HAL open science

# Distributed Computation with Continual Population Growth

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

**To cite this version:**

Da-Jung Cho, Matthias Függer, Corbin Hopper, Manish Kushwaha, Thomas Nowak, et al.. Distributed Computation with Continual Population Growth. DISC 2020 - 34th International Symposium on DIStributed Computing, 2020, virtual, Germany. p. 1-17, 10.4230/LIPIcs.DISC.2020.6 . hal-02946883

**HAL Id: hal-02946883**

**https://hal.science/hal-02946883v1**

Submitted on 23 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Distributed Computation with Continual Population Growth

**Da-Jung Cho**
University of Kassel, Kassel, Germany
dajung.cho@uni-kassel.de

**Matthias Függer**
CNRS, LSV, ENS Paris-Saclay, Université Paris-Saclay, Inria, Gif-sur-Yvette, France
mfuegger@lsv.fr

**Corbin Hopper**
ENS Paris-Saclay, Gif-sur-Yvette, France
Université Paris-Saclay, CNRS, Orsay, France
corbin.hopper@mail.mcgill.ca

**Manish Kushwaha**
Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France
manish.kushwaha@inrae.fr

**Thomas Nowak**
Université Paris-Saclay, CNRS, Orsay, France
thomas.nowak@lri.fr

**Quentin Soubeyran**
École polytechnique, Institut Polytechnique de Paris, Route de Saclay, Palaiseau, France
Université Paris-Saclay, CNRS, Orsay, France
quentin.soubeyran.16@polytechnique.org

## Abstract

Computing with 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. We demonstrate that combining the NAND gate protocol with the continuous-growth majority consensus protocol, using the latter as an amplifier, it is possible to implement circuits computing arbitrary Boolean functions.

**2012 ACM Subject Classification** Theory of computation → Distributed algorithms

**Keywords and phrases** microbiological circuits, majority consensus, birth-death processes

**Digital Object Identifier** 10.4230/LIPIcs.DISC.2020.6

**Related Version** The full version with proofs is available at `https://arxiv.org/abs/2003.09972`.

**Acknowledgements** We acknowledge support from the Digicosme working group HicDiesMeus, Ile-de-France region's DIM-RFSI, INRAE's MICA department, and the CNRS project ABIDE. We thank Joel Rybicki for feedback on an earlier version.

© D. Cho, M. Függer, C. Hopper, M. Kushwaha, T. Nowak, Q. Soubeyran;
licensed under Creative Commons License CC-BY
34th International Symposium on Distributed Computing (DISC 2020).
Editor: Hagit Attiya; Article No. 6; pp. 6:1–6:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 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 [25, 32], light [35], oxygen [1], pH [31]) and process them to produce desired outputs according to defined processing codes (primarily digital [11, 20], but occasionally analog [8]). Some potential applications of the cellular machines include production of metabolic compounds of interest [24], bio-remediation of toxic environments [37], sensing of disease bio-markers [32], and therapeutic intervention by targeted effector delivery [1]. 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 [27, 36], 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 [17, 26]. 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 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, a problem that has been studied in distributed computing for decades in different contexts [18]. 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 [3], where reactions are restricted to two reactants and two products with constant size populations, but also in computability results in more general CRNs [7, 33].

**Consistent cell states by competition among cells.** Competition among species naturally lends itself to solving consensus-type problems. Angluin *et al.* [3] analyzed a population protocol with three states: $A$, $B$, and blank. Encounters of opposing species $A$ and $B$ lead

to one of them becoming blank, and blank species that encounter a non-blank species copy its state. The population protocol by Angluin *et al.* [2] alternates phases of state duplication and cancellation, separated by a clock signal generated by a dedicated leader species. These protocols, however, rely on constant size populations and the latter on a dedicated leader, rendering them impractical for implementations in bacterial cultures.

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 [22, 30].

