbf{S}}_t = \widehat{N}_t$. This is a stopping time of the Markov chain $(\widehat{\mathbf{S}}, \widehat{N})$ [16, Chapter 2, Definition 7.1].

Lemma 11. *Suppose $\widehat{\mathbf{S}}_0 = \mathbf{S}_0$ and $\min \widehat{\mathbf{S}}_0 = \widehat{N}_0$. If N goes extinct almost surely, then $\tau(k+1) < \infty$ almost surely and the marginal distribution of $\widehat{\mathbf{S}}_{\tau(k+1)}$ equals the distribution of $\mathbf{S}_k$ for every $k \geq 0$.*

Proof. We proceed by induction on k . For the base case, by the hypothesis on the initial states, we have $\tau(1) = 0 < \infty$ and $\widehat{\mathbf{S}}_{\tau(1)} = \widehat{\mathbf{S}}_0 = \mathbf{S}_0$. This proves the base case.

For the induction step, assume that $k \geq 1$ and that $\widehat{\mathbf{S}}_{\tau(k)}$ and $\mathbf{S}_{k-1}$ are equally distributed. By the strong Markov property [16, Chapter 2, Theorem 7.1], the process $(\widehat{\mathbf{S}}_{\tau(k)+t}, \widehat{N}_{\tau(k)+t})_{t \geq 0}$ is a Markov chain with the same transition probabilities as $(\widehat{\mathbf{S}}, \widehat{N})$. By definition of $\tau(k)$, we have $\min \widehat{\mathbf{S}}_{\tau(k)} = \widehat{N}_{\tau(k)}$. By the definition of the transition probabilities of $(\widehat{\mathbf{S}}, \widehat{N})$ for this case, the states $\widehat{\mathbf{S}}_{\tau(k)+1}$ and $\mathbf{S}_{k+1}$ are equally distributed. We have that

$$\widehat{\mathbf{S}}_{\tau(k)+t} = \widehat{\mathbf{S}}_{\tau(k)+1} \text{ for all } 1 \leq t < \tau(k+1) - \tau(k) + 1 \quad (5)$$

almost surely by the definition of $\tau(k+1)$ and the transition probabilities of $(\widehat{\mathbf{S}}, \widehat{N})$. In particular, if $\tau(k+1) < \infty$, we deduce that $\widehat{\mathbf{S}}_{\tau(k+1)}$ has the same distribution as $\widehat{\mathbf{S}}_{\tau(k)+1}$, which, in turn, has the same distribution as $\mathbf{S}_k$. It remains to prove that $\tau(k+1) < \infty$ almost surely. By Lemma 10 and (5),

$$\min \widehat{\mathbf{S}}_{\tau(k)+1} < \widehat{N}_{\tau(k)+t} \text{ for all } 1 \leq t < \tau(k+1) - \tau(k) + 1. \quad (6)$$

Since N goes extinct almost surely, we have $N_t \rightarrow 0$ almost surely. Since $\widehat{N}_t$ and N_t are equally distributed, $\lim_{t \rightarrow \infty} \widehat{N}_{\tau(k)+t} = 0$ almost surely. Thus, there exists some $t \geq 1$ such that $\min \widehat{\mathbf{S}}_{\tau(k)+1} \geq \widehat{N}_{\tau(k)+t}$ almost surely, as $\min \widehat{\mathbf{S}}_{\tau(k)+1} \geq 0$. This contradicts $\Pr[\tau(k+1) = \infty] > 0$, so this probability is hence zero and so $\Pr[\tau(k+1) < \infty] = 1$. $\square$

Proof of the chain domination lemma (Lemma 9). We are now ready to give the proof of the chain domination lemma. For the first claim, observe that by Lemma 11, the marginal distribution of $(\widehat{S}_{\tau(k+1)})_{k \geq 0}$ has equal distribution with the chain $(S_k)_{k \geq 0}$ almost surely. In particular, since $\tau(k+1) \geq k$ always and the minimum cannot increase after hitting 0, we have that

$$\begin{aligned} \Pr[\min \mathbf{S}_k > 0] &= \Pr[\min \widehat{\mathbf{S}}_{\tau(k+1)+1} > 0] \\ &\leq \Pr[\min \widehat{\mathbf{S}}_k > 0] \leq \Pr[\widehat{N}_k > 0], \end{aligned}$$

as by Lemma 10 the event $\min \widehat{S}_k \leq \widehat{N}_k$ holds almost surely. This implies that $T(\widehat{\mathbf{S}}) \leq E(\widehat{N})$ almost surely. Therefore, for the consensus time, we get that $T(\mathbf{S})$ is stochastically dominated by $E(N)$, as

$$\begin{aligned} \Pr[T(\mathbf{S}) > k] &\leq \Pr[T(\widehat{\mathbf{S}}) > k] = \Pr[E(\widehat{N}) \geq T(\widehat{\mathbf{S}}) > k] \\ &\leq \Pr[E(\widehat{N}) > k] = \Pr[E(N) > k]. \end{aligned}$$

This establishes the first claim (a) that $T(\mathbf{S}) \preceq E(N)$.

The next step is to show the second claim, i.e., $J(\mathbf{S}) \preceq B(N)$, where $J(\mathbf{S})$ is the number of bad non-competitive reactions that occur in $\mathbf{S}$ and $B(N)$ is the number of births that occur in chain N . By Lemma 10 and Lemma 11, we have for any $k, x \geq 0$ that

$$\begin{aligned} \Pr[J_k(\mathbf{S}) > x] &= \Pr[J_{\tau(k+1)}(\widehat{\mathbf{S}}) > x] \\ &\leq \Pr[B_{\tau(k+1)}(\widehat{N}) \geq J_{\tau(k+1)}(\widehat{\mathbf{S}}) > x] \\ &\leq \Pr[\tau(k+1)(\widehat{N}) > x] \leq \Pr[B(\widehat{N}) > x] \\ &= \Pr[B(N) > x]. \end{aligned}$$

That is, for any $k \geq 0$, we have $J_k(\mathbf{S}) \preceq B(N)$ and so $J(\mathbf{S}) = \sup\{J_k(\mathbf{S}) : k \geq 0\} \preceq B(N)$.

5.2 Constructing a dominating chain for competitive Lotka–Volterra systems

Let be a two-species Lotka–Volterra chain $\mathbf{S}$ with $\gamma = 0$ and $\alpha_{\min} = \min\{\alpha_0, \alpha_1\} > 0$. We will now define a *nice* birth-death chain N that is a dominating chain for $\mathbf{S}$. By the chain domination lemma and results for nice chains given in Section 4, this implies that the consensus time of $\mathbf{S}$ is $O(n)$ in expectation and with high probability and that the number of bad non-competitive events is $O(\log n)$ in expectation and $O(\log^2 n)$ with high probability. Let $\vartheta = \beta + \delta$, $\alpha = \alpha_0 + \alpha_1$. Define the birth probability function p and death probability function q as

$$p(m) = \frac{\vartheta}{\alpha m + \vartheta} \quad \text{and} \quad q(m) = \frac{\alpha_{\min}}{\alpha + 2\vartheta}$$

for all $m > 0$ and $p(0) = q(0) = 0$. Since p attains its maximum at $p(1)$, we have that $p(m) + q(m) \leq p(1) + q(m) \leq 1$ for all $m \geq 0$. Let $(N_t)_{t \geq 0}$ be the birth-death chain defined by p and q and fix $N_0 \geq \min \mathbf{S}_0$. Since $p \in O(1/m)$ and $q(m)$ is a positive constant for $m > 0$, the chain is nice.

Lemma 12. *The nice birth-death chain N defined by p and q is a dominating chain for $\mathbf{S}$.*

Proof. Let $a, b \geq 1$. We need to show that the following conditions hold:

- (D1) $P(a, b) \leq p(\min\{a, b\})$, and
- (D2) $Q(a, b) \geq q(\min\{a, b\})$.

