# Distributed computation with continual population growth

Da-Jung Cho, Matthias Függer, Corbin Hopper, Manish Kushwaha, Thomas Nowak, Quentin Soubeyran

# Distributed Computation with Continual Population Growth

Da-Jung Cho · Matthias Függer · Corbin Hopper · Manish Kushwaha · Thomas Nowak · Quentin Soubeyran

**Abstract** Computing via synthetically engineered bacteria is a vibrant and active field with numerous applications in bio-production, bio-sensing, and medicine. Motivated by the lack of robustness and by resource limitation inside single cells, distributed approaches with communication among bacteria have recently gained in interest. In this paper, we focus on the problem of population growth happening concurrently, and possibly interfering, with the desired bio-computation. Specifically, we present a fast protocol in systems with continuous population growth for the majority consensus problem and prove that it correctly identifies the initial majority among two inputs with high probability if the initial difference is $\Omega(\sqrt{n \log n})$ where $n$ is the total initial population. We also present a fast protocol that correctly computes the NAND of two inputs with high probability. By combining NAND gates with the majority consensus protocol as an amplifier, it is possible to compute arbitrary Boolean functions. Finally, we extend the protocols to several biologically relevant settings. We simulate a plausible implementation of a noisy NAND gate with engineered bacteria. In the context of continuous cultures with a constant outflow and a constant inflow of fresh media, we demonstrate that majority consensus is achieved only if the flow is slower than the maximum growth rate. Simulations suggest that flow increases consensus time over a wide parameter range. The proposed protocols help set the stage for bio-engineered distributed computation that directly addresses continuous stochastic population growth.

D.-J. Cho
University of Kassel, School of Electrical Engineering and Computer Science, Kassel, Germany
E-mail: dajung.cho@uni-kassel.de

M. Függer
CNRS, LMF, ENS Paris-Saclay, Université Paris-Saclay, Inria, Gif-sur-Yvette, France
E-mail: mfuegger@lsv.fr

C. Hopper
ENS Paris-Saclay, Gif-sur-Yvette, France
Université Paris-Saclay, CNRS, Orsay, France
E-mail: corbinshopper@gmail.com

M. Kushwaha
Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France
E-mail: manish.kushwaha@inrae.fr

T. Nowak
Université Paris-Saclay, CNRS, Orsay, France
E-mail: thomas.nowak@lri.fr

Q. Soubeyran
Université Paris-Saclay, CNRS, Orsay, France
École polytechnique, Institut Polytechnique de Paris, Palaiseau, France
E-mail: quentin.soubeyran.16@polytechnique.org

## 1 Introduction

In the past few decades, synthetic biology has laid considerable focus on the re-programming of cells as computing machines. They have been engineered to sense a range of inputs (metabolites [49], light [53], oxygen [3], pH [47]) and process them to produce desired outputs according to defined processing codes (primarily digital [37], but occasionally analog [17]). Some potential applications of the cellular machines include production of metabolic compounds of interest [41], bioremediation of toxic environments [55], sensing of disease bio-markers [49], and therapeutic intervention by

targeted effector delivery [3]. Yet, the ability of single cells to reliably process multiple inputs is acutely constrained by their limited resources.

Adding too many processes into one cell leads to resource-stress and eventually the code is lost due to mutation, a baseline error mechanism present in all living systems. This has, in part, encouraged the notion of distributing the computational tasks across multiple cells [43,54], to reduce resource-stress and improve robustness. The value of the idea is corroborated by the success of the division of labor seen in multi-cellular organisms that have naturally evolved from their unicellular ancestors [30,42]. While task-distribution in cell populations solves some problems, it immediately leads to other challenges that must be tackled for the successful implementation of any complex distributed program. Some of these challenges include: the orthogonality/specificity of communication signals, the rate and bandwidth of communication channels, cellular growth and its effect on signal amplification or dissipation, and effect of cross-talk between different signals.

In this work we focus on signal amplification and Boolean function computation in distributed systems whose agents are duplicating bacteria. A central problem in this setting is to maintain a consistent state of circuit values among the bacteria. The problem of maintaining a consistent state among agents has been studied in distributed computing for decades in different contexts [33], e.g., for replicated state machines [48] and mobile networks [7]. Starting from a mathematical computing model, analysis of a system's behavior has led to correctness proofs and performance bounds of proposed solutions, also shedding light on how protocol parameters influence the quality of the outcome. In distributed systems with biological agents, the cellular behavior is usually expressed in the language of chemical reaction networks (CRNs). A CRN is defined by a set of reactions, each consuming members of one or several species and producing members of others at a given rate.

The two most commonly used kinetics for CRNs are deterministic and stochastic approaches. The deterministic approach models the kinetics of a CRN as systems of ordinary differential equations (ODEs) with continuous real-valued concentrations of each species, whereas the stochastic approach models the CRN as a continuous-time Markov chain with discrete integer-valued counts of each species. While ODE modeling can capture important behavioral characteristics, in particular expected-value large-population limits, some phenomena can only be explained by stochastic-process kinetics. In particular, ODE kinetics cannot elucidate the probability of certain population-level events occurring

in a system of two competing species, e.g., the extinction of one species due to a series of random events. The stochastic-process kinetics of CRNs are much more common in distributed computing, in particular in population protocols [5], where reactions are restricted to two reactants and two products with constant-size populations and identical reaction rate constants equal to 1. They are also used in computability results in general CRNs [50] where non-constant population sizes are exploited. A model that allows non-constant population sizes via "split" and "die" reactions, but otherwise restricts reactions as in population protocols, was studied by Goldwasser *et al.* [24]. The presented algorithm with the goal to maintain a stable population size, however, uses leader election and synchronized phases via non-constant states per agent, rendering it impractical for implementations in bacterial cultures.

Computation of Boolean predicates has been extensively studied both in CRNs and population protocols. Early work on Boolean computation in CRNs is by Magnasco [34]. Signal values are encoded with low and high concentrations of corresponding species. Chen *et al.* [15] generalized computation in CRNs to functions on natural numbers. Arguments are given in terms of input species counts and the computed value is encoded in the number of output species counts. The output in population protocols is typically provided by all agents reaching consensus on a common output state [6]. These works cited above, however, do not include obligatory duplication reactions as we do.

*Consistent cell states by competition among cells.* Birth-death processes track species counts within a population with "birth" and "death" events over time. For each such population state there are transitions that move from one population state to the other with respect to "birth" and "death" events. Birth-death processes have been used to model competition, predation, or infection in evolutionary biology, ecology, genetics, and queueing theory [39,46].

Competition among species naturally lends itself to solving consensus-type problems. Angluin *et al.* [5] analyzed a population protocol with three states: $A$, $B$, and blank $N$. Encounters of opposing species $A$ and $B$ lead to one of them becoming blank via the rules $(A, B) \to (A, N)$ and $(B, A) \to (B, N)$, and blank species that encounter a non-blank species copy their state via rules $(A, N) \to (A, A)$, $(N, A) \to (A, A)$, $(B, N) \to (B, B)$, and $(N, B) \to (B, B)$. The latter can be viewed as duplication reactions for $A$ and $B$. By contrast to our model, with a constant birth rate per cell and potentially unbounded growth, the model by Angluin *et al.*

implies bounded population sizes and varying growth rates per cell.

The population protocol by Angluin *et al.* [4] alternates phases of state duplication and cancellation, separated by a clock signal generated by a dedicated leader species. These protocols, however, rely on non-varying populations sizes and the latter on a dedicated leader, and are thus impractical for implementations in bacteria. More recent works have developed leaderless phase clocks in population protocols (e.g., [1]), which can be used to solve the majority problem [10,11,8,19] in population protocols.

An early mention of problems requiring a stochastic analysis of two competing species is by Volterra [56] and Feller [21] although only the growth of a single species is analyzed therein. For an overview of single species birth-death Markov chains, see, e.g., [13]. Extensions for multiples species, with applications to genetic mutations, are found in the literature on competition and branching processes [44,12,29]. For example, Ridler-Rowe [45] considers a stochastic process between two competing species. However, the process in that work differs from ours in that death reactions therein are $A + B \longrightarrow A$ and $A + B \longrightarrow B$, leaving a winner after an encounter between two competing individuals. The paper presents an approximation for long-term distributions and bounds the probability that starting from initial $A, B$ sizes, species $A$ goes extinct. However, the analysis is for initial population sizes approaching infinity, only, and assumes an initial gap between species counts that is linear in the population size. By contrast our analysis holds for finite population sizes $n$, and requires a gap of $\Omega(\sqrt{n \log n})$, only. A complementary approach for the same asymmetric process proposed in [25] is to numerically solve a finite size cut-off of the infinite linear equation systems.

*Computation in birth systems.* In this work, we introduce and study *protocols for birth systems* where all species inherently duplicate. To simplify our model, we do not consider death reactions as they occur at far lower rates than duplication reactions in typical bacterial colonies. Such protocols are thus different from population protocols, which have population sizes that remain constant over the course of an execution. Further, our protocols do not rely on exact species counts, they are not leader-based, and they require small and constant state space per cell, lending themselves readily for future biological implementation.

For simplicity we assume that all duplication reactions of our birth systems have the same rate. We leave the question of natural selection due to differing growth rates to future work. In particular, we study two protocols within birth systems.

(i) We introduce the Pairwise Annihilation protocol for two species $A$ and $B$ and show that it solves majority consensus with high probability: If the initial difference $\Delta$ between sizes $A$ and $B$ grows weakly with the population size $n$ according to $\Delta = \Omega(\sqrt{n \log n})$, then the protocol identifies the initial majority with high probability. Since it amplifies the difference between the two species, we also refer to the Pairwise Annihilation protocol as an *amplifier*. Further, we will show that the protocol reaches consensus in expected constant time. The protocol's reactions are deceptively simple. Besides the obligatory birth reactions $A \longrightarrow 2A$ and $B \longrightarrow 2B$, it comprises a single death reaction $A + B \longrightarrow \emptyset$.

(ii) We demonstrate how to implement the components of *feed-forward Boolean circuits*. Each Boolean gate in our implementation is a NAND gate, followed by an amplifier. Note that while we focus on the universal NAND gate for the sake of a lighter notation, our construction and its analysis holds for any arbitrary two-input Boolean function. The latter will be important for optimization and follow-up with biological implementations. Signals between the NAND gates are encoded using two species each, the difference of which determines whether a signal is a logical 0, 1, or neither. A NAND gate is a protocol that maps two input signals $X$ and $Y$ to an output signal $Z$ that is the logical NAND of $X$ and $Y$.

While NAND gates are used to implement the circuit's Boolean behavior, the successive amplifiers regenerate the gate's output signal by amplifying the difference between the two output signal species. Repeated, successive invocation of the NAND protocol followed by the amplifier protocol for time $O(\log n)$, where $n$ is the total initial population, can finally be used to compute the circuit's output values layer by layer.

*Organization.* The rest of the paper is organized as follows: In Section 2, we define the computational model. In Section 3, we introduce and analyze our protocol for majority consensus, both analytically and via simulations. In Section 4, we define and analyze the NAND gate protocol. In Section 5, we present simulations of a biologically plausible implementation of the NAND gate with amplifiers. In Section 6, we consider the context of continuous cultures and suggest that majority consensus can become slower in these cultures. Finally, Section 7 concludes the paper.

## 2 Model

We write $\mathbb{N} = \{0, 1, \dots\}$, $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$, and $\mathbb{R}_0^+ = [0, \infty)$. When analyzing our protocols, we employ the term "with high probability" relative to the total initial population. That is, event $E$ happens with high probability if there exists some $c > 0$ such that $\mathbb{P}(E) = 1 - O(1/n^c)$, where $n$ is the total initial population.

### 2.1 Chemical Reaction Networks

We use the standard stochastic kinetics for chemical reaction networks. A reader familiar with the model can safely skip this subsection.

A *chemical reaction network* is described by a set $\mathcal{S}$ of species and a set of reactions. A *reaction* is a triple $(\mathbf{r}, \mathbf{p}, \alpha)$ where $\mathbf{r}, \mathbf{p} \in \mathbb{N}^{\mathcal{S}}$ and $\alpha \in \mathbb{R}_0^+$. The species with positive count in $\mathbf{r}$ are called the reaction's *reactants* and this with positive count in $\mathbf{p}$ are called its *products*. The parameter $\alpha$ is called the reaction's *rate constant*. A *configuration* of a CRN is simply an element of $\mathbb{N}^{\mathcal{S}}$. A reaction $(\mathbf{r}, \mathbf{p}, \alpha)$ is *applicable* to configuration $\mathbf{c}$ if $\mathbf{r}(S) \leq \mathbf{c}(S)$ for all $S \in \mathcal{S}$.

We write $\mathbf{r} \xrightarrow{\alpha} \mathbf{p}$ to denote a reaction $(\mathbf{r}, \mathbf{p}, \alpha)$. For instance, the reaction $(\{A, B\}, \{2B, C\}, \alpha)$ will simply be denoted $A + B \xrightarrow{\alpha} 2B + C$. Here, we used the shorthand notations $\{A, B\}$ and $\{2B, C\}$ for functions $\mathcal{S} \to \mathbb{N}$. For instance, the notation $\{2B, C\}$ represents the function $\mathbf{p} : \mathcal{S} \to \mathbb{N}$ defined by $\mathbf{p}(B) = 2$, $\mathbf{p}(C) = 1$, and $\mathbf{p}(S) = 0$ for all other species $S \notin \{B, C\}$.

The *stochastic kinetics* of a CRN are a continuous-time Markov chain (see a textbook [13] for auxiliary definitions). Given some volume $v \in \mathbb{R}_0^+$, which we will normalize to $v = 1$ for most of the paper, the propensity of a reaction $(\mathbf{r}, \mathbf{p}, \alpha)$ in configuration $\mathbf{c}$ is equal to $\frac{\alpha}{v} \prod_{S \in \mathcal{S}} \binom{\mathbf{c}(S)}{\mathbf{r}(S)}$, where $\binom{\mathbf{c}(S)}{\mathbf{r}(S)}$ denotes the binomial coefficient of $\mathbf{c}(S)$ and $\mathbf{r}(S)$. The binomial coefficient is 1 if $\mathbf{r}(S) = 0$, i.e., if the species $S$ is not a reactant of the reaction. It is 0 if $\mathbf{r}(S) > \mathbf{c}(S)$. The propensity of a non-applicable reaction is thus 0. For example, the propensity of reaction $A + B \xrightarrow{\alpha} 2B + C$ in configuration $\mathbf{c}$ is equal to $\frac{\alpha}{v} \cdot \mathbf{c}(A) \cdot \mathbf{c}(B)$. The propensity of $A \xrightarrow{\gamma} 2A$ is equal to $\frac{\gamma}{v} \cdot \mathbf{c}(A)$. The new configuration after an applicable reaction is equal to $\mathbf{c}' = \mathbf{c} - \mathbf{r} + \mathbf{p}$.

We will use the notation $Q(x, y)$ for the propensity of the transition from state $x$ to state $y$ in a continuous-time Markov chain. To each continuous-time Markov chain corresponds a discrete-time Markov chain that only keeps track of the sequence of state changes, but not of their timing. We use $P(x, y)$ for the transition probability from state $x$ to state $y$ in the discrete-time

chain. We have the formula

$$P(x, y) = Q(x, y) / \sum_z Q(x, z) \ .$$

### 2.2 Birth Systems

A *protocol for a birth system*, or protocol, with input species $\mathcal{I}$ and output species $\mathcal{O}$, for finite, not necessarily disjoint, sets $\mathcal{I}$ and $\mathcal{O}$ is a CRN specified as follows. Its set of species $\mathcal{S}$ comprises input/output species $\mathcal{I} \cup \mathcal{O}$ and a finite set of internal species $\mathcal{L}$. Further, the protocol defines the initial species counts $X_0$ for internal and output species $X \in \mathcal{L} \cup \mathcal{O}$ and a finite set of reactions $\mathcal{R}$ on the species in $\mathcal{S}$. For each species $X \in \mathcal{S}$, there is a duplication reaction of the form $X \xrightarrow{\gamma} 2X$. All duplication reactions have the same rate constant $\gamma > 0$.