An early mention of problems requiring a stochastic analysis of two competing species is by Volterra [38] and Feller [10] although only the growth of a single species is analyzed therein. For an overview of single species birth-death Markov chains, see, e.g., [5]. Extensions for multiples species, with applications to genetic mutations, are found in the literature on competition and branching processes [4, 16, 28]. For example, Ridler-Rowe [29] considers a stochastic process between two competing species. However, the process in that work differs from ours in that death reactions 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 [13] 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. 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 A-B protocol for two species $A$ and $B$ and show that it solves majority consensus with high probability: If the initial difference between $A$ and $B$ sizes $\Delta$ 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 A-B 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. In Section 4, we define and analyze the NAND gate protocol. In Section 5, we present simulations of the A-B protocol as well as a biologically plausible implementation of the NAND gate with amplifiers. Finally, Section 6 summarizes our results.

## 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 will 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 [5] for auxiliary definitions). Given some volume $v \in \mathbb{R}_0^+$, which we will normalize to $v = 1$, 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 will write $P(x, y)$ 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 A-B 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 A-B protocol is thus:

$$A \xrightarrow{\gamma} 2A \qquad B \xrightarrow{\gamma} 2B \qquad A + B \xrightarrow{\delta} \emptyset$$

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 exactly one species is extinct.

We show that the A-B 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 A-B protocol reaches consensus in expected time $O(1)$ and in time $O(\log n)$ with high probability. It reaches majority consensus with probability $1 - e^{-\Omega(\Delta^2/n)}$.*

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

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

Without duplication reactions, it is obvious that the A-B 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 [3, 6].

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 A-B protocol evolves as a continuous-time Markov chain with state space $S = \mathbb{N}^2$. Its state-transition rates are $Q\big((A, B), (A + 1, B)\big) = \gamma A$, $Q\big((A, B), (A, B + 1)\big) = \gamma B$, and

$Q\big((A,B)\,,\,(A-1,B-1)\big) = \delta AB$. 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 probabilities $P\big((A,B)\,,\,(A+1,B)\big) = \frac{\gamma A}{\gamma(A+B)+\delta AB}$, $P\big((A,B)\,,\,(A,B+1)\big) = \frac{\gamma B}{\gamma(A+B)+\delta AB}$, and $P\big((A,B)\,,\,(A-1,B-1)\big) = \frac{\delta AB}{\gamma(A+B)+\delta AB}$ 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 A-B protocol Markov chain with a single-species birth-death process. We show that the A-B protocol reaches consensus when the single-species process reaches its extinction state and then bound this time in the single-species process. Figure 1 visualizes the idea of the proof.

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 A-B 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 A-B protocol chain.

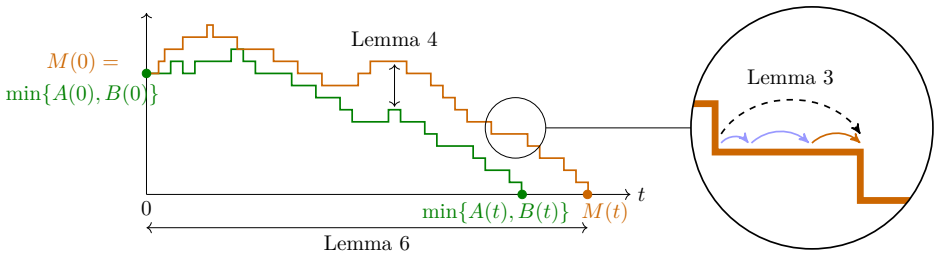


**Figure 1** Idea of the proof: Construction of a continuous-time coupling of the A-B protocol and the single species birth-death M chain. Stuttering steps are mapped to effective steps (Lemma 3). An execution of the coupling process fulfills the deterministic guarantee $\min\{A(t),B(t)\} \leq M(t)$ for all times $t \geq 0$ (Lemma 4). From the coupling it follows that $\mathbb{P}\big(M(t)=0\big) \leq \mathbb{P}\big(A(t)=0 \vee B(t)=0\big)$ for the uncoupled processes (Lemma 5). The time until consensus then follows from the time until extinction in the M chain (Lemma 6).