Without loss of generality, assume $a \geq b \geq 1$. For the first condition (D1), note that

$$P(a, b) = \frac{\delta a + \beta b}{\alpha ab + \vartheta(a + b)} \leq \frac{\vartheta a}{\alpha ab + \vartheta a + \vartheta b} \leq \frac{\vartheta}{\alpha b + \vartheta + \vartheta b/a} \leq \frac{\vartheta}{\alpha b + \vartheta} = p(b).$$

For the second condition (D2), note that the probability of a good competitive interaction is at least

$$Q(a, b) \geq \frac{\alpha_{\min} \cdot ab}{\alpha ab + \vartheta(a + b)} = \frac{\alpha_{\min}}{\alpha + \vartheta(1/b + 1/a)} \geq \frac{\alpha_{\min}}{\alpha + 2\vartheta} = q(b). \quad \square$$

The following is a straightforward consequence of the chain domination lemma (Lemma 9), Lemma 12, and the results given in Section 4.

Theorem 13. *Let $\mathbf{S}$ be a Lotka–Volterra system with $\alpha_{\min} > 0$ and $\gamma = 0$ and an initial population of size $n > 0$. The consensus time $T(\mathbf{S})$ and the number $J(\mathbf{S})$ of bad non-competitive events satisfy*

- (a) $\mathbf{E}[T(\mathbf{S})] \in O(n)$ and $T(\mathbf{S}) \in O(n)$ with high probability, and
- (b) $\mathbf{E}[J(\mathbf{S})] \in O(\log n)$ and $J(\mathbf{S}) \in O(\log^2 n)$ with high probability.

Proof. By construction, N is a nice birth-death chain and a dominating chain for the two-species chain $\mathbf{S}$. Set $N_0 = n$. By Lemma 9, we have that $T(\mathbf{S}) \preceq E(N)$ and $J(\mathbf{S}) \preceq B(N)$. The claim (a) follows now from the fact that $E(N) \in \Theta(n)$ in expectation and with high probability by Lemma 5 and Lemma 8. The claim (b) follows from Lemma 6 and Lemma 7. $\square$

6 Lotka–Volterra systems with self-destructive competition

In this section, we show that the threshold for high probability majority consensus lies between $\Omega(\sqrt{\log n})$ and $O(\log^2 n)$ for neutral Lotka–Volterra systems with self-destructive interspecific competition and no intraspecific competition as given by Eq. (1) with $\alpha > 0$, $\beta, \delta \geq 0$ and $\gamma = 0$.

Demographic noise under self-destructive competition. Under self-destructive competition, competitive events cannot change the difference between the two species. Thus, only demographic noise from the individual birth and death reactions play a role. Let $I(\mathbf{S})$ denote the total number of such events before the chain reaches consensus by time $T(\mathbf{S})$. Let $F_t = \Delta_{t-1} - \Delta_t$. Recall from Eq. (3) that the probability of reaching majority consensus is given by $\rho(\mathbf{S}) = 1 - \Pr[F \leq \Delta_0]$, where

$$F = \sum_{t=1}^{T(\mathbf{S})} F_t = \sum_{k=1}^{I(\mathbf{S})} F_{t(k)} \quad (7)$$

and $t(k)$ is the time when the k th non-competitive event occurs. The equality follows from the fact that competitive reactions cannot change the gap under self-destructive competition and for any time step t for which a competitive interaction occurs we have $F_t = 0$.

6.1 Upper bound for self-destructive competition

Theorem 14. *Let $k \geq 0$ be a constant. Suppose $\mathbf{S}$ is a Lotka–Volterra system with self-destructive competition and $n = \mathbf{S}_0 + \mathbf{S}_1$. There is a constant $C(k)$ such that if $\Delta_0 > C(k) \log^2 n$, then*

$$\rho(\mathbf{S}) \geq 1 - 1/n^k.$$

Proof. By Theorem 13, the consensus time $T(\mathbf{S})$ is finite with probability one. From Eq. (7) we have that the probability of reaching majority consensus is given by $\rho(\mathbf{S}) = 1 - \Pr[F \leq \Delta_0]$. Now

$$1 - \rho(\mathbf{S}) = \Pr[F \geq \Delta_0] \leq \Pr[J(\mathbf{S}) \geq \Delta_0] \leq 1/n^k,$$

whenever $\Delta_0 \geq C(k) \log^2 n$, where $C(k)$ is the constant given by Theorem 13. This is because the $J(\mathbf{S})$ non-competitive reactions that reduce the gap between (current) minority and majority species will not exceed Δ_0 with high probability. $\square$

6.2 Lower bound for self-destructive competition

We now show that with self-destructive competition, a Lotka–Volterra system $\mathbf{S}$ can fail to reach majority consensus with constant probability if the initial gap is $o(\sqrt{\log n})$. The idea is to show that when the gap is this small, with at least constant probability, the noise from the individual events will bring the chain into a state (a, a) , for some $a > 0$, where both species have equal counts. From such a state, a system with identical species (i.e., equal rate parameters), both species have equal probability of winning majority consensus (i.e., going extinct last). The next lemma expresses this observation more formally.

Lemma 15. *For a Lotka–Volterra system $\mathbf{S}$ with identical species,*

$$1 - \rho(\mathbf{S}) \geq \frac{1}{2} \cdot \Pr[\Delta_t = 0 \text{ for some } t < T(\mathbf{S})].$$

The next lemma applies to Lotka–Volterra systems with self-destructive and non-self-destructive competition and establishes a lower bound on the number $I(\mathbf{S})$ of individual events.

Lemma 16. *Let $\vartheta = \beta + \delta$, $\alpha > 0$ and $\gamma = 0$. If $\vartheta > 0$, then there exist constants $f, g > 0$ such that $I(\mathbf{S}) \geq f \log m$ with probability at least $1 - 1/m^g$, where m is the initial count of the minority species.*

Proof. Define $M_t = \min \mathbf{S}_t$ for every $t \geq 0$. Note that $M_0 = m$. For every $0 \leq k \leq m$, define the stopping time $t(k) = \min\{t : M_t = k\}$ to be the *first* time the chain reaches a state where the (current) minority species has count k . Note that $t(k) \leq T(\mathbf{S})$ and $T(\mathbf{S})$ is finite with probability 1 by Theorem 13. Let X_t be the indicator variable denoting if a non-competitive event occurs at time step $t \geq 0$. Since the minimum species count can decrease by at most one in each step, the sequence $(M_t)_{0 \leq t < T(\mathbf{S})}$ has to visit every state $m, m-1, \dots, 1$ at some point before time $T(\mathbf{S})$. Therefore,

$$I(\mathbf{S}) = \sum_{t=1}^{T(\mathbf{S})} X_t \geq \sum_{k=1}^m X_{t(k)} = X,$$

where $X = X_{t(1)} + \dots + X_{t(m)}$. Thus, to show the claim of the lemma, it suffices to establish that $X \in \Omega(\log m)$ holds with probability $1 - 1/m$. Define $p(k) = \vartheta/(\alpha k + 2\vartheta)$ and observe that when the chain is in state (a, k) or (k, a) , where $a \geq k \geq 1$, the probability a non-competitive event is

$$\frac{\vartheta(a+k)}{\alpha a k + \vartheta(a+k)} = \frac{\vartheta(1+k/a)}{\alpha k + \vartheta(1+k/a)} \geq \frac{\vartheta}{\alpha k + 2\vartheta} = p(k).$$

In particular, this implies that for every $0 < k \leq m$, we have $\Pr[X_{t(k)} = 1 \mid X_{t(1)}, \dots, X_{t(k-1)}] \geq p(k)$. Define $Y = Y_1 + \dots + Y_m$ as the sum of *independent* Bernoulli random variables with $\Pr[Y_k = 1] = p(k)$. By Lemma 4, we get that $Y \preceq X$ and so $Y \preceq X \preceq I(\mathbf{S})$. Since $p(k) \in \Omega(1/k)$, there is some constant $0 < c < 1$ such that $p(k) \geq c/k$ for all k . Therefore, by linearity of expectation,

$$\mathbf{E}[Y] = \sum_{k=1}^m \mathbf{E}[Y_k] \geq c \sum_{k=1}^m \frac{1}{k} = cH_m \geq c \ln m.$$