Given a protocol and an initial species count for its inputs, an execution of the protocol is given by the stochastic process of the CRN with species $\mathcal{S}$, reactions $\mathcal{R}$, and respective initial species counts.

## 3 Majority Consensus

The *Pairwise Annihilation protocol* is defined for two species, $A$ and $B$, both of which are inputs and outputs. It contains, apart from the obligatory duplication reactions, the single reaction of $A$ and $B$ eliminating each other with rate constant $\delta > 0$. The complete list of reactions of the Pairwise Annihilation protocol is thus:

$$A \xrightarrow{\gamma} 2A$$
$$B \xrightarrow{\gamma} 2B$$
$$A + B \xrightarrow{\delta} \emptyset$$
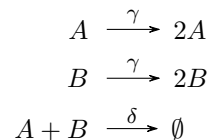
We say that *consensus* is reached if one of the two species becomes extinct. If the initial population counts differ, we say that *majority consensus* is reached if consensus is reached and the species that was initially in majority is not extinct. If the initial counts of both species are equal, then majority consensus is reached when one species is extinct and the other is not.

We show that the Pairwise Annihilation protocol reaches consensus in constant time and majority consensus with high probability.

**Theorem 1** *For initial population $n = A(0) + B(0)$ and initial gap $\Delta = |A(0) - B(0)|$, the Pairwise Annihilation protocol reaches consensus in expected time $O(1)$ and in time $O(\log n)$ with high probability. It reaches majority consensus with probability $1 - e^{-\Omega(\Delta^2/n)}$.*

4

From Theorem 1 we immediately obtain a bound on the initial gap sufficient for majority consensus with high probability.

**Corollary 2** *For an initial population $n$ and an initial gap $\Delta$, if $\Delta = \Omega\left(\sqrt{n \log n}\right)$, then the Pairwise Annihilation protocol reaches majority consensus with high probability.*

Without duplication reactions, it is obvious that the Pairwise Annihilation protocol reaches consensus and that majority consensus is always reached if the two species have different initial population counts. We are thus not only able to show that we can achieve majority consensus in spite of continual population growth via duplication reactions of all species, but also that a sub-linear gap in the initial population counts suffices. The required initial gap of $\Omega(\sqrt{n \log n})$ matches that of the best protocols without obligatory duplications [5,16,2]. The time complexity of our protocol is different to population protocols, however: in the $O(n)$ expected transitions that it takes the Pairwise Annihilation protocol to achieve consensus, asymptotically almost surely, there are agents that never interacted once in a population protocol of the same size. This is not a contradiction: In contrast to population protocols, the population size decreases during the initial stages of the Pairwise Annihilation protocol.

We will prove Theorem 1 in the following sections; first the time upper bound, then correctness with high probability.

### 3.1 Markov-Chain Model

The evolution of the Pairwise Annihilation protocol is described by a continuous-time Markov chain with state space $S = \mathbb{N}^2$. Its state-transition rates are:

$$Q\big((A,B)\,,\,(A+1,B)\big) = \gamma A$$
$$Q\big((A,B)\,,\,(A,B+1)\big) = \gamma B$$
$$Q\big((A,B)\,,\,(A-1,B-1)\big) = \delta AB$$

Note that the death transition $(A,B) \to (A-1,B-1)$ has rate zero if $A = 0$ or $B = 0$. Both axes $\{0\} \times \mathbb{N}$ and $\mathbb{N} \times \{0\}$ are absorbing, and so is the state $(A,B) = (0,0)$. This chain is regular, i.e., its sequence of transition times is unbounded with probability 1. Indeed, as we will show, the discrete-time chain reaches consensus with probability 1, from which time on the chain is equal to a linear pure-birth process, which is regular.

The corresponding discrete-time jump chain has the same state space $S = \mathbb{N}^2$ and the state-transition prob-

abilities

$$P\big((A,B)\,,\,(A+1,B)\big) = \frac{\gamma A}{\gamma(A+B) + \delta AB}$$
$$P\big((A,B)\,,\,(A,B+1)\big) = \frac{\gamma B}{\gamma(A+B) + \delta AB}$$
$$P\big((A,B)\,,\,(A-1,B-1)\big) = \frac{\delta AB}{\gamma(A+B) + \delta AB}$$

if $A > 0$ or $B > 0$. The axes as well as state $(A,B) = (0,0)$ is absorbing, as in the continuous-time chain.

As a convention, we will write $X(t)$ for the state of the continuous-time process $X$ at time $t$, and $X_k$ for the state of the discrete-time jump process after $k$ state transitions. The time to reach consensus is the earliest time $T$ such that $A(T) = 0$ or $B(T) = 0$.

### 3.2 Time to Reach Consensus

In this section we prove the first part of Theorem 1, i.e., the bounds on the time to reach consensus, both in expected time and with high probability. For that, we will employ a coupling of the Pairwise Annihilation protocol Markov chain with a single-species birth-death process. We show that the Pairwise Annihilation protocol reaches consensus when the single-species process reaches its extinction state and then bound this time in the single-species process. Fig. 1 visualizes the idea.

We denote the single-species process by $M(t)$. It is a birth-death chain with state space $S = \mathbb{N}$ and transition rates $Q(M, M+1) = \gamma M$ and $Q(M, M-1) = \delta M^2$. State 0 is absorbing. Note that the death rate $\delta M^2$ depends quadratically on the current population $M$, and not linearly like the birth rate $\gamma M$. The reason is that we want $M(t)$ to bound the minimum of the populations $A(t)$ and $B(t)$ and that the death transition in the Pairwise Annihilation protocol is quadratic in this minimum.

We will crucially use the fact that $\mathbb{P}\big(M(t) = 0\big) \leq \mathbb{P}\big(A(t) = 0 \vee B(t) = 0\big)$ for all times $t$. This, together with a bound on the time until $M(t) = 0$, then gives a bound on the time until consensus in the Pairwise Annihilation protocol chain.

*Continuous-time coupling.* The coupling is defined as follows. For sequences $(\xi_k)_{k \geq 1}$ of i.i.d. (independent and identically distributed) uniform random variables in the unit interval $[0, 1)$ and $(\eta_k)_{k \geq 1}$ of i.i.d. exponential random variables with normalized rate 1, we define the coupled process $(A(t), B(t), M(t))$ as follows. Initially, $M(0) = \min\{A(0), B(0)\}$. For $k \geq 0$, the $(k+1)^{\text{th}}$ transition happens after time $\eta_k / \Lambda(A_k, B_k, M_k)$ where $\Lambda(A, B, M) = \max\{\lambda(A, B), \lambda(M)\}$ is the maximum of
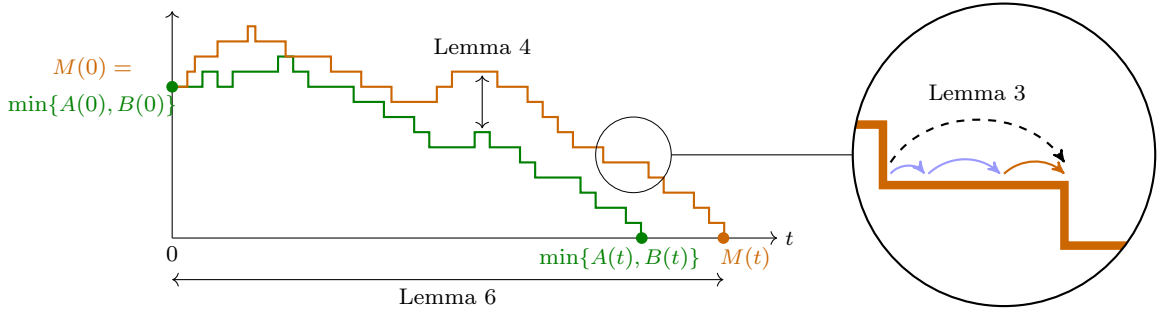
**Fig. 1** Idea of the proof: Construction of a continuous-time coupling of the Pairwise Annihilation protocol and the single species birth-death M chain. Stuttering steps are mapped to effective steps (Lemma 3). An execution of the coupling process fulfills the deterministic guarantee $\min\{A(t), B(t)\} \le M(t)$ for all times $t \ge 0$ (Lemma 4). From the coupling it follows that $\mathbb{P}\big(M(t) = 0\big) \le \mathbb{P}\big(A(t) = 0 \vee B(t) = 0\big)$ for the uncoupled processes (Corollary 5). The time until consensus then follows from the time until extinction in the M chain (Lemma 6).

the sums of transition rates of the individual chains in states $(A, B)$ and $M$, respectively, i.e., $\lambda(A, B) = \gamma(A + B) + \delta AB$ and $\lambda(M) = \gamma M + \delta M^2$. The new state $(A_{k+1}, B_{k+1}, M_{k+1})$ of the coupled chain is then determined by the following update rules. The state $(0, 0, 0)$ is absorbing. Otherwise, if $A_k \le B_k$, then: $(A_{k+1}, B_{k+1}) =$

$$
\begin{cases}
(A_k + 1, B_k) \\
\quad \text{if } \xi_{k+1} \in \left[0, \dfrac{\gamma A_k}{\Lambda(A_k, B_k, M_k)}\right) \\
(A_k, B_k + 1) \\
\quad \text{if } \xi_{k+1} \in \left[\dfrac{\gamma A_k}{\Lambda(A_k, B_k, M_k)}, \dfrac{\gamma A_k + \gamma B_k}{\Lambda(A_k, B_k, M_k)}\right) \\
(A_k - 1, B_k - 1) \\
\quad \text{if } \xi_{k+1} \in \left[1 - \dfrac{\delta A_k B_k}{\Lambda(A_k, B_k, M_k)}, 1\right) \\
(A_k, B_k) \quad \text{otherwise}
\end{cases}
\tag{1}
$$

If $A_k > B_k$ then the roles of $A_k$ and $B_k$ in (1) are exchanged. The update rule for $M_{k+1}$ is:

$$
M_{k+1} =
\begin{cases}
M_k + 1 & \text{if } \xi_{k+1} \in \left[0, \frac{\gamma M_k}{\Lambda(A_k, B_k, M_k)}\right) \\
M_k - 1 & \text{if } \xi_{k+1} \in \left[1 - \frac{\delta M_k^2}{\Lambda(A_k, B_k, M_k)}, 1\right) \\
M_k & \text{otherwise}
\end{cases}
\tag{2}
$$

*Analysis for time until consensus.* Note that in the coupling "stuttering steps" for $(A_k, B_k)$ or $M_k$ are possible in the definition of the coupled process, making the underlying discrete-time jump chains of, e.g., chain $(A(t), B(t))$ and the Pairwise Annihilation protocol, potentially differ.
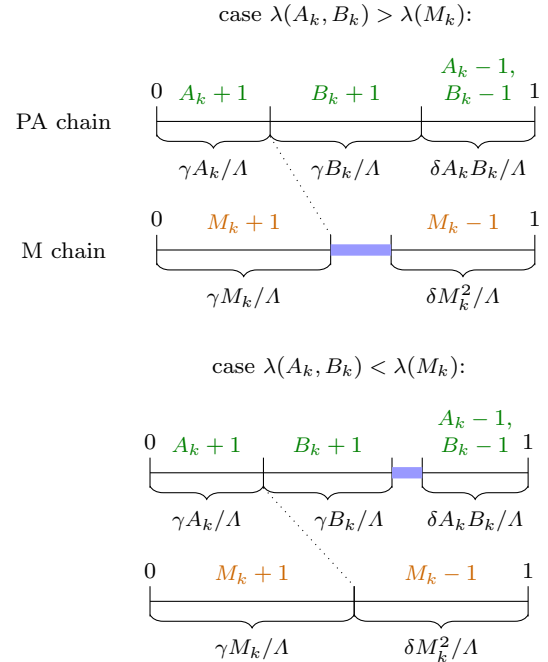


**Fig. 2** Continuous-time coupling of the Pairwise Annihilation (PA) chain and the single-species birth-death M-chain, given that $A_k \le B_k$, with $\Lambda = \Lambda(A_k, B_k, M_k)$. The intervals for the cases of $\xi_{k+1}$ and their effect on the Pairwise Annihilation chain and the M-chain are shown in green and orange, respectively. Cases that lead to stuttering steps are shown in blue. The dotted relation between intervals is proven in Lemma 4.

Indeed, the event $(A_{k+1}, B_{k+1}) = (A_k, B_k)$ is possible with positive probability if $\lambda(A_k, B_k) < \lambda(M_k)$, and $M_{k+1} = M_k$ has positive probability if $\lambda(M_k) < \lambda(A_k, B_k)$; see Fig. 2. The following Lemma 3, however, shows that the continuous-time chain $(A(t), B(t))$ and the Pairwise Annihilation protocol chain have identical transition rates, and are thus identically distributed. Its proof is folklore and given in the appendix.

**Lemma 3** *Let $T_1, T_2, \ldots$ be a sequence of i.i.d. exponential random variables with rate parameter $\lambda$ and let $k$ be an independent geometric random variable with success probability $p$. Then $T = T_1 + \cdots + T_k$ is exponentially distributed with rate parameter $p\lambda$.*

By construction of the coupled process, the single-species birth-death process $M(t)$ indeed dominates the minimum of the population counts $A(t)$ and $B(t)$ in the following way:

**Lemma 4** *In the coupled process, $\min\{A(t), B(t)\} \leq M(t)$ for all times $t \geq 0$.*

*Proof* Let $K$ be the step number of the discrete-time coupled process such that $t_K \leq t < t_{K+1}$, where $t_k$ is the time of the $k^{\text{th}}$ step. We show by induction that $\min\{A_k, B_k\} \leq M_k$ for all $k \in \mathbb{N}$. The inequality holds initially, for $k = 0$, by definition of the coupled process. Now assume that $\min\{A_k, B_k\} \leq M_k$. Without loss of generality, by symmetry, assume that $A_k \leq B_k$, so that $A_k = \min\{A_k, B_k\} \leq M_k$. Then $\gamma A_k \leq \gamma M_k$ and thus $A_{k+1} = A_k + 1$ implies $M_{k+1} = M_k + 1$ by the definition of the coupling in (1) and (2); see Fig. 2. We distinguish the two cases $A_k < M_k$ and $A_k = M_k$.

If $A_k < M_k$, then the only way to have $A_{k+1} > M_{k+1}$ is to have $A_{k+1} = A_k + 1$ and $M_{k+1} = M_k - 1$. But this is impossible since $A_{k+1} = A_k + 1$ implies $M_{k+1} = M_k + 1$.