**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

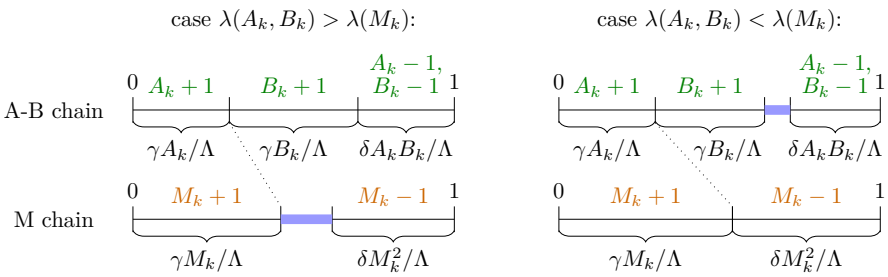


**Figure 2** Continuous-time coupling of the A-B chain and the single-species birth-death M-chain, given that $A_k \leq B_k$, with $\Lambda = \Lambda(A_k, B_k, M_k)$. The intervals for the cases of $\xi_{k+1}$ and their effect on the A-B chain and the M-chain are shown in green and orange, respectively.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

$[0, 1)$ 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 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 $A_k \leq B_k$, then:

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

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:

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

**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 A-B protocol, potentially differ. 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 Figure 2. The following elementary Lemma 3, however, shows that the continuous-time chain $(A(t), B(t))$ and the A-B protocol chain have identical transition rates, and are thus identically distributed. The same holds true for the continuous-time chain $M(t)$ and the birth-death M chain.

▶ **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 species 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* $t \geq 0$.

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

▶ **Lemma 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 A-B 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 processes [15, p. 149] we obtain, when starting from initial population $M(0) = M$, that

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

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

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

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 $M(t) = 0$, Lemma 5 is equivalent to $\mathbb{P}(T_M \leq t) \leq \mathbb{P}(T_{AB} \leq t)$, which, in turn, is equivalent to $\mathbb{P}(T_M > t) \geq \mathbb{P}(T_{AB} > t)$. 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 A-B 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* $\mathbb{P}(T \leq C\log_2 n) = 1 - O(1/n)$.

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

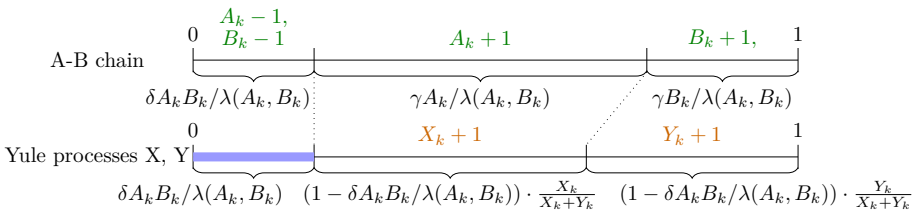


**Figure 3** Discrete-time coupling of A-B chain and two Yule processes X and Y with $\lambda(A_k, B_k) = \gamma(A_k + B_k) + \delta A_k B_k$. Cases for $\xi_{k+1}$ that lead to stuttering steps are shown in blue. The interval relations indicated by the dotted lines are proven by induction in Lemma 8.

## 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 A-B 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 coupled process $(A_k, B_k, X_k, Y_k)$ by $A_0 = X_0$, $B_0 = Y_0$, and

$$(A_{k+1}, B_{k+1}) = \begin{cases} (A_k - 1, B_k - 1) & \text{if } \xi_{k+1} \in \left[0\,,\ \frac{\delta A_k B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\right) \\ (A_k + 1, B_k) & \text{if } \xi_{k+1} \in \left[\frac{\delta A_k B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\,,\ 1 - \frac{\gamma B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\right) \\ (A_k, B_k + 1) & \text{if } \xi_{k+1} \in \left[1 - \frac{\gamma B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\,,\ 1\right) \end{cases}$$

$$(X_{k+1}, Y_{k+1}) = \begin{cases} (X_k, Y_k) & \text{if } \xi_{k+1} \in \left[0\,,\ \frac{\delta A_k B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\right) \\ (X_k + 1, Y_k) & \text{if } \xi_{k+1} \in \left[\frac{\delta A_k B_k}{\gamma(A_k+B_k)+\delta A_k B_k}\,,\ 1 - \frac{\gamma(A_k+B_k)}{\gamma(A_k+B_k)+\delta A_k B_k} \cdot \frac{Y_k}{X_k+Y_k}\right) \\ (X_k, Y_k + 1) & \text{if } \xi_{k+1} \in \left[1 - \frac{\gamma(A_k+B_k)}{\gamma(A_k+B_k)+\delta A_k B_k} \cdot \frac{Y_k}{X_k+Y_k}\,,\ 1\right) \end{cases}$$

if $\max\{A_k, B_k\} > 0$ and $\max\{X_k, Y_k\} > 0$. Otherwise the process remains constant. Figure 3 visualizes the construction.