Since $Y \preceq X \preceq I(\mathbf{S})$, applying the Chernoff bound (Lemma 1) with $\varepsilon = 1/2$, yields

$$\begin{aligned} \Pr[I(\mathbf{S}) < \frac{c}{2} \ln m] &\leq \Pr[I(\mathbf{S}) \leq (1 - \varepsilon)c \ln m] \leq \Pr[X \leq (1 - \varepsilon)c \ln m] \\ &\leq \Pr[Y \leq (1 - \varepsilon) \cdot \mathbf{E}[Y]] \leq \exp\left(-\frac{\varepsilon^2 \cdot \mathbf{E}[Y]}{2}\right) \leq 1/m^{c/8}. \end{aligned}$$

Setting $f = c/2$ and $g = c/8$ now concludes the proof. $\square$

We are now ready to prove that if the initial gap $\Delta_0 \in o(\sqrt{\log n})$, then Lotka–Volterra systems with self-destructive interspecific competition can fail to reach majority consensus with constant positive probability.

Theorem 17. *Suppose $\mathbf{S}$ is a neutral Lotka–Volterra system with $\beta = \delta$, self-destructive competition $\alpha > 0$, and $\gamma = 0$. Let $\varepsilon > 0$ be a constant. Then there exists a constant $\phi > 0$ such that if $\Delta_0 \leq \phi\sqrt{\log n}$, then $\rho(\mathbf{S}) \leq 1/2 + \varepsilon$ for all sufficiently large n .*

Proof. Let $\mathbf{S}_0 = (m', m)$ be the initial state of the Lotka–Volterra chain, where $m' \geq m > 0$ and $m' \leq m + \phi\sqrt{\log n}$ so that the initial gap is $\Delta_0 \leq \phi\sqrt{\log m}$. By Lemma 16, there is a constant $f > 0$ such that there are $I(\mathbf{S}) \geq K = f \ln m$ individual events before the chain reaches consensus with probability at least $1 - 1/m^g$. Let $t(k)$ be the time exactly $k > 0$ individual events have happened. As before, the probability of reaching majority consensus is given by $\rho(\mathbf{S}) = 1 - \Pr[F \geq \Delta_0]$, where

$$F = \sum_{t=1}^{T(\mathbf{S})} F_t = \sum_{k=1}^{I(\mathbf{S})} F_{t(k)},$$

as competitive reactions cannot change the gap under self-destructive competition. For $1 \leq k \leq K$, let $X_k = F_{t(k)}$ conditioned on the event that $I(\mathbf{S}) \geq K$. By assumption $\beta = \delta$, so we have

$$\begin{aligned} \Pr[X_k = 1 \mid \mathbf{S}_{t(k)} = (a, b)] &= \frac{\delta a + \beta b}{(\beta + \delta)(a + b)} = \frac{1}{2} \\ &= \frac{\beta a + \delta b}{(\beta + \delta)(a + b)} \\ &= \Pr[X_k = -1 \mid \mathbf{S}_{t(k)} = (a, b)]. \end{aligned}$$

Thus, conditioned on $I(\mathbf{S}) \geq K$, the random variables $X_1, \dots, X_K$ are (conditionally) independent. Moreover, they have mean $\mathbf{E}[X_k] = 0$ and variance $\mathbf{Var}[X_k] = 1$. Let $X = X_1 + \dots + X_K$. By Lemma 3, for any constant $\varepsilon > 0$, we can choose a constant $\theta > 0$ such that

$$\Pr[X > \theta\sqrt{K} \mid I(\mathbf{S}) \geq K] \geq 1 - \varepsilon/2$$

for all sufficiently large K . Set $\phi = 3\theta\sqrt{f}$. For large enough n , we have $n \leq 3m$ and hence $\phi\sqrt{\log n} \leq \theta\sqrt{K}$. Conditioned on the event $I(\mathbf{S}) \geq K$, we have that with probability $1 - \varepsilon/2$ the

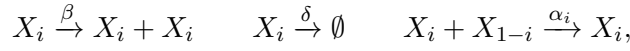
two-species chain reaches some state (a, a) , where $a > 0$. From this point onward, by Lemma 15, the probability that the majority species not winning is at least $1/2$, as the system is neutral. Therefore, when n , and thus m , is sufficiently large, the probability to reach majority consensus is at most

$$\begin{aligned}\rho(\mathbf{S}) &\leq \Pr[I(\mathbf{S}) < K] + \Pr[X \leq \theta\sqrt{K} \mid I(\mathbf{S}) \geq K] + 1/2 \\ &\leq 1/m^g + \varepsilon/2 + 1/2 \leq 1/2 + \varepsilon.\end{aligned}$$

□

7 Lotka–Volterra systems with non-self-destructive competition

In this section, we show that the threshold for high probability majority consensus lies between $\Omega(\sqrt{n})$ and $O(\sqrt{n \log n})$ for Lotka–Volterra systems with non-self-destructive interspecific competition and no intraspecific competition. That is, we consider the model given by the reactions



where $\alpha_i, > 0$, $\beta, \delta \geq 0$ and $i \in \{0, 1\}$. For the upper bound, we allow for non-symmetric competition $\alpha_0 \neq \alpha_1$. That is, the minority species can be a better competitor than the majority species.

Unlike in the previous section with self-destructive competition, when we have non-self-destructive competition, also the competition events give arise to noise which influences the gap between the two species. In particular, there will be $\Theta(m)$ competition events in a system with initial minority of size m , so the noise term F will essentially be $\Omega(\sqrt{m})$ and at most $O(\sqrt{m \log m})$ as we will see.

7.1 Upper bound

Theorem 18. *Suppose $\mathbf{S}$ is a Lotka–Volterra system with non-self-destructive interspecific competition and $\mathbf{S}_0 + \mathbf{S}_1 = n$. For any constant $k \geq 0$, there is a constant $\theta(k)$ such that if $\Delta_0 > \theta(k) \cdot \sqrt{n \log n}$, then $\rho(\mathbf{S}) \geq 1 - 1/n^k$.*

Proof. Let $I(\mathbf{S})$ be the number of individual events before the chain reaches consensus and $K(\mathbf{S})$ be the number of competitive events before the chain reaches consensus. Clearly, $T(\mathbf{S}) = I(\mathbf{S}) + K(\mathbf{S})$. Now

$$F = \sum_{t=1}^{T(\mathbf{S})} F_t = X + Y, \text{ where } X = \sum_{i=1}^{I(\mathbf{S})} X_i \text{ and } Y = \sum_{i=1}^{K(\mathbf{S})} Y_i,$$

where F and F_t are defined as before in Eq. (3), and X_i is the outcome of the i th non-competitive event and Y_i the outcome of the i th competitive event. That is, $X_i = 1$ if the gap between the initial minority and majority species decreases during the i th non-competitive event before the chain reaches consensus, and $X_i = -1$ otherwise. Similarly, $Y_i = 1$ if the gap between the initial minority and majority species decreases during the i th competitive event before the chain reaches consensus, and $Y_i = -1$ otherwise.

Recall that, as before, the probability of majority consensus is $\rho(\mathbf{S}) = \Pr[F < \Delta_0]$. By Theorem 13(b), we have that $\Pr[J(\mathbf{S}) > \theta\sqrt{n \log n}] \leq \Pr[J(\mathbf{S}) > \theta \log^2 n] \leq 1/n^{k+1}$ for some constant $\theta > 0$. That is, there are with high probability $O(\log^2 n)$ events that decrease the gap between the (current) majority and minority species before the chain reaches consensus.