Otherwise, $A_k = M_k$. The case is shown in Fig. 3. We have, $\delta M_k^2 = \delta A_k^2 \leq \delta A_k B_k$. Thus, as easily verified by the alignment of the intervals in Fig. 3, $M_{k+1} = M_k - 1$ implies $A_{k+1} = A_k - 1$ and $B_{k+1} = B_k - 1$. Hence, combined with the above implication which remains true, we have $A_{k+1} \leq M_{k+1}$ in all possible cases for $\xi_{k+1}$.

since $\lambda(A_k, B_k) > \lambda(M_k)$:
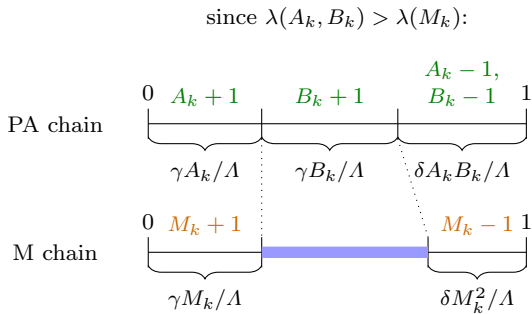
**Fig. 3** The case $A_k = M_k$ in the proof of Lemma 4, with $\Lambda = \Lambda(A_k, B_k, M_k)$. The case's assumption implies that $\lambda(A_k, B_k) > \lambda(M_k)$. The dotted relation between intervals is shown in the proof.

Lemma 4 allows to compare the probabilities of extinction in the single-species chain and of consensus in the Pairwise Annihilation protocol chain:

**Corollary 5** $\mathbb{P}(M(t) = 0) \leq \mathbb{P}(A(t) = 0 \vee B(t) = 0)$ *for all times $t \geq 0$.*

It thus suffices to prove bounds on the time until the population goes extinct in the single-species M chain. For that, we leverage known results on birth-death processes, which are not applicable to the two-species Pairwise Annihilation protocol chain.

**Lemma 6** *If $T$ denotes the time until extinction in the single-species process $M(t)$, then $\mathbb{E}\,T = O(1)$.*

*Proof* The birth rate in state $M(t) = i$ is equal to $\alpha(i) = i\gamma$ and the death rate is equal to $\beta(i) = i^2\delta$. From known general results on birth-death process (Theorem 25 in the appendix) we obtain, when starting from initial population $M(0) = M$, that

$$\mathbb{E}\,T = \sum_{j=1}^{M} \sum_{k=j-1}^{\infty} \frac{\alpha(j) \cdots \alpha(k)}{\beta(j) \cdots \beta(k)} \cdot \frac{1}{\beta(k+1)}$$
$$= \sum_{j=1}^{M} \sum_{k=j-1}^{\infty} \frac{\gamma^{k-j+1}}{\delta^{k-j+1} k!/(j-1)!} \cdot \frac{1}{(k+1)^2 \delta}$$

Setting $\alpha = \gamma/\delta$, we have

$$\mathbb{E}\,T = \frac{1}{\delta} \sum_{j=1}^{M} \sum_{k=j-1}^{\infty} \alpha^{k-j+1} \frac{(j-1)!}{(k+1)!(k+1)}$$
$$= \frac{1}{\delta} \sum_{j=1}^{M} \frac{(j-1)!}{\alpha^j} \sum_{k=j}^{\infty} \frac{\alpha^k}{k!k}$$
$$= \frac{1}{\delta} \sum_{j=1}^{M} \frac{(j-1)!}{\alpha^j} \cdot \frac{\alpha^j}{j!j} \sum_{k=j}^{\infty} \frac{\alpha^{k-j}}{k!/j! \cdot k/j}$$
$$\leq \frac{1}{\delta} \sum_{j=1}^{M} \frac{(j-1)!}{\alpha^j} \cdot \frac{\alpha^j}{j!j} \sum_{k=j}^{\infty} \frac{\alpha^{k-j}}{(k-j)!}$$

since for $k \geq j \geq 1$, it is $k!/j! \geq (k-j)!$ and $k/j \geq 1$. Thus,

$$\mathbb{E}\,T \leq \frac{1}{\delta} \sum_{j=1}^{M} \frac{(j-1)!}{\alpha^j} \cdot \frac{\alpha^j}{j!j} \cdot e^{\alpha}$$
$$= \frac{e^{\alpha}}{\delta} \sum_{j=1}^{M} \frac{1}{j^2}$$
$$\leq \frac{e^{\alpha} \pi^2}{6\delta} = O(1) \ .$$

This concludes the proof.

Denoting with $T_{AB}$ the earliest time $t$ such that $A(t) = 0$ or $B(t) = 0$, and with $T_M$ the earliest time $t$ such that $M(t) = 0$, Corollary 5 is equivalent to the inequality $\mathbb{P}(T_M \leq t) \leq \mathbb{P}(T_{AB} \leq t)$, which, in turn, is

equivalent to the inequality $\mathbb{P}(T_M > t) \geq \mathbb{P}(T_{AB} > t)$. Using the formula $\mathbb{E}\,T = \int_0^\infty \mathbb{P}(T > t)\,dt$, we further have

$$\mathbb{E}\,T_M = \int_0^\infty \mathbb{P}(T_M > t)\,dt$$
$$\geq \int_0^\infty \mathbb{P}(T_{AB} > t)\,dt = \mathbb{E}\,T_{AB}\ .$$

Combining this with Lemma 6, shows that the expected time until consensus in the Pairwise Annihilation protocol is also $O(1)$. For the high-probability result in the first part of Theorem 1, we simply make $\Theta(\log n)$ consecutive tries to achieve extinction in an interval of constant time:

**Lemma 7** *If $T$ denotes the time until extinction in the singles-species process $M(t)$, then there exists a constant $C$ such that $\mathbb{P}(T \leq C \log_2 n) = 1 - O(1/n)$.*

*Proof* Let $C_1$ be the $O(1)$ constant from Lemma 6 and set $C = \max\{2C_1, 2\}$. Then, by Markov's inequality, we have $\mathbb{P}(T > C) \leq C_1/C \leq 1/2$. Thus, the probability of the event $T > C \log_2 n$ is dominated by the probability of failing $\log_2 n$ consecutive tries with a Bernoulli random variable with parameter $p = 1/2$. But this probability is $2^{-\log_2 n} = 1/n$.

A simple combination of Corollary 5 and 7 completes the proof of the first part of Theorem 1.

### 3.3 Probability of Reaching Majority Consensus

We now turn to the proof of the second part of Theorem 1, i.e., the bound on the probability to achieve majority consensus. We use a coupling of the Pairwise Annihilation protocol chain with a different process than for the time bound. Namely we couple it with two parallel independent Yule processes. A Yule process, also known as a pure birth process, has this single state-transition rule $X \to X + 1$ with linear transition rate $\gamma X$. Since we already showed the upper bound on the time until consensus, it suffices to look at the discrete-time jump process. In particular, the coupling we define is discrete-time.

*Discrete-time coupling.* For an i.i.d. sequence $(\xi_k)_{k \geq 1}$ of uniformly distributed random variables in the unit interval $[0, 1)$, we define the coupling as the process $(A_k, B_k, X_k, Y_k)$ with $A_0 = X_0$, $B_0 = Y_0$ such that $(A_{k+1}, B_{k+1})$ is equal to

– $(A_k - 1, B_k - 1)$ if $\xi_{k+1} < \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$
– $(A_k + 1, B_k)$ if $\xi_{k+1} \geq \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$ and $\xi_{k+1} < 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$

– $(A_k, B_k + 1)$ if $\xi_{k+1} \geq 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$

and $(X_{k+1}, Y_{k+1})$ is equal to

– $(X_k, Y_k)$ if $\xi_{k+1} < \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$
– $(X_k + 1, Y_k)$ if $\xi_{k+1} \geq \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$ and $\xi_{k+1} < 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$
– $(X_k, Y_k + 1)$ if $\xi_{k+1} \geq 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$

if $\max\{A_k, B_k\} > 0$ and $\max\{X_k, Y_k\} > 0$. Otherwise the process remains constant. Fig. 4 visualizes the construction.

*Analysis for probability of reaching majority consensus.* We start with two simple technical lemmas that we will use for the comparison of the coupled processes.

**Lemma 8** *Let $a, b, x, y \in \mathbb{R}_0^+$ with $\max\{a, b\} > 0$ and $\max\{x, y\} > 0$. Then $\frac{b}{a+b} \leq \frac{y}{x+y}$ if and only if $bx \leq ay$.*

*Proof* Multiplying both sides by $(a + b) \cdot (x + y)$, we see that the first inequality is equivalent to $bx + by \leq ay + by$, which is in turn equivalent to $bx \leq ay$.

**Lemma 9** *Let $a, b, x, y, m \in \mathbb{R}_0^+$ with $\max\{a, b\} > 0$, $\max\{x, y\} > 0$, and $x \geq y$. If $x \leq a + m$ and $y \geq b + m$, then $\frac{b}{a+b} \leq \frac{y}{x+y}$.*

*Proof* By Lemma 8 it suffices to prove $bx \leq ay$. From the inequality chain $a + m \geq x \geq y \geq b + m$ we get $a \geq b$. We thus have $bx \leq b(a + m) = ab + bm \leq ab + am = a(b + m) \leq ay$.

The crucial property of this coupling is that the initial minority in the Pairwise Annihilation process cannot overtake the initial majority before the initial minority overtakes the initial majority in the parallel Yule processes. We now prove that our construction indeed has this property.

**Lemma 10** *If $X_0 = A_0 \geq B_0 = Y_0$ and $X_k \geq Y_k$ for all $0 \leq k \leq K$, then $X_k - Y_k \leq A_k - B_k$ for all $0 \leq k \leq K$.*

*Proof* We first show by induction on $k$ that $X_k \leq A_k + m_k$ and $Y_k \geq B_k + m_k$ for all $0 \leq k \leq K$, where $m_k$ is the number of death reactions up to step $k$. In the base case $k = 0$ we even have equality. For the induction step $k \mapsto k + 1$, we distinguish four cases; see Fig. 4.

1. The case $\xi_{k+1} < \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$: Then $m_{k+1} = m_k + 1$, $A_{k+1} = A_k - 1$, $B_{k+1} = B_k - 1$, $X_{k+1} = X_k$, and $Y_{k+1} = Y_k$. Hence, $X_{k+1} = X_k \leq A_k + m_k = (A_{k+1} + 1) + m_k = A_{k+1} + m_{k+1}$ and $Y_{k+1} = Y_k \geq B_k + m_k = (B_{k+1} + 1) + m_k = B_{k+1} + m_{k+1}$ by the induction hypothesis.
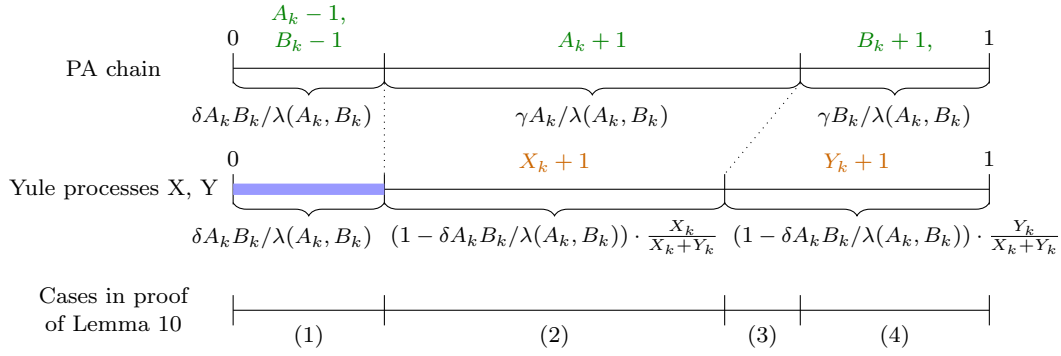
**Fig. 4** Discrete-time coupling of Pairwise Annihilation (PA) chain and two Yule processes X and Y with $\lambda(A_k, B_k) = \gamma(A_k + B_k) + \delta A_k B_k$. Cases for $\xi_{k+1}$ that lead to stuttering steps are shown in blue. The interval relations indicated by the dotted lines are proven by induction in Lemma 10. The four cases for the induction step are indicated.

2. The case $\xi_{k+1} \geq \frac{\delta A_k B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$ and $\xi_{k+1} < 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$: In particular we have

$$\xi_{k+1} \leq 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$$
$$\leq 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{B_k}{A_k + B_k}$$
$$= 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$$

by the induction hypothesis and Lemma 9. This implies the interval relation indicated in Fig. 4.
Hence, $m_{k+1} = m_k$, $A_{k+1} = A_k + 1$, $B_{k+1} = B_k$, $X_{k+1} = X_k + 1$, and $Y_{k+1} = Y_k$. But this means $X_{k+1} = X_k + 1 \leq A_k + m_k + 1 = (A_k + 1) + m_k = A_{k+1} + m_{k+1}$ and $Y_{k+1} = Y_k \geq B_k + m_k = B_{k+1} + m_{k+1}$ by the induction hypothesis.

3. The case $\xi_{k+1} \geq 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$ and $\xi_{k+1} < 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$: We have $m_{k+1} = m_k$, $A_{k+1} = A_k + 1$, $B_{k+1} = B_k$, $X_{k+1} = X_k$, and $Y_{k+1} = Y_k + 1$. But this means $X_{k+1} = X_k < X_k + 1 \leq A_k + m_k + 1 = (A_k + 1) + m_k = A_{k+1} + m_{k+1}$ and $Y_{k+1} = Y_k + 1 > Y_k \geq B_k + m_k = B_{k+1} + m_{k+1}$ by the induction hypothesis. $X_{k+1} = X_k < X_k + 1 \leq A_k + m_k + 1 = (A_k + 1) + m_k = A_{k+1} + m_{k+1}$ $Y_{k+1} = Y_k + 1 > Y_k \geq B_k + m_k = B_{k+1} + m_{k+1}$

4. The case $\xi_{k+1} \geq 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$: In particular we have

$$\xi_{k+1} \geq 1 - \frac{\gamma B_k}{\gamma(A_k + B_k) + \delta A_k B_k}$$
$$= 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{B_k}{A_k + B_k}$$
$$\geq 1 - \frac{\gamma(A_k + B_k)}{\gamma(A_k + B_k) + \delta A_k B_k} \cdot \frac{Y_k}{X_k + Y_k}$$

by the induction hypothesis and Lemma 9. Hence $m_{k+1} = m_k$, $A_{k+1} = A_k$, $B_{k+1} = B_k + 1$, $X_{k+1} = X_k$, and $Y_{k+1} = Y_k + 1$. But this means $X_{k+1} =$

$X_k \leq A_k + m_k = A_{k+1} + m_{k+1}$ and $Y_{k+1} = Y_k + 1 \geq B_k + m_k + 1 = (B_k + 1) + m_k = B_{k+1} + m_{k+1}$ by the induction hypothesis.
The lemma now follows via $X_k - Y_k \leq (A_k + m_k) - (B_k + m_k) = A_k - B_k$. $\blacksquare$

**Corollary 11** *If $A_0 = X_0$ and $B_0 = Y_0$, then*

$$\mathbb{P}(\exists k \colon A_k = B_k) \leq \mathbb{P}(\exists k \colon X_k = Y_k).$$

*Proof* By Lemma 10, if $k$ is minimal such that $A_k = B_k$, then $X_k = Y_k$. $\blacksquare$

As defined in the coupling the parallel Yule processes $(X_k, Y_k)$ can have stuttering steps where

$$(X_{k+1}, Y_{k+1}) = (X_k, Y_k).$$

However, this happens only finitely often almost surely. This allows us to analyze a version of the process $(X_k, Y_k)$ without stuttering steps in the rest of the proof.