**Analysis for probability of reaching majority consensus.** The crucial property of this coupling is that the initial minority in the A-B process cannot overtake the initial majority before the initial minority overtakes the initial majority in the parallel Yule processes.

▶ **Lemma 8.** *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 $0 \leq k \leq K$.*

▶ **Lemma 9.** *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 8, if $k$ is minimal such that $A_k = B_k$, then $X_k = Y_k$. ◀

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 sequel.

▶ **Lemma 10.** *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 $\mathbb{P}(\exists k\colon \tilde{X}_k = \tilde{Y}_k) = \mathbb{P}(\exists k\colon X_k = Y_k)$.*

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

▶ **Lemma 11.** *If $X_0 > Y_0$, then $\mathbb{P}\left(\exists k\colon X_k = Y_k\right) = 2 \cdot I_{1/2}(X_0, Y_0)$.*

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

▶ **Lemma 12.** *If $A_0 > B_0$, then $\mathbb{P}\left(\exists k\colon A_k = B_k\right) = 2 \cdot \mathbb{P}(B \text{ wins})$.*

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

▶ **Lemma 13.** *If $A_0 \geq B_0$, then the A-B 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 Lemmas 9, 11, and 12, 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 13, it only remains to upper-bound the term $I_{1/2}(\alpha, \beta)$. Lemma 14 provides such a bound.

▶ **Lemma 14.** *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.

## 4 Boolean Gates

In terms of circuit design, the A-B protocol can be viewed as a differential signal amplifier. Differential signaling has applications in systems that require high resilience to noise, and thus an application for our inherently growing systems is natural.

In this section we study a protocol that allows to compute the logical NAND of two signals, however with a loss of signal quality at the output. The A-B 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.

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 $x(t) \in \{0, 1, \perp\}$. Following a technique from clockless circuit design [34, 21] 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 $\perp$. We will make the assumptions on the input signals precise in the sequel, 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

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

The *initial gap* $X^x(0) - X^{\neg x}(0)$ of signal $X$ is thus bounded by $X^x(0) - X^{\neg x}(0) = X^x(0) + X^{\neg x}(0) - 2X^{\neg x}(0) \geq \Delta$.

## 4.1 Dual-Rail NAND Gate

A dual-rail implementation of a NAND gate with input signals $A, B$ and output signal $Y$ 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

$$A^a + B^b \xrightarrow{\alpha} A^a + B^b + Y^y \quad ,$$

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 $i \in \{0, 1\}$. Figure 4 depicts the NAND gate with the subsequent amplification protocol.

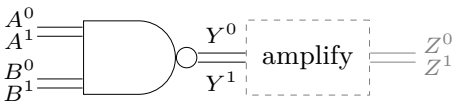


**Figure 4** Dual-rail NAND gate with input signals $A$ and $B$ an output signal $Y$. Successive amplification of $Y$ to signal $Z$ shown in gray.

In Section 4.2 we will show that the NAND gate ensures the following input-output specification:

▶ **Theorem 15.** *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.*

## 4.2 Gate Correctness and Performance

We now turn to the proof of Theorem 15. For our analysis we need a bound on the regularized incomplete beta function $I_{3/4}$.

▶ **Lemma 16.** *For $X \geq Y$, it is $I_{3/4}(X, Y) \leq \frac{1}{2} \exp\left(-\frac{(X-Y+1)^2}{4(Y-1)} + (X + Y - 1) \log \frac{3}{2}\right)$. In particular, for $m, \Delta \geq 0$, $I_{3/4}(m + \Delta, m) \leq \frac{1}{2} \exp\left(-\frac{(\Delta+1)^2}{4(m-1)} + \frac{2m+\Delta}{2}\right)$.*

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 17.** *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 $X_\ell/(X_\ell+Y_\ell) \le 3/4$.