We now show that Y is $O(\sqrt{n \log n})$ with high probability. By Theorem 13(a), we have that $\Pr[K(\mathbf{S}) \geq cn] \leq \Pr[T(\mathbf{S}) \geq cn] \leq 1/n^{k+1}$ for some constant $c > 0$ and all sufficiently large n . Conditioned on the event $K(\mathbf{S}) = \ell$, the sum Y is a sum of ℓ (conditionally) independent random

variables taking values between $[-1, 1]$. Thus, by the law of total probability and applying Hoeffding's inequality with $t = \sqrt{(k+1)cn \ln n}$, we get

$$\begin{aligned}
\Pr[Y \geq t] &= \sum_{\ell=0}^{\infty} \Pr[K(\mathbf{S}) = \ell] \cdot \Pr[Y \geq t \mid K(\mathbf{S}) = \ell] \\
&\leq \sum_{\ell=t}^{cn} \Pr[K(\mathbf{S}) = \ell] \cdot \Pr[Y \geq t \mid K(\mathbf{S}) = \ell] + \sum_{\ell > cn} \Pr[K(\mathbf{S}) = \ell] \\
&\leq 2 \cdot \sum_{\ell=t}^{cn} \Pr[K(\mathbf{S}) = \ell] \cdot \exp\left(-\frac{2t^2}{\ell}\right) + \Pr[K(\mathbf{S}) > cn] \\
&\leq 2cn \cdot \exp\left(-\frac{2t^2}{cn}\right) + 1/n^{k+1} \leq \frac{2c}{n^{2k+1}} + 1/n^{k+1} \leq 2/n^{k+1}
\end{aligned}$$

for all sufficiently large n . Note that $t + \theta\sqrt{n \log n} \leq \theta(k) \cdot \sqrt{n \log n}$ holds for some sufficiently large constant $\theta(k) \geq \theta'$ depending only on k . Therefore, if $\Delta_0 \geq \theta(k)\sqrt{n \log n}$, then

$$\begin{aligned}
1 - \rho(\mathbf{S}) &= \Pr[F \geq \Delta_0] \leq \Pr[J(\mathbf{S}) + Y \geq \Delta_0] \\
&\leq \Pr[J(\mathbf{S}) \geq \theta \cdot \sqrt{n \log n}] + \Pr[Y \geq t] \leq 3/n^{k+1} \leq 1/n^k
\end{aligned}$$

for all sufficiently large n , proving the claim. $\square$

7.2 Lower bound

Theorem 19. *Let $\varepsilon > 0$ be a constant. Suppose $\mathbf{S}$ is a neutral Lotka–Volterra system with $\beta = \delta$, non-self-destructive competition and $\gamma = 0$. Then there exists a constant $\phi > 0$ such that if $\Delta_0 \leq \phi\sqrt{n}$, then $\rho(\mathbf{S}) \leq 1/2 + \varepsilon$ for all sufficiently large $n > 0$.*

Proof. Suppose $\mathbf{S}_0 = (m + \Delta_0, m)$ is the initial state of the chain $\mathbf{S}$ with $m > 0$ and $\Delta_0 \leq \phi\sqrt{m}$, where $\phi > 0$ is a constant we fix later. Now the total initial population size is $n = 2m + \Delta_0 \in \Theta(m)$ and the initial gap is $\Delta_0 \in O(\sqrt{n})$.

Let $X_1, \dots, X_m$ denote the outcomes of the first m events, where $X_i = -1$ if Δ_i decreases in the event and $X_i = 1$ otherwise. By definition, the random variables are independent and identically distributed: $\Pr[X_i = 1] = 1/2 = \Pr[X_i = -1]$ since the competition rates satisfy $\alpha_0 = \alpha_1 > 0$, as the system is neutral (i.e., the species have identical rate parameters) and $\beta = \delta$. Thus, for each $1 \leq i \leq m$, we have $\mathbf{E}[X_i] = 0$ and $\mathbf{Var}[X_i] = 1$. Letting $X = X_1 + \dots + X_m$ and applying Lemma 3, we get

$$\Pr[X \geq \Delta_0] = \Pr[X \geq \theta\sqrt{m}] \geq 1 - \varepsilon$$

Thus, before reaching consensus, the chain reaches some state (a, a) , where $a > 0$, with probability at least $1 - \varepsilon$. From this point onward, by Lemma 15, the probability that the majority species losing is at least $1/2$, as the species have identical birth, death and competition rates. Therefore, when m is sufficiently large, the probability to reach majority consensus is at most

$$\rho(\mathbf{S}) \leq \Pr[X < \theta\sqrt{m}] + 1/2 \leq 1/2 + \varepsilon. \quad \square$$

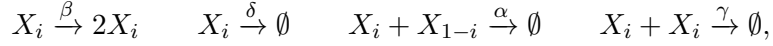
8 Lower bounds for systems with intraspecific competition

In contrast to our earlier results for systems without intraspecific competition, we now show that Lotka–Volterra systems with intraspecific competition (i.e., $\gamma > 0$) can be *much* worse signal

amplifiers than systems with only interspecific competition. Namely, the probability of reaching majority consensus can be very small.

We show that when the system has both intraspecific and interspecific competition, that is, $\alpha, \gamma > 0$ holds, then the threshold for majority consensus can be $\Omega(n)$. In systems with only intraspecific competition, that is, $\gamma > \alpha = 0$, we show that no such threshold exists.

The models with intra- and interspecific competition. In this section, we consider the neutral model with self-destructive interspecific and intraspecific competition given by the reactions



where $i \in \{0, 1\}$ and $\alpha, \gamma, \beta, \delta \geq 0$.

8.1 Systems with both intra- and interspecific competition

Self-destructive competition. We consider first the case of self-destructive competition. We now establish the following result.

Theorem 20. *Let $\mathbf{S}$ be a Lotka–Volterra chain with self-destructive competition and initial configuration $\mathbf{S}_0 = (a, b)$, where $a \geq b > 0$, $\alpha = \gamma \geq 0$ and $\beta, \delta \geq 0$. If the chain reaches consensus with probability 1, then*

$$\rho(\mathbf{S}) = \frac{a}{a+b}.$$

The above immediately implies that the $\Delta_0 = n - 1$ is the only positive gap that can guarantee majority consensus under self-destructive competition with probability at least $1 - 1/n$ if $\gamma = \alpha$ holds. The above result is a generalization of a similar result by Andaur et al. [6], who established the same bound for the case $\alpha = \gamma = 0$ two *independent*, non-competing species, to systems with two competing *non-independent* species. Let us write $\rho(a, b) = \rho(\mathbf{S})$ and $T(a, b) = T(\mathbf{S})$ for a chain $\mathbf{S}$ with initial state (a, b) . Let P be the transition probability matrix of the chain $\mathbf{S}$. By the law of total probability and the fact that the chain is Markovian, we have that ρ satisfies the recurrence

$$\rho(a, b) = \sum_{x, y \geq 0} P((a, b), (x, y)) \cdot \rho(x, y). \quad (8)$$

with boundary conditions $\rho(a, 0) = 1$ for all $a > 0$ and $\rho(0, b) = 0$ for all $b \geq 0$. A straightforward but tedious calculation shows if $\alpha = \gamma$, then $\rho(a, b) = a/(a+b)$ is a solution for the recurrence. Furthermore, assuming that the chain reaches consensus with probability 1, we can show that the solution is unique.