**Lemma 12** *If $(\tilde{X}_k, \tilde{Y}_k)$ is the product of two independent pure-birth processes with $\tilde{X}_0 = X_0$ and $\tilde{Y}_0 = Y_0$, then $\mathbb{P}(\exists k \colon \tilde{X}_k = \tilde{Y}_k) = \mathbb{P}(\exists k \colon X_k = Y_k)$.*

*Proof* Lemma 6 implies that there are only finitely many deaths in the coupled chain almost surely. There are hence only finitely many stuttering steps in $(X_k, Y_k)$ almost surely. $\blacksquare$

Because of Lemma 12, by slight abuse of notation, we will use $(X_k, Y_k)$ to refer to the parallel Yule processes without any stuttering steps.

Two parallel independent Yule processes are known to be related to a beta distribution, which we will use below. The regularized incomplete beta function $I_z(\alpha, \beta)$ is defined as

$$I_z(\alpha, \beta) = \int_0^z t^{\alpha-1}(1-t)^{\beta-1}\, dt \Big/ \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt \ .$$

9

**Lemma 13** If $X_0 > Y_0$, then

$$\mathbb{P}\left(\exists k\colon X_k = Y_k\right) = 2 \cdot I_{1/2}(X_0, Y_0) \ .$$

*Proof* The sequence of ratios $\frac{X_k}{X_k+Y_k}$ converges with probability 1 and the limit is distributed according to a beta distribution with parameters $\alpha = X_0$ and $\beta = Y_0$ (Theorem 26 in the appendix). In particular, the probability that the limit is less than $1/2$ is equal to the beta distribution's cumulative distribution function evaluated at $1/2$, i.e., equal to $I_{1/2}(X_0, Y_0)$. Because initially we have $X_0 > Y_0$, the law of total probability gives:

$$\begin{aligned}
I_{1/2}(X_0, Y_0) &= \mathbb{P}\left(\lim_{k\to\infty} \frac{X_k}{X_k + Y_k} < \frac{1}{2}\right) \\
&= \mathbb{P}\left(\lim_{k\to\infty} \frac{X_k}{X_k + Y_k} < \frac{1}{2}\Big|\exists k\colon X_k = Y_k\right) \\
&\quad \cdot \mathbb{P}\left(\exists k\colon X_k = Y_k\right) \\
&\quad + \mathbb{P}\left(\lim_{k\to\infty} \frac{X_k}{X_k + Y_k} < \frac{1}{2} \land \forall k\colon X_k > Y_k\right)
\end{aligned} \quad (3)$$

Now, if $\forall k\colon X_k > Y_k$, then $\lim_k \frac{X_k}{X_k+Y_k} \geq 1/2$, which shows that the second term in the sum in (3) is zero. Further, under the condition $\exists k\colon X_k = Y_k$, it is equiprobable for the limit of $\frac{X_k}{X_k+Y_k}$ to be larger or smaller than $1/2$ by symmetry and the strong Markov property. This shows that the right-hand side of (3) is equal to $\frac{1}{2} \cdot \mathbb{P}\left(\exists k\colon X_k = Y_k\right)$, which concludes the proof.

We define the event "$B$ wins" as $A$ eventually becoming extinct. Then, we have:

**Lemma 14** If $A_0 > B_0$, then $\mathbb{P}\left(\exists k\colon A_k = B_k\right) = 2 \cdot \mathbb{P}(B \text{ wins})$.

*Proof* Similarly to the proof of Lemma 13, by the law of total probability, we have:

$$\begin{aligned}
\mathbb{P}\left(B \text{ wins}\right) &= \mathbb{P}\left(B \text{ wins} \mid \exists k\colon A_k = B_k\right) \cdot \\
&\quad \mathbb{P}\left(\exists k\colon A_k = B_k\right) + \mathbb{P}\left(B \text{ wins} \land \forall k\colon A_k > B_k\right)
\end{aligned} \quad (4)$$

If $\forall k\colon A_k > B_k$, then $B$ cannot win, i.e., the second term in the right-hand side of (4) is zero. Also, by symmetry and the strong Markov property, it is

$$\mathbb{P}\left(B \text{ wins} \mid \exists k\colon A_k = B_k\right) = 1/2 \ .$$

A simple algebraic manipulation now concludes the proof.

Combining the previous two lemmas with the coupling, we get an upper bound on the probability that the Pairwise Annihilation protocol fails to reach majority consensus. This upper bound is in terms of the regularized incomplete beta function.

**Lemma 15** If $A_0 \geq B_0$, then the Pairwise Annihilation protocol fails to reach majority consensus with probability at most $I_{1/2}(A_0, B_0)$.

*Proof* Setting $X_0 = A_0$ and $Y_0 = B_0$, and combining Corollary 11 and Lemmas 13 and 14, we get $\mathbb{P}(B \text{ wins}) = \frac{1}{2}\cdot\mathbb{P}(\exists k\colon A_k = B_k) \leq \frac{1}{2}\cdot\mathbb{P}(\exists k\colon X_k = Y_k) = I_{1/2}(A_0, B_0)$.

Due to Lemma 15, it only remains to upper-bound the term $I_{1/2}(\alpha, \beta)$. Lemma 16 provides such a bound. Its proof is given in the appendix.

**Lemma 16** For $m, \Delta \in \mathbb{N}$, it holds that

$$I_{1/2}(m + \Delta, m) = \exp\left(-\Omega\left(\frac{\Delta^2}{m}\right)\right) \ .$$

Combining the above lemmas proves the second part of Theorem 1.

3.4 Simulation of the Pairwise Annihilation Protocol

Stochastic simulations [22,26] of the Pairwise Annihilation protocol are shown in Fig. 5 for the probability that species $A$ survives, while species $B$ goes extinct. The birth and death rates, $\gamma$ and $\delta$, are both set to 1. The probability that the protocol converges on $A$ surviving and $B$ going extinct is primarily dependent on the difference in initial population size $A_0 - B_0$. Larger populations are only slightly less sensitive to the difference: Fig. 5 demonstrates that the total population size across two orders of magnitude has a small effect compared to the difference between species. Indeed, this behavior qualitatively matches the bound in Theorem 1 with $-\Omega(\Delta^2/n)$ in the exponent.

The dependence of expected convergence time for the Pairwise Annihilation protocol is explored using stochastic simulation over its reaction rate constants and initial conditions in Fig. 6. Exponential changes in rate constants yield exponential changes in convergence time. As expected, the convergence time is more strongly dependent on the death rate constant $\delta$ than the birth rate constant $\gamma$. Convergence time sharply increases if the initial concentrations of the two species $A$ and $B$ are closer to each other. The off-diagonal initial concentrations converge faster for larger population sizes since the absolute difference in concentrations is larger.

**4 Boolean Gates**

In terms of circuit design, the Pairwise Annihilation protocol can be viewed as a differential signal amplifier; see also the sharp S-shaped curve in Fig. 5 that is typical for an amplifier. Differential signaling has applications in systems that require high resilience to noise, and thus an application for our inherently growing systems is natural.
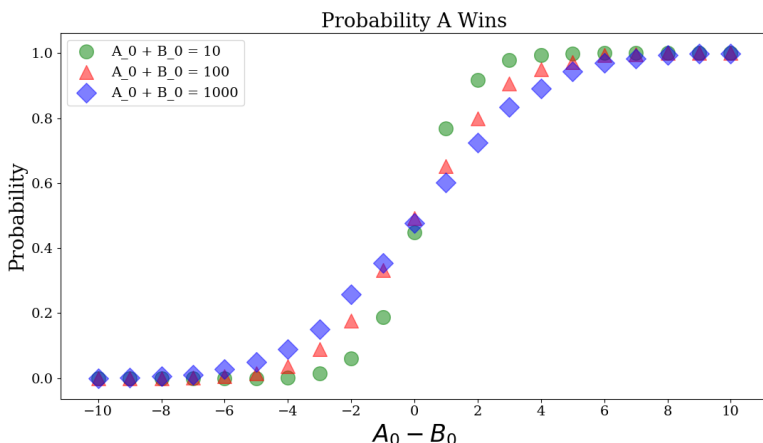
**Fig. 5** The probability that species $A$ survives while species $B$ goes extinct is sharply dependent on their initial difference in population count $A_0 - B_0$. The sharpness of the transition is inversely proportional to initial population size $A_0 + B_0$.
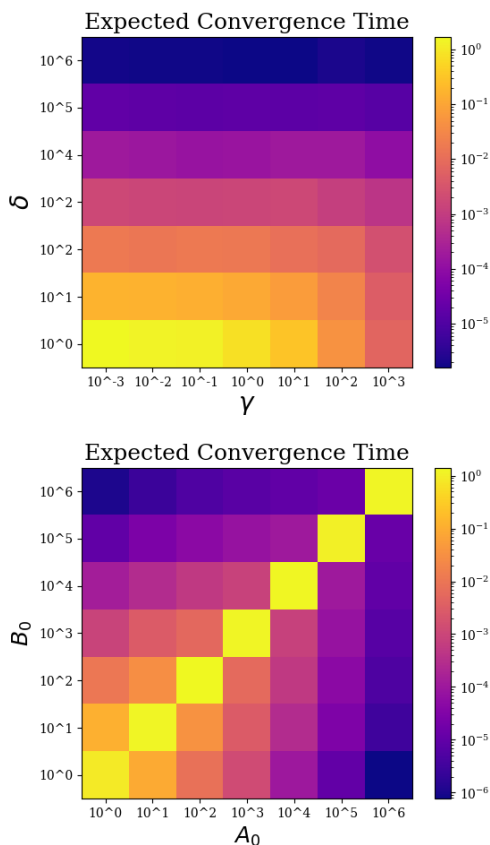




**Fig. 6** Log-scaled expected convergence time (in minutes) of the Pairwise Annihilation protocol is represented by color. Corresponding values are shown on the adjacent vertical bar. **Top:** birth rate coefficient $\gamma$ and death rate coefficient $\delta$ with $A_0 = B_0 = 100$. **Bottom:** initial populations sizes with $\gamma = 0.01$ and $\delta = 1$.

In this section we study a protocol for computing the logical NAND of two signals, despite a loss of signal quality at the output. The Pairwise Annihilation protocol is then applied to regenerate the signal, obtaining a clear 0 or 1 with high probability. Note that the NAND gate protocol is easily generalized to arbitrary two-input Boolean functions, and so is its analysis.

### 4.1 Dual-rail Signals

We start with some notation. A *signal* is from a finite alphabet $\Sigma = \{X, Y, \dots\}$. At each time $t \geq 0$, a signal $X \in \Sigma$ has a value $x(t) \in \{0, 1, \bot\}$. Following a technique from clockless circuit design [51, 38] we encode the value of a signal as a dual-rail signal in the following way. For each signal $X$, there are two species $X^0$ and $X^1$. Intuitively, for $v \in \{0, 1\}$, a large count of $X^v(t)$ and a low count of $X^{\neg v}(t)$ encodes for $x(t) = v$. In fact, we will ask for a minimum gap in species counts between $X^v(t)$ and $X^{\neg v}(t)$. If the signal is neither 0 nor 1, we will say that it has value $\bot$. We will make the assumptions on the input signals precise and discuss guarantees on output signals when specifying the gate input/output behavior.

Let $X^0, X^1$ be species of a dual-rail encoding of signal $X$. For convenience we write $X(t)$ for $X^0(t) + X^1(t)$. For $n, \Delta \in \mathbb{N}$, we say signal $X$ is *initially $(n, \Delta)$-correct with value* $x \in \{0, 1\}$ if

$$X(0) \geq n \quad \text{and} \quad X^{\neg x}(0) \leq \frac{n - \Delta}{2} \ . \tag{5}$$

The *initial gap* $X^x(0) - X^{\neg x}(0)$ of signal $X$ is thus bounded by

$$X^x(0) - X^{\neg x}(0) = X^x(0) + X^{\neg x}(0) - 2X^{\neg x}(0) \geq \Delta \ .$$

11

## 4.2 Dual-Rail NAND Gate

A dual-rail implementation of a NAND gate with input signals $A, B$ and output signal $Y$, the so called NAND *gate protocol*, is as a protocol with input species $\mathcal{I} = \{A^0, A^1, B^0, B^1\}$, output species $\mathcal{O} = \{Y^0, Y^1\}$, and no internal species. Initial counts for outputs that are not inputs are $Y^0(0) = Y^1(0) = 0$. Further, for all $a, b \in \{0, 1\}$ and $y = \neg(a \wedge b)$, the protocol contains a reaction

$$A^a + B^b \xrightarrow{\alpha} A^a + B^b + Y^y \ ,$$
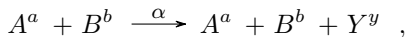
where $\alpha > 0$ is the gate's rate constant. Since all species are permanently replicating, we further have the obligatory duplication reactions $A^i \xrightarrow{\gamma} 2A^i$, $B^i \xrightarrow{\gamma} 2B^i$, and $Y^i \xrightarrow{\gamma} 2Y^i$ for $i \in \{0, 1\}$. Fig. 7 depicts the NAND gate with the subsequent amplification protocol.
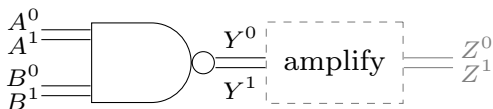


**Fig. 7** Dual-rail NAND gate with input signals $A$ and $B$ an output signal $Y$. Successive amplification of $Y$ to signal $Z$ shown in gray.

In Section 4.3 we will show that the NAND gate ensures the following input-output specification:

**Theorem 17** *Assume that the* NAND *gate's input signals $A, B$ are dual-rail encoded signals, and that they are initially $(n, \Delta)$-correct with values $a, b \in \{0, 1\}$, respectively, where*

*$n \in \mathbb{N}^+$ and*
*$\Delta \geq 0.62 \cdot \max\{A(0), B(0)\}$ .*

*Then with high probability, there exists some time $t = O(1)$ such that $Y(t) = n$ and $Y^y(t) - Y^{\neg y}(t) = \Omega(n)$ for the output signal $Y$ where $y = \neg(a \wedge b)$ is the correct* NAND *output based on the initial values $a, b$ of signals $A$ and $B$, respectively.*

## 4.3 Gate Correctness and Performance

To show the correct of the dual-rail two-input gate, we proceed as follows. Starting with technical lemmas, we first show that the initial value of a single dual-rail input signal is not lost due to growth of the input species (Lemma 20). Making use of independence of growth of the species encoding the rails of the two input signals, we then bound the probability that both input

signals remain correct (Lemma 22). We finally show that the gate's dual-rail output is correct in two steps: We first assume a simplified model where gate output species are produced by the gate but cannot duplicate (Lemma 23). The lemma bounds the number of incorrect output species that are produced by the gate, showing that its output signal has the correct value. In a second step, we prove that duplication of the output species, as required by our model, does not invalidate the output correct value (Lemma 24).

We now turn to the proof of Theorem 17. For our analysis we need a bound on the regularized incomplete beta function $I_{3/4}$. The proofs of the next two lemmas are given in the appendix.

**Lemma 18** *For $X \geq Y$, it holds that $I_{3/4}(X, Y) \leq \frac{1}{2} \exp\left(-\frac{(X-Y+1)^2}{4(Y-1)} + (X+Y-1)\log\frac{3}{2}\right)$. If $m, \Delta \geq 0$,*

$$I_{3/4}(m + \Delta, m) \leq \frac{1}{2} \exp\left(-\frac{(\Delta+1)^2}{4(m-1)} + \frac{2m+\Delta}{2}\right) \ .$$

The following lemma shows that for $z = 3/4$, the function $(x, y) \mapsto I_z(x, y)$ is non-decreasing in $(x, y)$ along the discretized line with slope $1/3$.