▶ **Lemma 18.** *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* $\mathbb{P}\left(\lim_{k\to\infty}\frac{X_k}{X_k+Y_k} < \frac{3}{4} \;\middle|\; \exists \ell : \frac{X_\ell}{X_\ell+Y_\ell} \le \frac{3}{4}\right) \ge \omega\big(X_0,Y_0\big)$ *where* $\omega\big(X_0,Y_0\big) = \inf\big\{I_{3/4}(x,y) \,\big|\, x \ge X_0 \wedge y \ge Y_0+1 \wedge x \in 3y - \{0,1,2\}\big\}$. *Moreover,* $\omega\big(X_0,Y_0\big) > 0.444$

Making use of Lemma 18, 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.

▶ **Lemma 19.** *Let $X$ and $Y$ be species from a Yule process. Assume that $\frac{X_0}{X_0+Y_0} > \frac{3}{4}$. Then* $\mathbb{P}\left(\exists k : \frac{X_k}{X_k+Y_k} \le \frac{3}{4}\right) < \frac{I_{3/4}(X_0,Y_0)}{0.444}$ .

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.

▶ **Lemma 20.** *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* $\mathbb{P}\left(\forall t \ge 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}\right)$ *is lower bounded by*

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

We next show in Lemma 21 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 22.

▶ **Lemma 21.** *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 $a,b \in \{0,1\}$, $\forall t \ge 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}$. Then, with $y = \neg(a \wedge b)$ being the correct Boolean output of the gate, for any $t \ge 0$ and $\Delta, n \in \mathbb{N}$ with $\Delta \le n/8$,* $\mathbb{P}\left(Y^y(t) - Y^{\neg y}(t) > \Delta \;\middle|\; Y(t) = n\right) \ge 1 - \exp\left(-\frac{\left(\frac{n}{8}-\Delta\right)^2}{2n}\right)$.

▶ **Lemma 22.** *Consider the* NAND *gate with dual-rail encoded input signals $A, B$ and output signal $Y$. If for some $a,b \in \{0,1\}$, $\forall t \ge 0 : \frac{A^a(t)}{A(t)} > \frac{3}{4} \wedge \frac{B^b(t)}{B(t)} > \frac{3}{4}$, $A(0) \ge n$, and $B(0) \ge n$ 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 $Y(t) = n$.*

We are now in the position to prove Theorem 15, 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 of Theorem 15.** The theorem follows from Lemma 22 if its assumption holds with high probability. The latter follows from Lemma 20 if the exponent $\frac{1}{2}\left(-\frac{\Delta^2}{(n-\Delta)} + \max\{A(0),B(0)\}\right)$ is in $\Omega\left(-\max\{A(0),B(0)\}\right)$. We next show that this is the case.

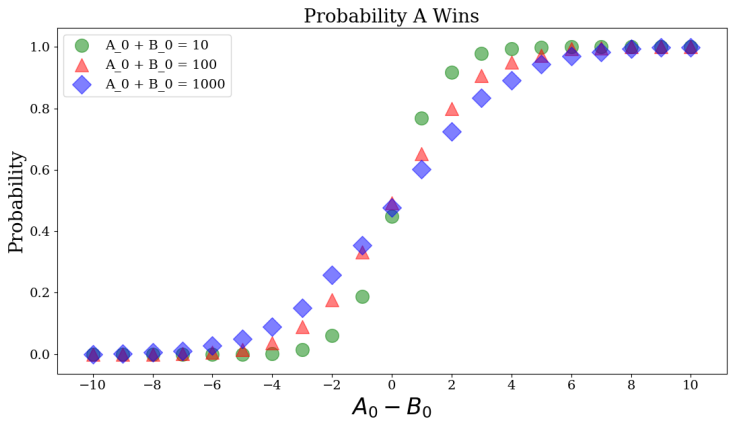


**Figure 5** The probability that species A survives while species B goes extinct is sharply dependent on their initial difference in population count $A_0 - B_0$. The sharpness of the transition is inversely proportional to initial population size $A_0 + B_0$.