Lemma 21. *The function $a/(a+b)$ is a solution for the recurrence given in Eq (8).*

Proof. We now give the proof in the case of self-destructive competitive interactions, i.e., a competitive interaction removes both participating individuals from the system. There are in total seven types of events: for both species, there are individual birth and death events, which change the population count by one, and *intraspecific* competitive events, where two individuals of the same species interact and are removed. Finally, there are the *interspecific* competition events that remove one individual of both species. With this in mind, we can rewrite the recurrence as

$$\rho(a, b) = \sum_{(j, k) \in \mathcal{T}_1 \cup \mathcal{T}_2} P((a, b), (a+j, b+k)) \cdot \rho(a+j, b+k),$$

where

$$\mathcal{T}_1 = \{(1, 0), (0, 1), (-1, 0), (0, -1)\} \quad \text{and} \quad \mathcal{T}_2 = \{(-2, 0), (0, -2), (-1, -1)\}.$$

Let $r_i(a, b)$ be the conditional probability that the first species wins when starting from state (a, b) conditioned on the event that the next step involves i individuals. Let $\mathcal{F}$ denote the event that the next step involves exactly one individual and observe that

$$\rho(a, b) = \Pr[\mathcal{F}] \cdot r_1(a, b) + (1 - \Pr[\mathcal{F}]) \cdot r_2(a, b).$$

We now show that $r_i(a, b) = a/(a + b)$ for both $i \in \{1, 2\}$, which then implies that $a/(a + b)$ is a solution for the recurrence. First, a simple calculation shows that

$$r_1(a, b) = \frac{\beta a \cdot \rho(a + 1, b) + \delta a \cdot \rho(a - 1, b) + \beta b \cdot \rho(a, b + 1) + \delta b \cdot \rho(a, b - 1)}{(\beta + \delta)(a + b)}$$

Plugging in $\rho(a, b) = a/(a + b)$ yields

$$\begin{aligned} r_1(a, b) &= \frac{1}{(\beta + \delta)(a + b)} \left(\beta a \cdot \frac{a + 1}{a + b + 1} + \delta a \cdot \frac{a - 1}{a + b - 1} + \beta b \cdot \frac{a}{a + b + 1} + \delta b \cdot \frac{a}{a + b - 1} \right) \\ &= \frac{a}{a + b}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} r_2(a, b) &= \frac{\alpha ab \cdot \rho(a - 1, b - 1) + \gamma a(a - 1)/2 \cdot \rho(a - 2, b) + \gamma b(b - 1)/2 \cdot \rho(a, b - 2)}{\alpha ab + \gamma a(a - 1)/2 + \gamma b(b - 1)/2} \\ &= \frac{2ab \cdot \rho(a - 1, b - 1) + a(a - 1) \cdot \rho(a - 2, b) + b(b - 1) \cdot \rho(a, b - 2)}{2ab + a(a - 1) + b(b - 1)} \end{aligned}$$

using the assumption $\alpha = \gamma$. Using the identity $(a + b)(a + b - 1) = 2ab + a(a - 1) + b(b - 1)$ twice, and plugging in $\rho(a, b) = a/(a + b)$ we get that

$$r_2(a, b) = \frac{a}{(a + b)(a + b - 1)} \cdot \frac{2(a - 1)b + (a - 1)(a - 2) + b(b - 1)}{(a + b - 2)} = \frac{a}{a + b}. \quad \square$$

Lemma 22. *If the chain reaches consensus with probability one, then the recurrence given by Eq. (8) has a unique solution.*

Proof. Define $d(a, b) = f(a, b) - g(a, b)$, where f and g are two solutions for the recurrence. We will prove the claim by showing that $|d(a, b)| < \varepsilon$ for any $\varepsilon > 0$. Under the assumption that the chain reaches consensus with probability 1, we have that the consensus time $T(a, b)$ satisfies

$$\lim_{k \rightarrow \infty} \Pr[T(a, b) > k] = 0.$$

Therefore, for any $\varepsilon > 0$, there exists k such that $\Pr[T(\mathbf{S}) > k] < \varepsilon$. Now for any $k > 0$, $d(a, b)$ satisfies the recurrence

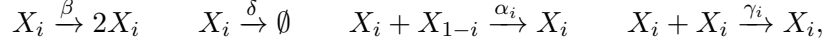
$$d(a, b) = \sum_{x, y \geq 0} P^k((a, b), (x, y)) \cdot d(x, y) = \sum_{x, y > 0} P^k((a, b), (x, y)) \cdot d(x, y),$$

where the last equality follows fact that $d(0, x) = d(x, 0) = 0$ for any $x \geq 0$ due to the boundary conditions of the recurrence for ρ . Since $|d(x, y)| \leq 1$, we have that

$$|d(a, b)| \leq \sum_{x, y > 0} P^k((a, b), (x, y)) \cdot |d(x, y)| \leq \Pr[T(a, b) > k] < \varepsilon. \quad \square$$

Now the two lemmas imply Theorem 20.

Non-self-destructive competition. We next consider the case of non-self-destructive competition. The model with non-self-destructive competition is given by



where we assume that $\alpha = \alpha_0 + \alpha_1 = 2\alpha_0$ and $\gamma = \gamma_0 + \gamma_1 = 2\gamma_0$, i.e., that both species have equal rate parameters.

Theorem 23. *Let $\mathbf{S}$ be a neutral Lotka–Volterra system with non-self-destructive competition and $\mathbf{S}_0 = (a, b)$, where $a \geq b > 0$, and $\gamma = 2\alpha$. If the chain reaches consensus with probability 1, then*

$$\rho(\mathbf{S}) = \frac{a}{a+b}.$$

Proof. With non-self-destructive competition, the recurrence for ρ simplifies to

$$\rho(a, b) = \sum_{(j,k) \in \mathcal{T}} P((a, b), (a+j, b+k)) \cdot \rho(a+j, b+k),$$

where $\mathcal{T} = \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$, as the counts of species can change by at most one under non-self-destructive competition. We now show that $a/(a+b)$ is a solution for the recurrence $\rho(a, b)$ also under non-self-destructive competition. We proceed as in the proof of Lemma 21 and compute the conditional probabilities r_i for $i \in \{0, 1\}$. The analysis for the function $r_1(a, b) = a/(a+b)$ is identical as in Lemma 21. The case for the function $r_2(a, b)$ is different. If the chain is in state (a, b) , conditioned on the event that the next step has a competitive interaction, the probability that the first species with count a decreases is

$$\frac{\alpha_0 ab + \gamma_0 a(a-1)/2}{\alpha ab + \gamma_0 a(a-1)/2 + \gamma_1 b(b-1)/2} = \frac{ab + a(a-1)}{2ab + a(a-1) + b(b-1)} = \frac{a(a+b-1)}{(a+b)(a+b-1)} = \frac{a}{a+b}.$$

Thus, we get that

$$\begin{aligned} r_2(a, b) &= \frac{a}{a+b} \cdot \rho(a-1, b) + \left(1 - \frac{a}{a+b}\right) \cdot \rho(a, b-1) \\ &= \frac{a}{a+b} \cdot \frac{a-1}{a+b-1} + \frac{b}{a+b} \cdot \frac{a}{a+b-1} = \frac{a}{a+b} \cdot \frac{a+b-1}{a+b-1} = \frac{a}{a+b}. \end{aligned}$$

This implies that $\rho(a, b) = a/(a+b)$ is a solution for the recurrence ρ . By Lemma 22, the solution is unique under the assumption that the chain reaches consensus with probability 1. $\square$

8.2 Systems with intraspecific competition only

Finally, to complete the picture, we consider the limiting case with intraspecific competition only ($\gamma > 0$ and $\alpha = 0$). As expected, this case is even worse: for *any* initial configuration, where both species are present, the system fails to reach majority consensus with constant probability.

In the case $\alpha = 0$, it is convenient to consider the continuous-time version of the Lotka–Volterra chain, as in this case the two chains for the individual species are independent, continuous-time single-species processes. We first introduce the following lemma, which readily follows from standard absorption time results on continuous-time birth-death chains [51]. For example, the next result can be proven using a straightforward adaption of the proof of given by Andaur et al. [21, Lemma 6].