**Lemma 19** *If $X \geq 3Y \geq 0$, then*

$$I_{3/4}(X, Y) \leq I_{3/4}(X + 3, Y + 1).$$

We are now in the position to show a lower bound on the probability for a discrete time Yule process with two species $X$ and $Y$, that $\lim_{k \to \infty} X_k/(X_k + Y_k) < 3/4$, given that the initial values fulfill $X_0/(X_0 + Y_0) > 3/4$ and that there is a step $\ell$ with $X_\ell/(X_\ell + Y_\ell) \leq 3/4$.

**Lemma 20** *Let $X$ and $Y$ be species from a Yule process. Assume that $X_0/(X_0 + Y_0) > 3/4$ for the initial values. Then*

$$\mathbb{P}\left(\lim_{k \to \infty} \frac{X_k}{X_k + Y_k} < \frac{3}{4} \ \Big| \ \exists \ell : \frac{X_\ell}{X_\ell + Y_\ell} \leq \frac{3}{4}\right) \geq \omega(X_0, Y_0) \ ,$$

*where $\omega(X_0, Y_0)$ equals to*

$$\inf\left\{I_{3/4}(x, y) \ \big| \ x \geq X_0 \wedge y \geq Y_0 + 1 \right.$$
$$\left. \wedge x \in 3y - \{0, 1, 2\}\right\} \ .$$

*Moreover, $\omega(X_0, Y_0) > 0.444$*

*Proof* By assumption $X_0/(X_0 + Y_0) > 3/4$. Let $\ell \geq 1$ be minimal such that $X_\ell/(X_\ell + Y_\ell) \leq 3/4$. By assumption such an $\ell$ exists. By minimality of $\ell$, we have

$$X_\ell \leq 3Y_\ell \quad \text{and} \quad X_{\ell-1} > 3Y_{\ell-1} \ .$$

From the fact that $X, Y$ follow a Yule process, this can only be the case if $Y$ has increased from step $\ell - 1$ to $\ell$, i.e.,

$$X_\ell = X_{\ell-1} \geq X_0 \tag{6}$$
$$Y_\ell = Y_{\ell-1} + 1 \geq Y_0 + 1 \ . \tag{7}$$

Thus, $X_\ell > 3Y_\ell - 3$ from which $X_\ell \geq 3Y_\ell - 3$ and further,

$$X_\ell \in 3Y_\ell - \{0, 1, 2\} \ . \tag{8}$$

For a Yule process with species $X'$ and $Y'$, and arbitrary initial counts $X_0' = x$ and $Y_0' = y$, we have

$$\mathbb{P}\left(\lim_{k \to \infty} \frac{X_k'}{X_k' + Y_k'} < \frac{3}{4}\right) = I_{3/4}(x, y) \ . \tag{9}$$

The first inequality of the lemma now follows from (6), (7), (8), and (9).

We next show the second inequality of the lemma. For that purpose, we remark that any $(x, y)$ in

$$S = \{x \geq X_0 \wedge y \geq Y_0 + 1 \wedge x \in 3y - \{0, 1, 2\}\}$$

with $X_0 \geq 1$ and $Y_0 \geq 1$ is of the form

$$s_0 + m \cdot (3, 1) \ , \tag{10}$$

where $s_0 \in \{(4, 2), (5, 2), (6, 2)\}$ and $m \in \mathbb{N}$.

Assume $x = 3y - \Delta$ with $\Delta \in \{0, 1, 2\}$. Choosing $s_0 = (6 - \Delta, 2)$ and $m = y - 2 \geq 0$, and applying (10) yields

$$(6 - \Delta, 2) + (y - 2) \cdot (3, 1) = (3y - \Delta, y) = (x, y) \ ,$$

from which the claim follows.

By repeatedly applying Lemma 19 to an element $(x, y)$ in $S$, we have from (10) that

$$\omega(X_0, Y_0) \geq \min\{I_{3/4}(4, 2), I_{3/4}(5, 2), I_{3/4}(6, 2)\}$$
$$= I_{3/4}(6, 2) > 0.444 \ .$$

The lemma follows.

Making use of Lemma 20, we next prove an upper bound on the probability that the two-species discrete-time Yule process $X, Y$, with an initial large majority of $X$, eventually hits a step where its relative population size drops to $\frac{3}{4}$ or below. We will later on use this result, instantiating it with species $X^0, X^1$ of a dual-rail encoding of a signal $X$, to make sure both rails remain separated and the signal $X$ maintains its initial value.

**Lemma 21** *Let $X$ and $Y$ be species from a Yule process. Assume that $\frac{X_0}{X_0 + Y_0} > \frac{3}{4}$. Then*

$$\mathbb{P}\left(\exists k : \frac{X_k}{X_k + Y_k} \leq \frac{3}{4}\right) < \frac{I_{3/4}(X_0, Y_0)}{0.444} \ .$$

*Proof* By assumption $X_0 > 3Y_0$. Further, we have

$$I_{3/4}(X_0, Y_0) = \mathbb{P}\left(\lim_{k \to \infty} \frac{X_k}{X_k + Y_k} < \frac{3}{4}\right)$$

$$= \mathbb{P}\left(\lim_{k \to \infty} \frac{X_k}{X_k + Y_k} < \frac{3}{4} \ \Big| \ \exists k : \frac{X_k}{X_k + Y_k} \leq \frac{3}{4}\right) \cdot$$

$$\mathbb{P}\left(\exists k : \frac{X_k}{X_k + Y_k} \leq \frac{3}{4}\right) +$$

$$\mathbb{P}\left(\lim_{k \to \infty} \frac{X_k}{X_k + Y_k} < \frac{3}{4} \wedge \forall k : \frac{X_k}{X_k + Y_k} > \frac{3}{4}\right)$$

$$\overset{\text{Lemma 20}}{>} 0.444 \cdot \mathbb{P}\left(\exists k : \frac{X_k}{X_k + Y_k} \leq \frac{3}{4}\right) \ .$$

The lemma follows.

While Lemma 20 dealt with the correctness of a single dual-rail signal, the following lemma provides a lower bound on the probability that the dual-rail encoding of signals $A$ and $B$, that are both initially $(n, \Delta)$-correct, for $\Delta > n/2$, remains separated as their species grow. We will make use of this result in Boolean gates with two dual-rail inputs, making sure that the inputs of the gate remain their correct signal value.

**Lemma 22** *Let $A^0, A^1$ as well as $B^0, B^1$ be species of a dual-rail encoding of signals $A$ and $B$. Assume that each species follows a Yule processes. If signals $A$ and $B$ are initially $(n, \Delta)$-correct with $n, \Delta \in \mathbb{N}$ with $\Delta > \frac{n}{2}$ for some $a, b \in \{0, 1\}$, then*

$$\mathbb{P}\left(\forall t \geq 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}\right) \geq \Bigg(1 -$$

$$\frac{1}{2 \cdot 0.444} \exp\left(\frac{1}{2}\left(-\frac{\Delta^2}{(n - \Delta)} + \max\{A(0), B(0)\}\right)\right)\Bigg)^2 \tag{11}$$

*Proof* By Independence of the two Yule processes, we have

$$\mathbb{P}\left(\forall t \geq 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}\right) =$$

$$\mathbb{P}\left(\forall t \geq 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4}\right) \cdot \mathbb{P}\left(\forall t \geq 0 : \frac{B^b(t)}{B(t)} > \frac{3}{4}\right) \tag{12}$$

Further, since $A$ is $(n, \Delta)$-correct with $\Delta > \frac{n}{2}$,

$$A(0) = 2A^a(0) - (A^a(0) - A^{\neg a}(0)) \quad \Rightarrow$$

$$A^a(0) \geq \frac{A(0) + \Delta}{2} \quad \Rightarrow$$

$$\frac{A^a(0)}{A(0)} \geq \frac{n + \Delta}{2n} > \frac{3}{4} \ .$$

By analogous arguments, $\frac{B^b(0)}{B(0)} > \frac{3}{4}$. We may thus apply Lemma 21 twice to (12), obtaining

$$\mathbb{P}\left(\forall t \geq 0: \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}\right) >$$
$$\left(1 - \frac{I_{3/4}\big(A^a(0), A^{\neg a}(0)\big)}{0.444}\right) \cdot$$
$$\left(1 - \frac{I_{3/4}\big(B^b(0), B^{\neg b}(0)\big)}{0.444}\right) .$$

We can now apply Lemma 18 twice: for $X = A^a(0)$ and $Y = A^{\neg a}(0)$, and for $X = B^b(0)$ and $Y = B^{\neg b}(0)$. For $A^a$ and $A^{\neg a}$, we further have

$$-\frac{(A^a(0) - A^{\neg a}(0) + 1)^2}{4(A^{\neg a}(0) - 1)} + \frac{A(0)}{2}$$
$$\leq -\frac{(A^a(0) - A^{\neg a}(0))^2}{4A^{\neg a}(0)} + \frac{A(0)}{2}$$
$$\overset{(5)}{\leq} -\frac{\Delta^2}{4\frac{n-\Delta}{2}} + \frac{A(0)}{2}$$
$$= \frac{1}{2}\left(-\frac{\Delta^2}{(n - \Delta)} + A(0)\right) .$$

By analogous arguments for $B^b$ and $B^{\neg b}$, the bound in (11) follows.

We next show in Lemma 23 that when the NAND gates has produced $n$ output species $Y^0$ and $Y^1$, a certain gap $\Delta > 0$ is guaranteed with a probability that depends on $n$ and $\Delta$. However, instead of showing this for the original NAND gate, we first prove that the bound holds for an adapted version where $Y^0$ and $Y^1$ do not duplicate. We later extend the result to the original NAND gate in Lemma 24.

**Lemma 23** *Consider an adapted version of the* NAND *gate with dual-rail encoded input signals $A, B$ and output signal $Y$. In the adapted version, species $Y^0$ and $Y^1$ do not duplicate. Further, assume that for some $a, b \in \{0, 1\}$,*

$$\forall t \geq 0: \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4} .$$

*Then, with $y = \neg(a \wedge b)$ being the correct Boolean output of the gate, for any $t \geq 0$ and $\Delta, n \in \mathbb{N}$ with $\Delta \leq n/8$,*

$$\mathbb{P}\left(Y^y(t) - Y^{\neg y}(t) > \Delta \mid Y(t) = n\right) \geq$$
$$1 - \exp\left(-\frac{\left(\frac{n}{8} - \Delta\right)^2}{2n}\right) .$$

*Proof* From the assumption on the inputs, we have that the probability of the NAND gate to chose species $A^a$ and $B^b$ when producing an output species, is at least $p = \left(\frac{3}{4}\right)^2$. Likewise a wrong output is produced with probability at most $1 - p$.

Consider the discrete random walk on $\mathbb{Z}$, starting at position $D_0 = 0$, and at step $i \geq 1$, incrementing $D_{i-1}$ by one with probability $p$, and decrementing by one with probability $1 - p$. It is easy to construct a coupling such that $D_n \leq Y^y(t) - Y^{\neg y}(t)$, given that $Y(t) = n$.

Let $I_i$, $i \geq 1$, be a sequence of i.i.d. Bernoulli trials with success probability $p$, and $R_n = \sum_{i=1}^{n} I_i$. Then $R_n$ follows a Binomial distribution and $2R_n - n$ is identically distributed to $D_n$. Thus,

$$\mathbb{P}(D_n > \Delta) = 1 - \mathbb{P}\left(R_n \leq \frac{\Delta + n}{2}\right)$$
$$= 1 - \sum_{i=0}^{\frac{\Delta+n}{2}} \binom{n}{i} p^i (1-p)^{n-i} .$$

Applying Chernoff's inequality for sums of Bernoulli trials, we obtain for $\Delta \leq (2p - 1)n = \frac{n}{8}$,

$$\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \leq \exp\left(-2\frac{(np - k)^2}{n}\right)$$

where $k = \frac{\Delta + n}{2}$. Thus,

$$\mathbb{P}(D_n > \Delta) \geq 1 - \exp\left(-2\frac{\left(np - \frac{\Delta+n}{2}\right)^2}{n}\right)$$
$$= 1 - \exp\left(-\frac{\left(\frac{n}{8} - \Delta\right)^2}{2n}\right) .$$

The lemma follows.

**Lemma 24** *Consider the* NAND *gate with dual-rail encoded input signals $A, B$ and output signal $Y$. If for some $a, b \in \{0, 1\}$,*

$$\forall t \geq 0: \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4} ,$$

*$A(0) \geq n$, and $B(0) \geq n$ then, letting $y = \neg(a \wedge b)$ be the correct Boolean output of the gate, with high probability there exists a $t = O(1)$ such that $Y^y(t) - Y^{\neg y}(t) = \Omega(n)$ and $Y(t) = n$.*

*Proof* We first consider the variant of the NAND gate from Lemma 23 where $Y^0$ and $Y^1$ do not duplicate. Let $T > 0$ be the earliest time $t$ when $Y(t) = n$.

By assumption, for all $t' \geq 0$, $A(t') \geq n$ and $B(t') \geq n$. Thus the gate's production rate of $Y$ species is at least $n^2 \alpha$. It follows that with high probability $T \leq \frac{\log n}{\alpha n}$.

We will next upper bound the count of species $Y$ that would have been produced if duplication were in place during time $[0, T]$. For that purpose, assume that

all $Y$ species generated by the gate during $[0, T]$ are already produced at time 0. Then, the count of species $Y$ generated by duplication, let us call them $\hat{Y}$, follows a single species Yule process with initial count $\hat{Y}(0) = n$. Thus, $\hat{Y}(T)$ follows a negative binomial distribution with success probability $p = 1 - e^{-\gamma T}$ and $r = \hat{Y}(0)$, i.e.,

$$\mathbb{P}\left(\hat{Y}(T) = k\right) =$$
$$\binom{k-1}{\hat{Y}(0) - 1} e^{-\gamma T \hat{Y}(0)} \left(1 - e^{-\gamma T}\right)^{k - \hat{Y}(0)} \ .$$

Further, for the expected count of species generated by duplication, minus the initial $\hat{Y}(0)$ that were generated by the gate, we have,

$$\mathbb{E}\left(\hat{Y}(T) - \hat{Y}(0)\right) = \frac{r}{1-p} - r$$
$$= \hat{Y}(0)(e^{\gamma T} - 1) \le n \left(e^{\frac{\gamma}{\alpha} \frac{\log n}{n}} - 1\right) \ .$$

We next show that,

$$\mathbb{E}\left(\hat{Y}(T) - \hat{Y}(0)\right) = O(\log n) \ . \tag{13}$$

Setting $g = \gamma/\alpha$, and letting $C = ge^g$, Equation (13) follows from the fact that for all $n > 0$,

$$n \left(e^{\frac{g \log n}{n}} - 1\right) \le C \log n \qquad \Leftrightarrow$$
$$e^{\frac{g \log n}{n}} \le C \frac{\log n}{n} + 1 \ .$$

Substituting $z = \log n / n$, the latter follows from

$$\forall z \in [0, 1]: \quad e^{gz} \le Cz + 1 \ . \tag{14}$$

Inequality (14), follows by observing that it holds for $z = 0$, and that, by taking the $z$-derivative on both sides, we obtain $ge^{gz} \le C$ which is true for $z \in [0, 1]$ by choice of $C = ge^g$; Equation (13) follows.

Noting that the variance $\sigma^2 = \text{Var}\left(\hat{Y}(T) - \hat{Y}(0)\right)$ of a negative binomial distribution, with $r$ and $p$ as above, is

$$\sigma^2 = \frac{\mathbb{E}\left(\hat{Y}(T) - \hat{Y}(0)\right)}{1 - p} \ ,$$

and setting $\mu = \mathbb{E}(\hat{Y}(T) - \hat{Y}(0))$, we next apply Chebyshev's bound $\mathbb{P}\left(|X - \mu| \ge \epsilon\right) \le \frac{\sigma^2}{\epsilon^2}$.