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 Simulations

## 5.1 A-B Protocol Simulations

Simulations corresponding to the A-B protocol complement the theoretical results above. The A-B protocol is simulated in Figure 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 is primarily dependent on the difference in initial population size $A_0 - B_0$. Larger populations are only slightly less sensitive to the difference: Figure 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 A-B protocol is explored over its reaction rate constants and initial conditions in Figure 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 proportional. The off-diagonal initial concentrations converge faster for larger population sizes since the absolute difference in concentrations is larger.

## 5.2 *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

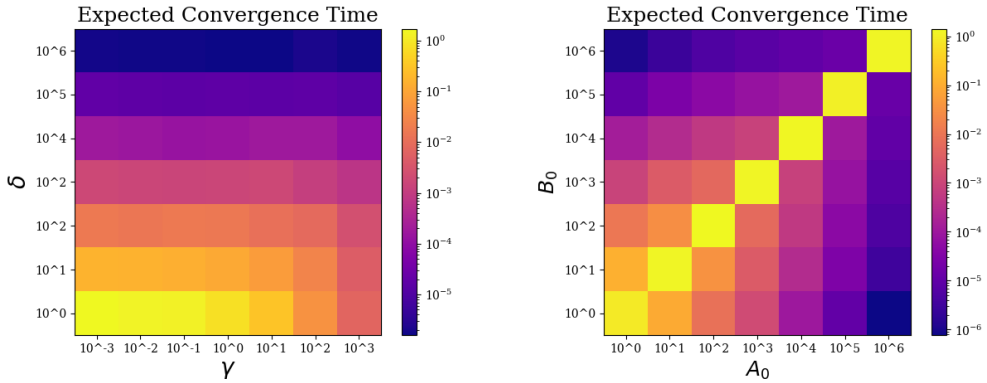


**Figure 6** Log-scaled expected convergence time of the A-B protocol is represented by color. Corresponding values are shown on the adjacent vertical bar. **Left:** rate constants $\gamma$ and $\delta$ with $A_0 = B_0 = 100$. **Right:** initial populations sizes with $\gamma = 0.01$ and $\delta = 1$.

between a sender and receiver bacterial cell [19]. 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 [23].

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. 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 [9, 19]. 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. Reaction rate constants for duplication $\gamma = 0.016$ and for conjugation $\delta = 10^{-11}$ have been taken from Dimitriu *et al.* [9]. 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 $Y^1$ and two reactions lead to $Y^0$.

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 [12, 14]. 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, at least 80% of the output species are correct at all times for all input choices.

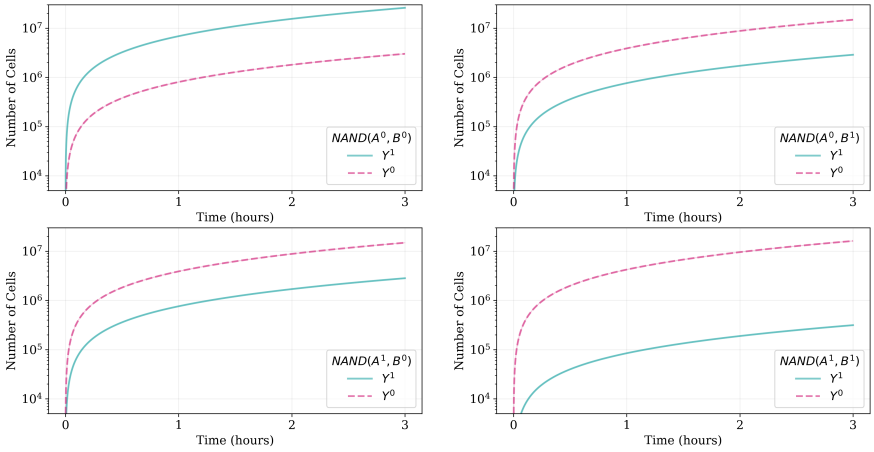


**Figure 7** A biologically plausible implementation of a NAND gate with amplifiers on inputs and outputs. Initial population size is $5 \times 10^8$, carrying capacity of $10^9$ cells. Reaction rate constants were set to $\gamma = 0.016$ and $\delta = 10^{-11}$ [9]. The output species is shown for each choice of inputs. The initial input error is 10%. For all inputs, at least 80% of the output species are correct at all times. Confidence intervals from 30 sample simulations are smaller than the width of the lines.