Lemma 24. *Let M be a continuous-time birth-death chain with 0 as the unique absorbing state. If the birth rate $\kappa(n)$ and death rate $\mu(n)$ satisfy $\kappa(n) \in \Theta(n)$ and $\mu(n) \in \Theta(n^2)$, then the mean absorption time from any state m is $O(1)$.*

Theorem 25. *Let $\mathbf{S}$ be a Lotka–Volterra chain with intraspecific competition and without interspecific competition, i.e., $\alpha = 0$ and $\gamma > 0$. Then $\mathbf{S}$ fails to reach majority consensus with at least constant probability from any starting state $\mathbf{S}_0 = (a, b)$, with $a \geq b > 0$.*

Proof. For the proof of this result, it is convenient to consider the continuous-time version $\mathbf{X} = (X_0, X_1)$ of the Lotka–Volterra chain instead of the discrete-time jump chain $\mathbf{S}$. Since $\alpha = 0$, the two continuous-time chains X_0 and X_1 are independent. Let T_i denote the time until the species $i \in \{0, 1\}$ goes extinct, i.e., the absorption time of the chain X_i . First, note that $\mathbf{E}[T_i] \in O(1)$. For non-self-destructive competition, this follows from Lemma 24, as the birth and death rates for the chains X_i satisfy $\kappa(n) = \beta n$ and $\mu(n) = \delta n + \gamma_i n(n-1)/2$. For self-destructive competition, the extinction time is also $O(1)$, as death events under self-destructive competition also always decrease the population count by at least one.

Let $Y(\lambda)$ denote the exponentially distributed random variable with mean $1/\lambda$. Observe that for the species i to go extinct, its chain transitions to state 0 via either

- (1) state 1, i.e., the last event before extinction is an individual death event, which happens at rate δ so the time between the second last and last transition is $Y(\delta)$, and
- (2) state 2, i.e., the last event before extinction is a (self-destructive) competitive interaction, which happens at rate γ so the time between the second last and last transition is $Y(\gamma)$.

Note that if the system does not have self-destructive competition, then the second case cannot occur. Let X be an indicator random variable for whether the first case occurs and set $\lambda = \max\{\alpha, \delta\}$. Now

$$\begin{aligned} \Pr[T_i \geq x] &= \Pr[X = 1] \cdot \Pr[T_i \geq x \mid X = 1] + \Pr[X = 0] \cdot \Pr[T_i \geq x \mid X = 0] \\ &\geq \Pr[X = 1] \cdot \Pr[Y(\delta) \geq x] + \Pr[X = 0] \cdot \Pr[Y(\gamma) \geq x] \\ &\geq \Pr[Y(\lambda) \geq x] = \exp(-\lambda x), \end{aligned}$$

where the last equality is given by the CDF of the exponential distribution. By Markov's inequality,

$$\Pr[T_0 < 2 \cdot \mathbf{E}[T_0]] \geq 1 - \Pr[T_0 \geq 2 \cdot \mathbf{E}[T_0]] \geq 1/2.$$

Since T_0 and T_1 are independent random variables, we have

$$\Pr[T_0 < T_1] \geq \Pr[T_0 < 2 \cdot \mathbf{E}[T_0]] \cdot \Pr[T_1 \geq 2 \cdot \mathbf{E}[T_0]] \geq \frac{1}{2} \cdot \exp(-2\lambda \cdot \mathbf{E}[T_0]),$$

where the last term is lower bounded by some positive constant $\theta > 0$ since $\lambda, \mathbf{E}[T_0] \in O(1)$. That is, the minority species survives longer with probability at least θ . $\square$

Acknowledgements

This research was supported by the ANR projects DREAMY (ANR-21-CE48-0003) and COSTX-PRESS (ANR-23-CE45-0013).

References

- [1] Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili, and Ronald L. Rivest. 2017. Time-space trade-offs in population protocols. In *Proc. 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2560–2579. <https://doi.org/10.1137/1.9781611974782.169>
- [2] Dan Alistarh, James Aspnes, and Rati Gelashvili. 2018. Space-optimal majority in population protocols. In *Proc. 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. <https://doi.org/10.1137/1.9781611975031.144>
- [3] Dan Alistarh and Rati Gelashvili. 2018. Recent algorithmic advances in population protocols. *ACM SIGACT News* (2018). <https://doi.org/10.1145/3289137.3289150>
- [4] Dan Alistarh, Rati Gelashvili, and Milan Vojnović. 2015. Fast and exact majority in population protocols. In *Proc. 2015 ACM Symposium on Principles of Distributed Computing (PODC)*. 47–56. <https://doi.org/10.1145/2767386.2767429>
- [5] Razan N. Alnahhas, Mehdi Sadeghpour, Ye Chen, Alexis A. Frey, William Ott, Krešimir Josić, and Matthew R. Bennett. 2020. Majority sensing in synthetic microbial consortia. *Nature Communications* 11, 1 (2020), 1–10. <https://doi.org/10.1038/s41467-020-17475-z>
- [6] Victoria Andaur, Janna Burman, Matthias Függer, Manish Kushwaha, Bilal Manssouri, Thomas Nowak, and Joel Rybicki. 2021. Reaching Agreement in Competitive Microbial Systems. *arXiv preprint arXiv:2103.07450* (2021).
- [7] Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. 2006. Computation in networks of passively mobile finite-state sensors. *Distributed Computing* 18, 4 (2006), 235–253. <https://doi.org/10.1007/s00446-005-0138-3>
- [8] Dana Angluin, James Aspnes, and David Eisenstat. 2008. A simple population protocol for fast robust approximate majority. *Distributed Computing* 21, 2 (2008), 87–102. <https://doi.org/10.1007/s00446-008-0059-z>
- [9] Gregor Bankhamer, Petra Berenbrink, Felix Biermeier, Robert Elsässer, Hamed Hosseinpour, Dominik Kaaser, and Peter Kling. 2022. Fast Consensus via the Unconstrained Undecided State Dynamics. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 3417–3429.
- [10] Gregor Bankhamer, Petra Berenbrink, Felix Biermeier, Robert Elsässer, Hamed Hosseinpour, Dominik Kaaser, and Peter Kling. 2022. Population protocols for exact plurality consensus. In *Proc. 2022 ACM Symposium on Principles of Distributed Computing (PODC)*. <https://doi.org/10.1145/3519270.3538447>
- [11] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Riccardo Silvestri. 2014. Plurality consensus in the gossip model. In *Proc. 26th ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 371–390. <https://doi.org/10.1007/s00446-016-0289-4>
- [12] Yaakov Benenson. 2012. Biomolecular computing systems: principles, progress and potential. *Nature Reviews Genetics* 13, 7 (2012), 455–468. <https://doi.org/10.1038/nrg3197>
- [13] Petra Berenbrink, George Giakkoupis, and Peter Kling. 2020. Optimal time and space leader election in population protocols. In *Proc. 52nd ACM Symposium on Theory of Computing (STOC)*. <https://doi.org/10.1145/3357713.3384312>