In particular, the fact that with high probability less than $\mu + \epsilon$ species of $Y$ are generated by duplication, follows from

$$\mathbb{P}\left(\hat{Y}(T) - \hat{Y}(0) \ge \mu + \epsilon\right) \le \frac{\sigma^2}{\epsilon^2} \le 1/n \ . \tag{15}$$

Solving for $\varepsilon$ gives,

$$\frac{\sigma^2}{\epsilon^2} \le 1/n$$
$$\Leftrightarrow \quad \epsilon \ge \sigma\sqrt{n}$$
$$\Leftrightarrow \quad \epsilon \ge \sqrt{n\mathbb{E}\left(\hat{Y}(T) - \hat{Y}(0)\right) e^{\gamma T}} \ .$$

Further, observing that $e^{\gamma T} = e^{\frac{g \log n}{n}} = O(1)$, and using (13), we obtain the existence of a function $h(\cdot)$, such that, if we choose

$$\epsilon \ge h(n) = O\left(\sqrt{n \log n}\right) \ ,$$

Inequality (15) is fulfilled.

Thus, together with (13), one obtains that with high probability $\hat{Y}(T) - \hat{Y}(0)$ is at most

$$\mathbb{E}\left(\hat{Y}(T) - \hat{Y}(0)\right) + \varepsilon = O\left(\sqrt{n \log n}\right) \ . \tag{16}$$

Applying Lemma 23 for $Y(T) = n$, we obtain a bound on the gap $\Delta = Y^y(T) - Y^{\neg y}(T)$, excluding those generated by duplication, that holds with high probability. Choosing

$$\Delta = \frac{n}{8} - \sqrt{2n \log n} \ ,$$

we apply Lemma 23 for $n$ and $\Delta \le \frac{n}{8}$, and obtain

$$\mathbb{P}\left(Y^y(t) - Y^{\neg y}(t) > \Delta \mid Y(t) = n\right) \ge$$
$$1 - \exp\left(-\frac{\left(\frac{n}{8} - \Delta\right)^2}{2n}\right) \ . \tag{17}$$

By choice of $\Delta$,

$$\Delta \le \frac{n}{8} - \sqrt{2n \log n} \qquad \Rightarrow$$
$$\left(\frac{n}{8} - \Delta\right)^2 \ge 2n \log n \qquad \Rightarrow$$
$$\exp\left(-\frac{\left(\frac{n}{8} - \Delta\right)^2}{2n}\right) \ge \frac{1}{n} \qquad \Rightarrow$$
$$1 - \exp\left(-\frac{\left(\frac{n}{8} - \Delta\right)^2}{2n}\right) \ge 1 - \frac{1}{n}$$

Together with (17), we have

$$\mathbb{P}\left(Y^y(t) - Y^{\neg y}(t) > \Delta \mid Y(t) = n\right) \ge 1 - \frac{1}{n} \ .$$

Additionally accounting for the $Y$ species that have been generated by duplication until time $T$, by using (16), we obtain that the gap $Y^y(T) - Y^{\neg y}(T)$ between correct output species $Y^y$ and incorrect output species $Y^{\neg y}$ at time $T$ in a gate with duplication, with high probability, fulfills

$$Y^y(T) - Y^{\neg y}(T) \ge \Delta - \left(\hat{Y}(T) - \hat{Y}(0)\right) = \Omega(n) \ .$$

The lemma follows.

We are now in the position to prove Theorem 17, showing the correctness of the Nand gate if each of the two dual-rail input signals has a sufficiently large gap between its rails.

*Proof (Proof of Theorem 17.)* The theorem follows from Lemma 24 if its assumption holds with high probability. The latter, however, follows from Lemma 22 if the exponent $\frac{1}{2}\left(-\frac{\Delta^2}{(n-\Delta)} + \max\{A(0), B(0)\}\right)$ holds to be in $\Omega\left(-\max\{A(0), B(0)\}\right)$. We next show that this is the case.

Let $M = \max\{A(0), B(0)\}$. From $\Delta \geq \mu M$ with $\mu = 0.62$ we have,

$$-\frac{\Delta^2}{n-\Delta} + M \leq -\frac{\mu^2 M^2}{M - \mu M} + M$$
$$\leq M\left(1 - \frac{\mu^2}{(1-\mu)}\right) \ .$$

It thus remains to show that $\left(1 - \frac{\mu^2}{(1-\mu)}\right) < 0$. By algebraic manipulation, this is the case if $\mu \in \left(\frac{1}{2}(\sqrt{5}-1), 1\right)$, which is true by assumption. The theorem follows.
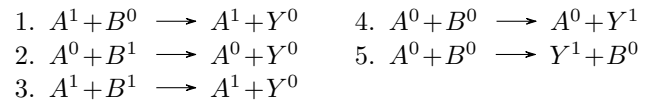
## 5 *In silico* Biological Implementation

While the studied model is a simplification, it represents core functions that constitute collective decision-making among biological species, and is readily adaptable for specific biological applications. If reactions are modified such that one of the two reactants does not change, the model could represent one-way messaging equivalent to a conjugation event between a sender and receiver bacterial cell [36]. Similarly, if the messages $A$ and $B$ are coded as free species diffusible between senders and receivers, it could represent communication between bacterial cells using bacteriophage particles as messages [40].

In this section, we discuss a plausible biological implementation with *E. coli* bacteria that use conjugation to communicate. Conjugation is a method of genetic communication in which circular DNA plasmids are transferred from a sender cell to a receiver cell. An F plasmid allows a cell to be a sender during conjugation. The receiver can be engineered to express a logical function using the received plasmid and its existing DNA, although the internal implementation is not detailed for this simulation. A conjugation reaction with a sender $S$ and a receiver $R$ is described by $R + S \xrightarrow{\delta} f(R, S) + S$, where $\delta$ is the conjugation rate constant, and $f$ is a function to the species. Both, the amplifier and the Nand gate follow this scheme. For the amplifier, $f(R, S) = \emptyset$ and for the Nand gate

$f(R, S) = Y$, where $Y$ is the gate's corresponding output species. While with wild-type F plasmids, *E. coli* are either senders (with F plasmid) or receivers (without F plasmid), there exist engineered systems that allow the same cell (with F plasmid) to be both a sender and a receiver [18,36]. Note that a single cell still cannot act as both the sender and the receiver during a single reaction.

The growth of the *E. coli* is modeled by a logistic model with a carrying capacity of $10^9$ cells. A limited resource is consumed when species duplicate and released when they die. Reaction rate constants for duplication $\gamma = 0.016$ and for conjugation $\delta = 10^{-11}$ have been taken from Dimitriu *et al.* [18]. For our implementation, amplification of the gate's inputs and outputs was executed in parallel to the gate's protocol. The simulations discussed in the following suggest that sequential execution is not required for correctness and performance, greatly simplifying the biological design. If all possible gate reactions were used, inputs that lead to $Y^1$ would be more susceptible to noise since there are more possible input pairs leading to $Y^1$ than $Y^0$ in a Nand gate. This was alleviated by selecting a subset of all possible gate reactions in which three reactions lead to $Y^1$ (see 1–3 below) and two reactions lead to $Y^0$ (see 4–5 below).

1. $A^1 + B^0 \longrightarrow A^1 + Y^0$    4. $A^0 + B^0 \longrightarrow A^0 + Y^1$
2. $A^0 + B^1 \longrightarrow A^0 + Y^0$    5. $A^0 + B^0 \longrightarrow Y^1 + B^0$
3. $A^1 + B^1 \longrightarrow A^1 + Y^0$

Simulation of our system for the four possible input choices are shown in Fig. 8. For performance with many individuals, simulations are done using the $\tau$-leaping approximation of stochastic simulation, in which multiple reactions occur during a dynamic time interval of $\tau$, before updating reaction rates [23,26]. The initial population size is set to $5 \times 10^8$, the carrying capacity to $1 \times 10^9$, and the initial input error to 10% of wrong input species per input. Despite the low rate of communication from conjugation, the correct output species rapidly out-competes the incorrect output species for all input choices.

## 6 Pairwise Annihilation in an Open System

Many biological applications require growing cells for prolonged periods of time [20,27]. Continuous cultures, also known as chemostats, can be used to constantly supply fresh nutrients while maintaining a fixed volume via a respective outflow. This in- and outflow transforms the system into an open system. We examined the effect of this setting on the Pairwise Annihilation
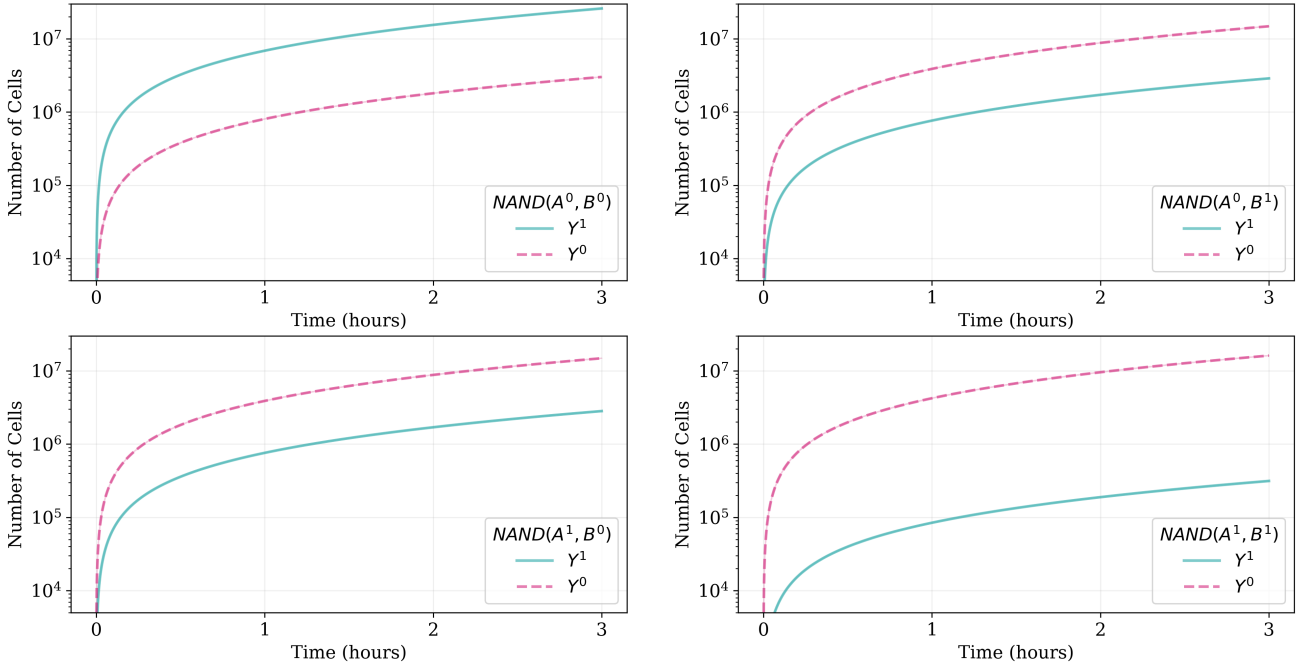
**Fig. 8** A biologically plausible implementation of a NAND gate with amplifiers on inputs and outputs. Note that a subset of all possible gate reactions has been implemented to balance production of $Y^0$ and $Y^1$. Initial population size is $5 \times 10^8$, carrying capacity of $10^9$ cells. Reaction rate constants were set to $\gamma = 0.016$ and $\delta = 10^{-11}$ [18]. The output species is shown for each choice of inputs. The initial input error is 10%. All choices lead to correct, clearly separable outputs within half an hour. Confidence intervals from 30 sample simulations are smaller than the width of the lines.

protocol. Since outflow contributes to the basal death rate of each species, we conjectured that it aids the annihilation process and accelerates consensus. Interestingly, this is not the case. We will show that flow tends to impede consensus.

The section is structured as follows: first, linear stability analysis of an ODE model is used to show that the system converges to majority consensus, assuming that both species do not die. In this range, the analysis suggests that flow impedes consensus, except in a narrow parameter range. Second, stochastic simulations demonstrate that flow increases the time until consensus across a wide range of parameters, including the questionable range isolated by linear analysis.

### 6.1 ODE Model

In addition to the species $A$ and $B$, we use a food species $F$ to model logistic growth due to depletion of resources like nutrients. We use the common notation of $\dot{X}$ for $\frac{d}{dt}X(t)$ and $X$ instead of $X(t)$ when no confusion can arise. Let $\gamma > 0$ be the growth rate constant within fresh medium, $\delta > 0$ the annihilation rate constant, and $\phi > 0$ the in- and outflow rate constant. The fresh medium supplied via the inflow is assumed to contain $\hat{F} > 0$ food per unit volume. As before, food

is consumed by duplication and released by death. We then obtain the following ODE model for the Pairwise Annihilation protocol:

$$\dot{A} = \left(\frac{\gamma F}{\hat{F}} - \phi\right) A - \delta AB$$

$$\dot{B} = \left(\frac{\gamma F}{\hat{F}} - \phi\right) B - \delta AB$$

$$\dot{F} = -\frac{\gamma F}{\hat{F}}(A + B) + 2\delta AB - \phi(F - \hat{F})$$

Letting $X = A + B$, $Y = A - B$, and $Z = A + B + F$ and assuming that $A(0) \geq B(0)$, we have that $Y(t) \geq 0$ for all $t \geq 0$, and we may rewrite the system, obtaining:

$$\dot{X} = \left(\frac{\gamma}{\hat{F}}(Z - X) - \phi\right) X - \frac{\delta}{2}\left(X^2 - Y^2\right)$$

$$\dot{Y} = \left(\frac{\gamma}{\hat{F}}(Z - X) - \phi\right) Y$$

$$\dot{Z} = \phi\left(\hat{F} - Z\right)$$

In the following we will analyze this transformed system.

### 6.2 Linearization at Fixed Points

Let $(X^*, Y^*, Z^*)$ be a fixed point of the above system. The Jacobian $J(X^*, Y^*, Z^*)$ at this point is a linear approximation for the ODEs in the neighborhood of that

fixed point, and its eigenvalues determine the stability of the point [14, 32, 52]: if all real parts of eigenvalues are negative, the point is a stable sink, while a strictly positive real part shows instability along the corresponding eigenvector making it an unstable source or saddle point. The magnitude of the largest real component of an eigenvalue controls the rate of attraction to a stable point or rate of repulsion from an unstable point, so long as the point is not approached precisely along another eigenvector.

In our case, the Jacobian $J(X, Y, Z)$ is equal to

$$
J = \begin{bmatrix} \frac{\gamma}{\hat{F}}Z - \phi - (2\frac{\gamma}{\hat{F}} + \delta)X & \delta Y & \frac{\gamma}{\hat{F}}X \\ -\frac{\gamma}{\hat{F}}Y & \frac{\gamma}{\hat{F}}(Z - X) - \phi & \frac{\gamma}{\hat{F}}Y \\ 0 & 0 & -\phi \end{bmatrix}
$$

and there exist 3 fixed points:

- The *washout* point $(0, 0, \hat{F})$ where $A = B = 0$.
- The *metastable* point $((2\gamma - 2\phi)/(2\gamma + \delta), 0, \hat{F})$ with $A = B > 0$.
- The *majority consensus* point $(\hat{F}(1 - \phi/\gamma), \hat{F}(1 - \phi/\gamma), \hat{F})$ with $\min(A, B) = 0$ and $\max(A, B) > 0$.