- [14] Philip Bittihn, M. Omar Din, Lev S. Tsimring, and Jeff Hasty. 2018. Rational engineering of synthetic microbial systems: from single cells to consortia. *Current Opinion in Microbiology* 45 (2018), 92–99. <https://doi.org/10.1016/j.mib.2018.02.009>
- [15] Andrew J Black and Alan J McKane. 2012. Stochastic formulation of ecological models and their applications. *Trends in ecology & evolution* 27, 6 (2012), 337–345. <https://doi.org/10.1016/j.tree.2012.01.014>
- [16] Pierre Brémaud. 1999. *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, Heidelberg.
- [17] Katie Brenner, Lingchong You, and Frances H. Arnold. 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology* 26, 9 (2008), 483–489. <https://doi.org/10.1016/j.tibtech.2008.05.004>
- [18] D. Ewen Cameron, Caleb J. Bashor, and James J. Collins. 2014. A brief history of synthetic biology. *Nature Reviews Microbiology* 12, 5 (2014), 381–390. <https://doi.org/10.1038/nrmicro3239>
- [19] Luca Cardelli and Attila Csikász-Nagy. 2012. The cell cycle switch computes approximate majority. *Scientific reports* 2, 1 (2012), 656. <https://doi.org/10.1038/srep00656>
- [20] Bernard Chazelle. 2012. Natural algorithms and influence systems. *Commun. ACM* 55, 12 (2012), 101–110. <https://doi.org/10.1145/2380656.2380679>
- [21] Da-Jung Cho, Matthias Függer, Corbin Hopper, Manish Kushwaha, Thomas Nowak, and Quentin Soubeyran. 2021. Distributed computation with continual population growth. *Distributed Computing* (2021), 1–23. <https://doi.org/10.1007/s00446-021-00404-8>
- [22] Anne Condon, Monir Hajiaghayi, David Kirkpatrick, and Ján Maňuch. 2020. Approximate majority analyses using tri-molecular chemical reaction networks. *Natural Computing* 19, 1 (2020), 249–270. <https://doi.org/10.1007/s11047-019-09756-4>
- [23] Colin Cooper, Robert Elsässer, and Tomasz Radzik. 2014. The power of two choices in distributed voting. In *International Colloquium on Automata, Languages, and Programming*. Springer, 435–446. https://doi.org/10.1007/978-3-662-43951-7_37
- [24] Jurek Czyzowicz, Leszek Gąsieniec, Adrian Kosowski, Evangelos Kranakis, Paul G. Spirakis, and Przemysław Uznański. 2022. On convergence and threshold properties of discrete Lotka-Volterra population protocols. *J. Comput. System Sci.* 130 (2022), 1–25. <https://doi.org/10.1016/j.jcss.2022.06.002>
- [25] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. 2010. A synchronized quorum of genetic clocks. *Nature* 463, 7279 (2010), 326–330. <https://doi.org/10.1038/nature08753>
- [26] Sandra Dedrick, Vaishnavi Warriar, Katherine P Lemon, and Babak Momeni. 2023. When does a Lotka-Volterra model represent microbial interactions? Insights from *in vitro* nasal bacterial communities. *Msystems* (2023), e00757–22. <https://doi.org/10.1128/msystems.00757-22>
- [27] Domitilla Del Vecchio, Yili Qian, Richard M. Murray, and Eduardo D. Sontag. 2018. Future systems and control research in synthetic biology. *Annual Reviews in Control* 45 (2018), 5–17. <https://doi.org/10.1016/j.arcontrol.2018.04.007>

- [28] Alexander Dobrinevski and Erwin Frey. 2012. Extinction in neutrally stable stochastic Lotka-Volterra models. *Physical Review E* 85, 5 (2012), 051903. <https://doi.org/10.1103/PhysRevE.74.051907>
- [29] David Doty. 2012. Theory of algorithmic self-assembly. *Commun. ACM* (2012). <https://doi.org/10.1145/2380656.2380675>
- [30] David Doty, Mahsa Eftekhari, Leszek Gąsieniec, Eric Severson, Przemysław Uznański, and Grzegorz Stachowiak. 2022. A time and space optimal stable population protocol solving exact majority. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 1044–1055. <https://doi.org/10.1109/FOCS52979.2021.00104>
- [31] Moez Draief and Milan Vojnović. 2012. Convergence speed of binary interval consensus. *SIAM Journal on Control and Optimization* 50, 3 (2012), 1087–1109. <https://doi.org/10.1137/110823018>
- [32] Michael B. Elowitz and Stanislas Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 6767 (2000), 335–338. <https://doi.org/10.1038/35002125>
- [33] Ofer Feinerman, Bernhard Haeupler, and Amos Korman. 2017. Breathe before speaking: efficient information dissemination despite noisy, limited and anonymous communication. *Distributed Computing* 30, 5 (2017), 339–355. <https://doi.org/10.1007/s00446-015-0249-4>
- [34] Ofer Feinerman and Amos Korman. 2013. Theoretical distributed computing meets biology: A review. In *Proc. 9th International Conference on Distributed Computing and Internet Technology (ICDIT)*. https://doi.org/10.1007/978-3-642-36071-8_1
- [35] Jasmin Fisher, David Harel, and Thomas A. Henzinger. 2011. Biology as reactivity. *Commun. ACM* 54, 10 (2011), 72–82. <https://doi.org/10.1145/2001269.2001289>
- [36] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 6767 (2000), 339–342. <https://doi.org/10.1038/35002131>
- [37] G. F. Gause. 1934. *The Struggle for Existence*. Williams & Wilkins Company, Baltimore.
- [38] Stefan Geritz and Éva Kisdi. 2012. Mathematical ecology: why mechanistic models? *Journal of mathematical biology* 65, 6-7 (2012), 1411. <https://doi.org/10.1007/s00285-011-0496-3>
- [39] Mohsen Ghaffari and Merav Parter. 2016. A polylogarithmic gossip algorithm for plurality consensus. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*. 117–126.
- [40] Daniel T. Gillespie. 1977. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81, 25 (1977), 2340–2361. <https://doi.org/10.1021/j100540a008>
- [41] Shafi Goldwasser, Rafail Ostrovsky, Alessandra Scafuro, and Adam Sealfon. 2018. Population stability: regulating size in the presence of an adversary. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing*. 397–406. <https://doi.org/10.1145/3212734.3212747>
- [42] Didier Gonze, Katharine Z. Coyte, Leo Lahti, and Karoline Faust. 2018. Microbial communities as dynamical systems. *Current Opinion in Microbiology* 44 (2018), 41–49. <https://doi.org/10.1016/j.mib.2018.07.004>

- [43] Elisa T. Granato, Thomas A. Meiller-Legrand, and Kevin R. Foster. 2019. The evolution and ecology of bacterial warfare. *Current Biology* 29, 11 (2019), R521–R537. <https://doi.org/10.1016/j.cub.2019.04.024>
- [44] Nicolas E. Grandel, Kiara Reyes Gamas, and Matthew R. Bennett. 2021. Control of synthetic microbial consortia in time, space, and composition. *Trends in Microbiology* (2021). <https://doi.org/10.1016/j.tim.2021.04.001>
- [45] Lewis Grozinger, Martyn Amos, Thomas E. Gorochofski, Pablo Carbonell, Diego A. Oyarzún, Ruud Stoof, Harold Fellermann, Paolo Zuliani, Huseyin Tas, and Angel Goñi-Moreno. 2019. Pathways to cellular supremacy in biocomputing. *Nature Communications* (2019). <https://doi.org/10.1038/s41467-019-13232-z>
- [46] Josef Hofbauer and Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge University Press. <https://doi.org/10.1017/CB09781139173179>
- [47] Jiliang Hu, Daniel R Amor, Matthieu Barbier, Guy Bunin, and Jeff Gore. 2022. Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science* 378, 6615 (2022), 85–89. <https://doi.org/10.1126/science.abm7841>
- [48] Weini Huang, Christoph Hauert, and Arne Traulsen. 2015. Stochastic game dynamics under demographic fluctuations. *Proceedings of the National Academy of Sciences* 112, 29 (2015), 9064–9069. <https://doi.org/10.1073/pnas.1418745112>
- [49] Peter Jagers. 2010. A plea for stochastic population dynamics. *Journal of Mathematical Biology* 60 (2010), 761–764. <https://doi.org/10.1007/s00285-010-0335-y>
- [50] Behzad D. Karkaria, Neythen J. Treloar, Chris P. Barnes, and Alex J. H. Fedorec. 2020. From microbial communities to distributed computing systems. *Frontiers in Bioengineering and Biotechnology* (2020). <https://doi.org/10.3389/fbioe.2020.00834>
- [51] S Karlin and H.M Taylor. 1975. *A First Course in Stochastic Processes* (2 ed.). Academic Press.
- [52] Russell Lande, Steinar Engen, and Bernt-Erik Saether. 2003. *Stochastic population dynamics in ecology and conservation*. Oxford University Press, USA.
- [53] Anselm Levskaya, Aaron A. Chevalier, Jeffrey J. Tabor, Zachary Booth Simpson, Laura A. Lavery, Matthew Levy, Eric A. Davidson, Alexander Scouras, Andrew D. Ellington, Edward M. Marcotte, and Christopher A. Voigt. 2005. Engineering *Escherichia coli* to see light. *Nature* 438, 7067 (2005), 441–442. <https://doi.org/10.1038/nature04405>
- [54] Shuyao Li, Jing Xiao, Tianzheng Sun, Fangjian Yu, Kaihang Zhang, Yuantao Feng, Chenchao Xu, Baojun Wang, and Lei Cheng. 2022. Synthetic microbial consortia with programmable ecological interactions. *Methods in Ecology and Evolution* 13, 7 (2022), 1608–1621. <https://doi.org/10.1111/2041-210X.13894>
- [55] Alfred James Lotka. 1925. *Elements of physical biology*. Williams & Wilkins.
- [56] Javier Macía, Francesc Posas, and Ricard V. Solé. 2012. Distributed computation: the new wave of synthetic biology devices. *Trends in Biotechnology* 30, 6 (2012), 342 – 349. <https://doi.org/10.1016/j.tibtech.2012.03.006>