Only the last point will be referred to as majority consensus, since it is not desirable in practice to reach consensus if both species eventually die (as in washout). Next, we show that irrespective of the parameters, at least one of the fixed points is stable. In particular, we will show that the metastable point is always unstable, whereas the washout and majority consensus points exchange stability during a transcritical bifurcation at $\gamma = \phi$.

The system will always converge to the only stable point since the system is bounded without periodic or chaotic behavior. No periodic behavior occurs since the imaginary components of all eigenvalues are always zero, as shown below. The system is bounded since $Z$ approaches a constant $\hat{F}$ independently of $X$ and $Y$, and $Y \leq X \leq Z$ by definition. Finally, chaos does not occur: since $Z$ approaches $\hat{F}$, the system is 2-dimensional and bounded in the limit, which cannot exhibit chaos [9, 52]. We detail the stability of each point in the following.

*Washout point $A = B = 0$.* At the fixed point where washout happens, we have

$$
J = \begin{bmatrix} \gamma - \phi & 0 & 0 \\ 0 & \gamma - \phi & 0 \\ 0 & 0 & -\phi \end{bmatrix}
$$

with eigenvalues $\lambda_{1,2} = \gamma - \phi$ and $\lambda_3 = -\phi$. All eigenvalues are negative if and only if $\gamma < \phi$, in which case the washout point is a sink. The rate of attraction near

the point is approximated as $\gamma - \phi$. Thus, not surprisingly, higher flow rates $\phi$ accelerate washout. Otherwise, if $\gamma > \phi$, the point becomes a source with $\phi$ impeding escape from its neighborhood whereas $\gamma$ aids escape.

*Metastable point $A = B \neq 0$.* In case both species have identical counts,

$$
J = \begin{bmatrix} -\gamma + \phi & 0 & 2\gamma\hat{F}\frac{\gamma-\phi}{2\gamma+\hat{F}\delta} \\ 0 & \frac{\delta\hat{F}(\gamma-\phi)}{2\gamma+\hat{F}\delta} & 0 \\ 0 & 0 & -\phi \end{bmatrix}
$$

with eigenvalues

$$
\lambda_1 = \phi - \gamma \ ,
$$
$$
\lambda_2 = \frac{\delta\hat{F}(\gamma - \phi)}{2\gamma + \hat{F}\delta} \ , \text{ and}
$$
$$
\lambda_3 = -\phi \ .
$$

Observe that $\lambda_1$ is negative if $\gamma > \phi$, and $\lambda_2$ is negative if $\gamma < \phi$; hence all eigenvalues are never simultaneously negative, so this fixed point is always unstable. If $\gamma < \phi$, trajectories near the metastable point leave it with a repulsion rate of $\phi - \gamma$. Thus a larger flow rate constant $\phi$ increases the rate of repulsion. Otherwise, if $\gamma > \phi$, trajectories leave with a rate of $\frac{\delta\hat{F}(\gamma-\phi)}{2\gamma+\hat{F}\delta}$. Hence a larger flow rate constant $\phi$ decreases the repulsion rate when washout does not occur. By contrast, the birth and death rate constants $\gamma$ and $\delta$ both increase the rate of repulsion.

*Majority consensus point $\min(A, B) = 0$, $\max(A, B) > 0$.* We have,

$$
J = \begin{bmatrix} -\left(\delta + \frac{\gamma}{\hat{F}}\right)\hat{F}\left(1 - \frac{\phi}{\gamma}\right) & \delta\hat{F}\left(1 - \frac{\phi}{\gamma}\right) & \gamma\left(1 - \frac{\phi}{\gamma}\right) \\ -\gamma\left(1 - \frac{\phi}{\gamma}\right) & 0 & \gamma\left(1 - \frac{\phi}{\gamma}\right) \\ 0 & 0 & -\phi \end{bmatrix}
$$

with eigenvalues

$$
\lambda_1 = \phi - \gamma \ , \quad \lambda_2 = \delta\hat{F}\left(\frac{\phi}{\gamma} - 1\right) \ , \text{ and } \lambda_3 = -\phi \ .
$$

All eigenvalues are negative if and only if $\gamma > \phi$. In case the largest eigenvalue is $-\phi$, higher $\phi$ increases the rate of attraction. In other words $\phi$ seems to accelerate majority consensus where $\phi < \min\left(\frac{\gamma}{2}, \frac{\delta\hat{F}\gamma}{\gamma+\delta\hat{F}}\right)$.

The preceding linear analysis demonstrates that if washout does not occur, majority consensus is achieved, since the majority consensus point would be the only sink in a bounded system without periodic or chaotic behavior. This stability result holds in the nonlinear,

deterministic case [14,32,52]. However, the implications of the linear analysis for the attraction and repulsion rates is an approximation for the overall system's behavior. This approximation suggests that $\gamma$ and $\delta$ accelerate majority consensus by increasing the attraction rate near majority consensus and the repulsion rate near the other fixed points. Conversely, $\phi$ tends to impede majority consensus, except possibly where $\phi < \min\left(\frac{\gamma}{2}, \frac{\delta\hat{F}\gamma}{\gamma+\delta\hat{F}}\right)$.
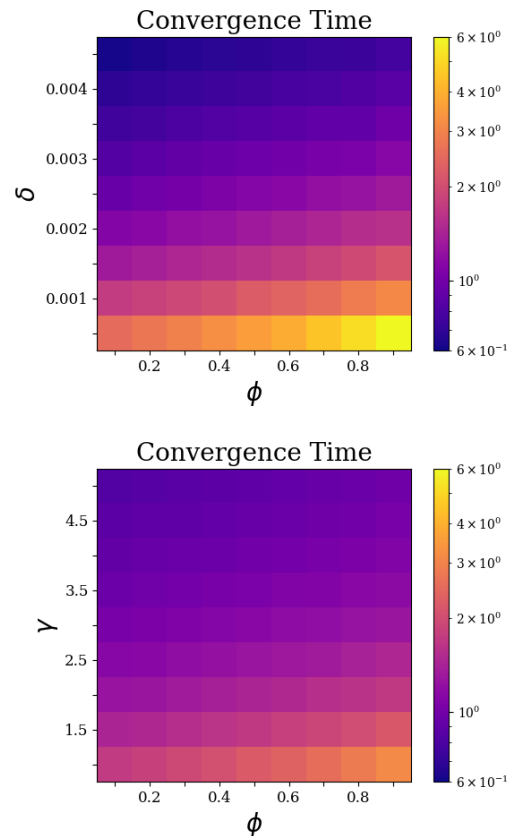
### 6.3 Simulation of the Open System

Next we ran stochastic simulations [22,26] to observe the overall behavior of the system. Two pairwise parameter sweeps are performed: $\delta$ with $\phi$ and $\gamma$ with $\phi$. The results in Fig. 9 are mostly in accordance with the observations from linearization: larger flow rate constants $\phi$ are seen to increase the times until majority consensus as $\phi$ approaches the birth rate constant $\gamma$ in fresh medium. However, both parameter sweeps cross $\phi = \min\left(\frac{\gamma}{2}, \frac{\delta\hat{F}\gamma}{\gamma+\delta\hat{F}}\right)$ without $\phi$ ever reducing majority consensus time, in contrast to the linear approximation around the majority consensus point.

The same two pairwise parameter sweeps are explored over a far wider parameter range in Fig. 10 in the appendix. The results reinforce that $\phi$ does not decrease the time until majority consensus. For all simulations, the measured probability for majority consensus is always 1 in the parameter ranges shown, which is consistent with the washout point being unstable for $\gamma > \phi$. Note that high flow rate constants $\phi$ could occasionally help knock out the smaller species $B$ for very small initial $B(0)$. However, in a biological setting, we expect numbers of at least several thousand cells, as used in the simulation.

The linear analysis and stochastic simulations confirm that majority consensus is still reached within plausible biological parameter ranges and sufficiently slow in- and outflow, but suggest that consensus takes longer to reach.

### 7 Conclusions

We considered the majority consensus problem with continuous population growth in a stochastic setting, and established the Pairwise Annihilation protocol between two competing species $A$ and $B$ with birth reactions $A \longrightarrow 2A$ and $B \longrightarrow 2B$, and death reaction $A + B \longrightarrow \emptyset$. In particular, the input of the Pairwise Annihilation protocol are two species $A$ and $B$ with an initial total population size $n = A(0) + B(0)$ and



**Fig. 9** Stochastic simulation of an open system: a limited parameter sweep emphasizes that higher flow rate constant $\phi$ increases majority consensus time as $\phi$ approaches the growth rate constant $\gamma$. Log-scaled expected convergence time (in minutes) is shown in color from 200 repetitions. Corresponding values are shown on the adjacent vertical bars. $A_0 = 18,000, B_0 = 2,000, F_0 = \hat{F} = 20,000$. **Top:** parameter sweep over $\delta$ and $\phi$, while $\gamma = 1$. **Bottom:** parameter sweep over $\gamma$ and $\phi$, while $\delta = 10^{-3}$.

an initial gap $\Delta = |A(0) - B(0)|$. We showed that the Pairwise Annihilation protocol reaches majority consensus with high probability if the gap weakly grows with the population size according to $\Delta = \Omega(\sqrt{n \log n})$. Expected convergence time until consensus is constant and in $O(\log n)$ with high probability. Simulations show that the qualitative behavior of our protocols matches the behavior expected from these asymptotic bounds.

We further demonstrated how to use dual-rail gates to implement digital circuits computing two-input Boolean functions in the example of a NAND gate. As opposed to thresholds of a single species, dual-rail encoding is particularly useful in our birth systems as the Pairwise Annihilation protocol allows us to amplify and thus regenerate such signals.

As a dual-rail gate implementation, we presented the NAND gate protocol that takes two dual-rail encoded input signals and produces a corresponding dual-

rail output signal. The protocol is simple, an important criterion for follow up in real-world biological implementations. We proved that, given a sufficiently large initial gap between the rails of the input signals, our gate produces the correct output with high probability in $O(\log n)$ time, where $n$ is a lower bound on the initial input population size. In particular, our gate guarantees an output signal gap of $\Omega(n)$ if both inputs have a gap of at least 0.62 times their initial population size. By alternating execution of the NAND gate protocol and the Pairwise Annihilation protocol, layer by layer, we finally arrive at computing the circuit's outputs.

Finally, several biologically motivated extensions of the Pairwise Annihilation protocol were explored. First, we simulated a potential biological implementation that is based on communication by conjugation among engineered *E. coli* that computes a noisy NAND gate. Second, a continuous culture setting with in- and outflow was analyzed. We demonstrated that majority consensus is reached with sufficiently slow in- and outflow and our simulations suggest that it takes longer to do so in presence of flow. While the studied protocols are simplifications of biological implementations, we believe that our results give a signpost for future research to implement complex distributed systems such as indirect inter-cellular communication.

# References

1. Dan Alistarh, James Aspnes, and Rati Gelashvili. Space-optimal majority in population protocols. In *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2018)*, pages 2221–2239. Society for Industrial and Applied Mathematics, 2018.
2. Talley Amir, James Aspnes, and John Lazarsfeld. Approximate Majority with Catalytic Inputs. In Quentin Bramas, Rotem Oshman, and Paolo Romano, editors, *24th International Conference on Principles of Distributed Systems (OPODIS 2020)*, volume 184 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
3. J. Christopher Anderson, Elizabeth J. Clarke, Adam P. Arkin, and Christopher A. Voigt. Environmentally controlled invasion of cancer cells by engineered bacteria. *Journal of Molecular Biology*, 355(4):619–627, January 2006.
4. Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(3):183–199, 2008.
5. Dana Angluin, James Aspnes, and David Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, March 2008.
6. Dana Angluin, James Aspnes, David Eisenstat, and Eric Ruppert. The computational power of population protocols. *Distributed Computing*, 20(4):279–304, August 2007.
7. Dana Angluin, Michael J. Fischer, and Hong Jiang. Stabilizing consensus in mobile networks. In Phillip B. Gibbons, Tarek Abdelzaher, James Aspnes, and Ramesh Rao, editors, *Distributed Computing in Sensor Systems*, volume 4026 of *LNCS*, pages 37–50. Springer Berlin Heidelberg, Heidelberg, 2006.
8. Stav Ben-Nun, Tsvi Kopelowitz, Matan Kraus, and Ely Porat. An $O(\log^{3/2} n)$ parallel time population protocol for majority with $O(\log n)$ states. In *Proceedings of the 39th Symposium on Principles of Distributed Computing*, PODC '20, page 191–199, New York, NY, USA, 2020. Association for Computing Machinery.
9. Ivar Bendixson. Sur les courbes définies par des équations différentielles. *Acta Mathematica*, 24(1):1, 1901.
10. Petra Berenbrink, Robert Elsässer, Tom Friedetzky, Dominik Kaaser, Peter Kling, and Tomasz Radzik. A Population Protocol for Exact Majority with $O(\log n^{5/3})$ Stabilization Time and $\Theta(\log n)$ States. In Ulrich Schmid and Josef Widder, editors, *32nd International Symposium on Distributed Computing (DISC 2018)*, volume 121 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:18, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
11. Petra Berenbrink, Robert Elsässer, Tom Friedetzky, Dominik Kaaser, Peter Kling, and Tomasz Radzik. Time-space trade-offs in population protocols for the majority problem. *Distributed Computing*, 34:91–111, 2021.
12. L Billard. Competition between two species. *Stochastic Processes and their Applications*, 2(4):391–398, 1974.
13. Pierre Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, Heidelberg, 1999.
14. Guanrong Chen. Stability of nonlinear systems. *Encyclopedia of RF and Microwave Engineering*, pages 4881–4896, 2004.
15. Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. *Natural computing*, 13(4):517–534, 2014.
16. Anne Condon, Monir Hajiaghayi, David Kirkpatrick, and Ján Maňuch. Approximate majority analyses using trimolecular chemical reaction networks. *Natural Computing*, 19:249–270, 2020.
17. Ramiz Daniel, Jacob R. Rubens, Rahul Sarpeshkar, and Timothy K. Lu. Synthetic analog computation in living cells. *Nature*, 497(7451):619–623, May 2013.
18. Tatiana Dimitriu, Chantal Lotton, Julien Bénard-Capelle, Dusan Misevic, Sam P Brown, Ariel B Lindner, and François Taddei. Genetic information transfer promotes cooperation in bacteria. *Proceedings of the National Academy of Sciences*, 111(30):11103–11108, 2014.
19. David Doty, Mahsa Eftekhari, and Eric Severson. A stable majority population protocol using logarithmic time and states, 2021.
20. Kevin M. Esvelt, Jacob C. Carlson, and David R. Liu. A system for the continuous directed evolution of biomolecules. *Nature*, 472(7344):499–503, April 2011.
21. William Feller. Die Grundlagen der volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung (1939). In *Selected Papers I*, pages 441–470. Springer, 2015.
22. Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977.

23. Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

24. Shafi Goldwasser, Rafail Ostrovsky, Alessandra Scafuro, and Adam Sealfon. Population stability. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing - PODC '18*, pages 397–406. ACM Press, 2018.

25. Antonio Gómez-Corral and M López García. Extinction times and size of the surviving species in a two-species competition process. *Journal of Mathematical Biology*, 64(1-2):255–289, 2012.

26. S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, October 2006.

27. Paul A Hoskisson and Glyn Hobbs. Continuous culture–making a comeback? *Microbiology*, 151(10):3153–3159, 2005.

28. Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 2 edition, 1975.