- [57] Junwen Mao, Andrew E Blanchard, and Ting Lu. 2015. Slow and steady wins the race: a bacterial exploitative competition strategy in fluctuating environments. *ACS Synthetic Biology* 4, 3 (2015), 240–248. <https://doi.org/10.1021/sb4002008>
- [58] John P Marken and Richard M Murray. 2023. Addressable and adaptable intercellular communication via DNA messaging. *Nature Communications* 14, 1 (2023), 2358. <https://doi.org/10.1038/s41467-023-37788-z>
- [59] Robert May and Angela McLean. 2007. *Theoretical Ecology: Principles and Applications* (third ed.). <https://doi.org/10.1093/oso/9780199209989.001.0001>
- [60] Alan J McKane and Timothy J Newman. 2004. Stochastic models in population biology and their deterministic analogs. *Physical Review E* 70, 4 (2004), 041902. <https://doi.org/10.1103/PhysRevE.70.041902>
- [61] George B. Mertzios, Sotiris E. Nikolettseas, Christoforos Raptopoulos, and Paul G. Spirakis. 2014. Determining Majority in Networks with Local Interactions and Very Small Local Memory. In *Proc. 41st International Colloquium on Automata, Languages, and Programming (ICALP)*. 871–882. https://doi.org/10.1007/978-3-662-43948-7_72
- [62] James D. Murray. 2002. *Mathematical Biology*. <https://doi.org/10.1007/b98868>
- [63] Saket Navlakha and Ziv Bar-Joseph. 2014. Distributed information processing in biological and computational systems. *Commun. ACM* (2014). <https://doi.org/10.1145/2678280>
- [64] Etienne Perron, Dinkar Vasudevan, and Milan Vojnović. 2009. Using three states for binary consensus on complete graphs. In *IEEE INFOCOM 2009*. IEEE, 2527–2535. <https://doi.org/10.1109/INFCOM.2009.5062181>
- [65] Sabrina Rashid, Gadi Taubenfeld, and Ziv Bar-Joseph. 2021. The epigenetic consensus problem. In *International Colloquium on Structural Information and Communication Complexity*. Springer, 146–163. https://doi.org/10.1007/978-3-030-79527-6_9
- [66] Sergi Regot, Javier Macia, Núria Conde, Kentaro Furukawa, Jimmy Kjellén, Tom Peeters, Stefan Hohmann, Eulália De Nadal, Francesc Posas, and Ricard Solé. 2011. Distributed biological computation with multicellular engineered networks. *Nature* (2011). <https://doi.org/10.1038/nature09679>
- [67] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. 2006. Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model. *Physical Review E* 74, 5 (2006), 051907. <https://doi.org/10.1103/PhysRevE.74.051907>
- [68] Bruno Sericola. 2013. Birth-and-Death Processes. *Markov Chains: Theory, Algorithms and Applications* (2013), 191–233.
- [69] Richard R Stein, Vanni Bucci, Nora C Toussaint, Charlie G Buffie, Gunnar Rätsch, Eric G Pamer, Chris Sander, and Joao B Xavier. 2013. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Computational Biology* 9, 12 (2013), e1003388. <https://doi.org/10.1371/journal.pcbi.1003388>
- [70] Alvin Tamsir, Jeffrey J Tabor, and Christopher A Voigt. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. *Nature* 469, 7329 (2011), 212–215. <https://doi.org/10.1038/nature09565>

- [71] Vito Volterra. 1926. Fluctuations in the abundance of a species considered mathematically. *Nature* 118, 2972 (1926), 558–560.
- [72] Darren J Wilkinson. 2018. *Stochastic Modelling for Systems Biology*. CRC press.
- [73] Jizhong Zhou and Daliang Ning. 2017. Stochastic community assembly: does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews* 81, 4 (2017), 10–1128. <https://doi.org/10.1128/mmb.00002-17>
- [74] Ali R. Zomorodi and Daniel Segrè. 2016. Synthetic ecology of microbes: mathematical models and applications. *Journal of Molecular Biology* 428, 5 (2016), 837–861. <https://doi.org/10.1016/j.jmb.2015.10.019>

9 Omitted proofs

9.1 Proof of Lemma 3

Proof. By the Central Limit Theorem, we have that the random variable $X/\sqrt{n}$ converges in distribution to the standard normal distribution, i.e., for any $x \in \mathbb{R}$, we have

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{X}{\sqrt{n}} \leq x \right] = \Phi(x),$$

where Φ is the CDF of the standard normal distribution. In particular, for any $\kappa > 0$ and $\theta > 0$, there exists some $n_0 > 0$ such that

$$\Pr [X > \theta\sqrt{n}] = 1 - \Phi(\theta) - \kappa \geq 1 - \frac{e^{-\theta^2/2}}{\sqrt{2\pi}(\theta + 1)} - \kappa \geq \varepsilon$$

for all $n \geq n_0$, where in the second inequality we used the bound $\Phi(\theta) \leq \frac{e^{-\theta^2/2}}{\sqrt{2\pi}(1+\theta)}$ for $\theta > 0$. The claim now follows by choosing θ and κ to be sufficiently small constants and n sufficiently large. $\square$

9.2 Proof of Lemma 4

Proof. For the case (a), we construct a coupling $(\hat{X}, \hat{Y})$ of X and Y such that $\hat{X} \leq \hat{Y}$. Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables sampled uniformly from the unit range $[0, 1)$. We construct the coupling inductively. For each step $1 \leq i \leq n$, the coupling uses the following rules:

- (1) Let $p_i = \Pr[Y_i]$. If $\xi_i \in [0, p_i)$, then set $\hat{Y}_i = 1$. Otherwise, set $\hat{Y}_i = 0$.
- (2) Let $P_i = \Pr[X_i \mid X_1 = \hat{X}_1, \dots, X_{i-1} = \hat{X}_{i-1}]$. If $\xi_i \in [0, P_i)$, then set $\hat{X}_i = 1$. Otherwise, set $\hat{X}_i = 0$.

By construction, the distribution of $\hat{X} = \hat{X}_1 + \dots + \hat{X}_n$ is the same as the distribution of X and $\hat{Y} = \hat{Y}_1 + \dots + \hat{Y}_n$ has the same distribution as Y . Moreover, $\hat{X}_i \leq \hat{Y}_i$, as by assumption we have $[0, P_i] \subseteq [0, p_i)$ for all $1 \leq i \leq n$. In particular, $\Pr[\hat{X}_i \geq x] \leq \Pr[\hat{Y}_i \geq x]$, so we get that

$$\Pr[Y \geq x] = \Pr[\hat{Y} \geq x] \geq \Pr[\hat{X} \geq x] = \Pr[X \geq x],$$

so $X \preceq Y$. The proof for the case (b) is symmetric and follows analogously by observing that $[0, p_i] \subseteq [0, P_i)$ implies that $\hat{Y}_i \leq \hat{X}_i$ for each $1 \leq i \leq n$. $\square$