29. David G Kendall. Branching processes since 1873. *Journal of the London Mathematical Society*, 1(1):385–406, 1966.

30. Eric Libby and William C Ratcliff. Ratcheting the evolution of multicellularity. *Science*, 346(6208):426–427, 2014.

31. L. Lovász, J. Pelikán, and K. Vesztergombi. *Discrete Mathematics: Elementary and Beyond*. Springer, Heidelberg, 2003.

32. Aleksandr Mikhailovich Lyapunov. The general problem of the stability of motion. *International journal of control*, 55(3):531–534, 1992.

33. Nancy A Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.

34. Marcelo O. Magnasco. Chemical kinetics is Turing universal. *Physical Review Letters*, 78(6):1190–1193, February 1997.

35. Hosam M. Mahmoud. *Pólya Urn Models*. CRC Press, Boca Raton, 2009.

36. John P Marken and Richard M Murray. Addressable," packet-based" intercellular communication through plasmid conjugation. *bioRxiv*, page 591552, 2019.

37. Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C. Stanton, and Christopher A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, October 2012.

38. Chris J Myers. *Asynchronous Circuit Design*. John Wiley & Sons, 2001.

39. Artem S Novozhilov, Georgy P Karev, and Eugene V Koonin. Biological applications of the theory of birth-and-death processes. *Briefings in Bioinformatics*, 7(1):70–85, 2006.

40. Monica E Ortiz and Drew Endy. Engineered cell-cell communication via DNA messaging. *Journal of Biological Engineering*, 6(1):16, December 2012.

41. C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, and J. D. Newman. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446):528–532, April 2013.

42. William C Ratcliff, R Ford Denison, Mark Borrello, and Michael Travisano. Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, 109(5):1595–1600, 2012.

43. Sergi Regot, Javier Macia, Núria Conde, Kentaro Furukawa, Jimmy Kjellén, Tom Peeters, Stefan Hohmann, Eulàlia de Nadal, Francesc Posas, and Ricard Solé. Distributed biological computation with multicellular engineered networks. *Nature*, 469(7329):207–211, 2010.

44. GEH Reuter. Competition processes. In *Proc. 4th Berkeley Symp. Math. Statist. Prob*, volume 2, pages 421–430, 1961.

45. CJ Ridler-Rowe. On competition between two species. *Journal of Applied Probability*, 15(3):457–465, 1978.

46. Thomas L Saaty. *Elements of Queueing Theory*, volume 34203. McGraw-Hill New York, 1961.

47. Sebastian R Schmidl, Felix Ekness, Katri Sofjan, Kristina N-M Daeffler, Kathryn R Brink, Brian P Landry, Karl P Gerhardt, Nikola Dyulgyarov, Ravi U Sheth, and Jeffrey J Tabor. Rewiring bacterial two-component systems by modular dna-binding domain swapping. *Nature Chemical Biology*, 15(7):690–698, 2019.

48. Fred B Schneider. Implementing fault-tolerant services using the state machine approach: A tutorial. *ACM Computing Surveys (CSUR)*, 22(4):299–319, 1990.

49. Shimyn Slomovic, Keith Pardee, and James J Collins. Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences*, 112(47):14429–14435, 2015.

50. David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, February 2008.

51. Jens Sparsø and Steve Furber. *Principles of asynchronous circuit design*. Springer, 2002.

52. Steven H Strogatz. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

53. Jeffrey J Tabor, Howard M Salis, Zachary Booth Simpson, Aaron A Chevalier, Anselm Levskaya, Edward M Marcotte, Christopher A Voigt, and Andrew D Ellington. A synthetic genetic edge detection program. *Cell*, 137(7):1272–1281, 2009.

54. Alvin Tamsir, Jeffrey J. Tabor, and Christopher A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469(7329):212–215, December 2010.

55. Pei Kun R Tay, Peter Q Nguyen, and Neel S Joshi. A synthetic circuit for mercury bioremediation using self-assembling functional amyloids. *ACS Synthetic Biology*, 6(10):1841–1850, 2017.

56. Vito Volterra. *Leçons sur la Theorie Mathematique de la Lutte pour la Vie*. Gauthier-Villars, Paris, 1931.

# A Proof of Lemma 3

By the law of total probability, for every $t \geq 0$, we have

$$\mathbb{P}(T \leq t) = \sum_{k=0}^{\infty} p(1-p)^k \mathbb{P}(T_1 + \cdots + T_{k+1} \leq t)$$

$$= \sum_{k=0}^{\infty} p(1-p)^k e^{-\lambda t} \sum_{i=k+1}^{\infty} \frac{1}{i!}(\lambda t)^i$$

$$= e^{-\lambda t} \sum_{i=0}^{\infty} \frac{1}{i!}(\lambda t)^i p \sum_{k=0}^{i-1}(1-p)^k$$

$$= e^{-\lambda t} \sum_{i=0}^{\infty} \frac{1}{i!}(\lambda t)^i (1-(1-p)^i)$$

$$= e^{-\lambda t}(e^{\lambda t} - e^{(1-p)\lambda t}) = 1 - e^{-p\lambda t} \ ,$$

which is equal to the cumulative distribution function of an exponential random variable with parameter $p\lambda$.

## B Expected Absorption Time in Birth-Death Chains

**Theorem 25 ([28, p. 149])** *Consider a birth and death process with birth and death parameters $\lambda_n$ and $\mu_n$, $n \geq 1$, where $\lambda_0 = 0$ so that 0 is an absorbing state.*

*Let $\rho_i = (\lambda_1 \lambda_2 \cdots \lambda_{i-1})/(\mu_1 \mu_2 \cdots \mu_i)$. If $\sum_{i=1}^{\infty} \rho_i < \infty$, then the mean time to absorption into state 0 from the initial state $m$ is*

$$\sum_{i=1}^{\infty} \rho_i + \sum_{r=1}^{m-1} \left( \prod_{k=1}^{r} \frac{\mu_k}{\lambda_k} \right) \sum_{j=r+1}^{\infty} \rho_j \ .$$

## C Urn Draws and Beta Distribution

A Pólya-Eggenberger urn is an urn containing white and black balls. At each discrete time step, we draw a ball uniformly at random from the urn, independently from the other draws. The color of the drawn ball is observed and the ball, as well as an additional ball of the same color, is returned to the urn. We denote by $W_n$ and $B_n$ the number of white and black balls in the urn after $n$ draws, respectively.

**Theorem 26 ([35, Theorem 3.2])** *Let $\tilde{W}_n$ be the number of white ball draws in a Pólya-Eggenberger urn after $n$ draws. Then $\tilde{W}_n/n$ converges in distribution to a beta distribution with parameters $\alpha = W_0$ and $\beta = B_0$.*

## D Regularized Incomplete Beta Function

The regularized incomplete beta function $I_z(\alpha, \beta)$ is given as

$$I_z(\alpha, \beta) = \int_0^z t^{\alpha-1}(1-t)^{\beta-1}\,dt \ \Big/ \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \ .$$

The following identities hold:

$$I_z(\alpha+1, \beta) - I_z(\alpha, \beta) = -\frac{z^{\alpha}(1-z)^{\beta}\Gamma(\alpha+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} \tag{18}$$

$$I_z(\alpha, \beta+1) - I_z(\alpha, \beta) = \frac{z^{\alpha}(1-z)^{\beta}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta+1)} \tag{19}$$

## E Proof of Lemma 16

We have the well-known formula

$$I_z(a, b) = \sum_{j=0}^{a-1} \binom{a+b-1}{j} z^{a+b-1-j}(1-z)^j \tag{20}$$

for $a, b \in \mathbb{N}$ with $a \geq b$. With $z = 1/2$, $a = m + \Delta$, and $b = m$, this implies

$$I_{1/2}(m+\Delta, m) = \frac{1}{2^{2m+\Delta-1}} \sum_{j=0}^{m+\Delta-1} \binom{2m+\Delta-1}{j} \ .$$

The sum of the first binomial coefficients can be upper-bounded (e.g., [31, Proof of Theorem 5.3.2]) via

$$\sum_{j=0}^{k} \binom{n}{j} \leq 2^{n-1} \exp\left(-\frac{(n-2k-2)^2}{4n-4k-4}\right) \ . \tag{21}$$

Setting $n = 2m + \Delta - 1$ and $k = m + \Delta - 1$ we get

$$I_{1/2}(m+\Delta, m) \leq \frac{1}{2}\exp\left(-\frac{(\Delta+1)^2}{4(m-1)}\right)$$

$$\leq \exp\left(-\Omega\left(\frac{\Delta^2}{m}\right)\right) \ .$$

This concludes the proof of the lemma.

## F Proof of Lemma 18

Instantiating (20) with $z = \frac{3}{4}$, $a = X$ and $b = Y$, we have

$$I_{3/4}(X, Y) = \sum_{j=0}^{X-1} \binom{X+Y-1}{j} \left(\frac{3}{4}\right)^{X+Y-1-j} \left(\frac{1}{4}\right)^j$$

$$= \left(\frac{3}{4}\right)^{X+Y-1} \sum_{j=0}^{X-1} \binom{X+Y-1}{j} 3^{-j}$$

$$\leq \left(\frac{3}{4}\right)^{X+Y-1} \sum_{j=0}^{X-1} \binom{X+Y-1}{j}$$

and from (21) with $n = X + Y - 1$ and $k = X - 1$,

$$\leq 3^{X+Y-1} \left(\frac{1}{4}\right)^{X+Y-1} \exp\left(-\frac{(X-Y+1)^2}{4(Y-1)}\right)$$

$$\leq \frac{1}{2}\left(\frac{3}{2}\right)^{X+Y-1} \exp\left(-\frac{(X-Y+1)^2}{4(Y-1)}\right)$$

$$= \frac{1}{2}\exp\left(-\frac{(X-Y+1)^2}{4(Y-1)} + (X+Y-1)\log\frac{3}{2}\right) \ .$$

The lemma's second inequality follows from setting $X = m + \Delta$ and $Y = m$ in the above inequality. Assuming that $m, \Delta \geq 0$, and noting that $\log\frac{3}{2} < \frac{1}{2}$, we obtain by algebraic manipulation

$$-\frac{(X-Y+1)^2}{4(Y-1)} + (X+Y-1)\log\frac{3}{2} \leq$$

$$-\frac{(\Delta+1)^2}{4(m-1)} + \frac{2m+\Delta}{2} \ ;$$

from which the lemma follows.

## G Proof of Lemma 19

We have

$$I_{3/4}(X+3, Y+1) - I_{3/4}(X,Y) =$$

$$I_{3/4}(X+3, Y+1) - I_{3/4}(X+2, Y+1)$$
$$+ I_{3/4}(X+2, Y+1) - I_{3/4}(X+1, Y+1)$$
$$+ I_{3/4}(X+1, Y+1) - I_{3/4}(X, Y+1)$$
$$+ I_{3/4}(X, Y+1) - I_{3/4}(X, Y) \ .$$

Let $T = \frac{\left(\frac{3}{4}\right)^X \left(\frac{1}{4}\right)^Y \Gamma(X+Y)}{\Gamma(X+3)\Gamma(Y+1)}$. Invoking (18), we further have

$$I_{3/4}(X+3, Y+1) - I_{3/4}(X+2, Y+1) =$$

$$- \frac{\left(\frac{3}{4}\right)^{X+2} \left(\frac{1}{4}\right)^{Y+1} \Gamma(X+Y+3)}{\Gamma(X+3)\Gamma(Y+1)} =$$

$$- T \cdot \left(\frac{3}{4}\right)^2 \frac{1}{4}(X+Y)(X+Y+1)(X+Y+2) \overset{Y \leq X/3}{\geq}$$

$$- T \cdot \left(\frac{3}{4}\right)^2 \frac{1}{4}\left(\frac{4X}{3}\right)\left(\frac{4X}{3}+1\right)\left(\frac{4X}{3}+2\right)$$

Similarly, with (18),

$$I_{3/4}(X+2, Y+1) - I_{3/4}(X+1, Y+1) =$$

$$- \frac{\left(\frac{3}{4}\right)^{X+1} \left(\frac{1}{4}\right)^{Y+1} \Gamma(X+Y+2)}{\Gamma(X+2)\Gamma(Y+1)} =$$

$$- T \cdot \frac{3}{4} \cdot \frac{1}{4}(X+Y)(X+Y+1)(X+2) \overset{Y \leq X/3}{\geq}$$

$$- T \cdot \frac{3}{4} \cdot \frac{1}{4}\left(\frac{4X}{3}\right)\left(\frac{4X}{3}+1\right)(X+2)$$

and

$$I_{3/4}(X+1, Y+1) - I_{3/4}(X, Y+1) =$$

$$- \frac{\left(\frac{3}{4}\right)^X \left(\frac{1}{4}\right)^{Y+1} \Gamma(X+Y+1)}{\Gamma(X+1)\Gamma(Y+1)} =$$

$$- T \cdot \frac{1}{4}(X+Y)(X+1)(X+2) \overset{Y \leq X/3}{\geq}$$

$$- T \cdot \frac{1}{4}\left(\frac{4X}{3}\right)(X+1)(X+2)$$

From (19) we have

$$I_{3/4}(X, Y+1) - I_{3/4}(X,Y) =$$

$$\frac{\left(\frac{3}{4}\right)^X \left(\frac{1}{4}\right)^Y \Gamma(X+Y)}{\Gamma(X)\Gamma(Y+1)} =$$

$$T \cdot X(X+1)(X+2)$$

Collecting all terms, we thus have

$$I_{3/4}(X+3, Y+1) - I_{3/4}(X, Y) \geq T \cdot \frac{1}{24}X(8X+11) \ ,$$

which is nonnegative since $X \geq 0$ and $T \geq 0$.

## H Wider Sweep of Parameters for Open Systems

Convergence time in the open system is explored for an exponentially wider range of parameters than in Fig. 9 to emphasize that flow never decreases convergence time. This scale is not visually thrilling, but one can faintly discern that the flow rate constant $\phi$ even increases convergence time as $\phi$ approaches the growth rate constant $\gamma$.
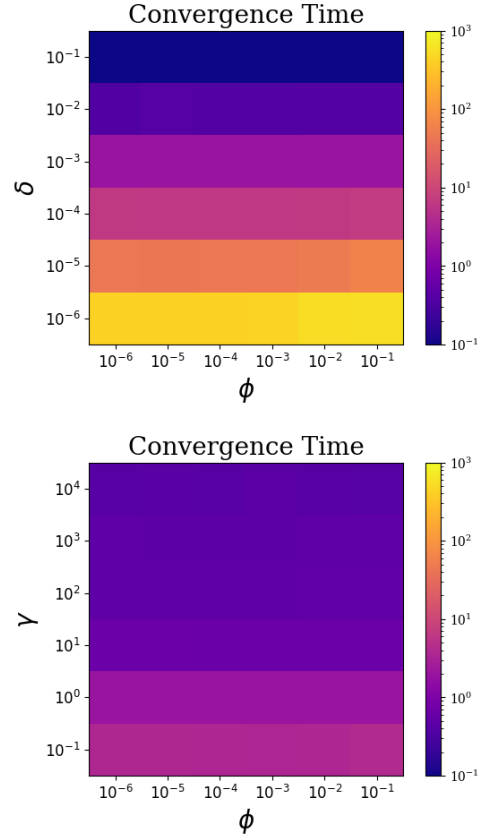


**Fig. 10** Stochastic simulation of an open system over wide range of parameters shows that flow rate constant $\phi$ does not decrease consensus time. Color represents log-scaled expected convergence time. All parameters are identical to Fig. 9, except for the values of the two dependent variables shown on the axis.