## 6 Conclusions

We considered the majority consensus problem with continuous population growth in a stochastic setting, and established the A-B 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 A-B protocol are two species $A$ and $B$ with an initial total population size $n = A(0) + B(0)$ and an initial gap $\Delta = |A(0) - B(0)|$. We showed that the A-B 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.

We further demonstrated how to use dual-rail gates to implement digital circuits computing arbitrary Boolean functions. As opposed to thresholds of a single species, dual-rail encoding is particularly useful in our birth systems as the A-B 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 A-B protocol, layer by layer, we finally arrive at computing the circuit's outputs.

Simulations show that the qualitative behavior of our protocols matches the behavior expected from the asymptotic bounds. While the studied A-B protocol and the NAND gate protocol are simplifications of biological implementations of consensus and gate evaluation protocols, we believe that our results give a signpost for future research on the successful implementation of complex distributed systems such as indirect inter-cellular communication via phages. We discussed a potential biological implementation based on communication by conjugation among engineered *E. coli*.

## References


1. 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. URL: https://doi.org/10.1016/j.jmb.2005.10.076, doi:10.1016/j.jmb.2005.10.076.
2. Dana Angluin, James Aspnes, and David Eisenstat. Fast computation by population protocols with a leader. *Distributed Computing*, 21(3):183–199, 2008.
3. Dana Angluin, James Aspnes, and David Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, March 2008. URL: https://doi.org/10.1007/s00446-008-0059-z, doi:10.1007/s00446-008-0059-z.
4. L Billard. Competition between two species. *Stochastic Processes and their Applications*, 2(4):391–398, 1974.
5. Pierre Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, Heidelberg, 1999.
6. Anne Condon, Monir Hajiaghayi, David Kirkpatrick, and Ján Maňuch. Approximate majority analyses using tri-molecular chemical reaction networks. *Natural Computing*, 19:249–270, 2020.
7. Rachel Cummings, David Doty, and David Soloveichik. Probability 1 computation with chemical reaction networks. *Natural Computing*, 15:245–261, 2016.
8. Ramiz Daniel, Jacob R. Rubens, Rahul Sarpeshkar, and Timothy K. Lu. Synthetic analog computation in living cells. *Nature*, 497(7451):619–623, May 2013. URL: https://doi.org/10.1038/nature12148, doi:10.1038/nature12148.
9. 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.
10. William Feller. Die Grundlagen der volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung (1939). In *Selected Papers I*, pages 441–470. Springer, 2015.
11. Matthias Függer, Manish Kushwaha, and Thomas Nowak. Digital circuit design for biological and silicon computers. In *Advances in Synthetic Biology*, pages 153–171. Springer, 2020.
12. Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.
13. 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.
14. 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. URL: https://doi.org/10.1093/bioinformatics/btl485, doi:10.1093/bioinformatics/btl485.
15. Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 2 edition, 1975.
16. David G Kendall. Branching processes since 1873. *Journal of the London Mathematical Society*, 1(1):385–406, 1966.
17. Eric Libby and William C Ratcliff. Ratcheting the evolution of multicellularity. *Science*, 346(6208):426–427, 2014.
18. Nancy A Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.
19. John P Marken and Richard M Murray. Addressable," packet-based" intercellular communication through plasmid conjugation. *bioRxiv*, page 591552, 2019.
20. 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. URL: https://doi.org/10.1038/nature11516, doi:10.1038/nature11516.
21. Chris J Myers. *Asynchronous Circuit Design*. John Wiley & Sons, 2001.
22. 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.

23 Monica E Ortiz and Drew Endy. Engineered cell-cell communication via DNA messaging. *Journal of Biological Engineering*, 6(1):16, December 2012. URL: https://doi.org/10.1186/1754-1611-6-16, doi:10.1186/1754-1611-6-16.

24 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. URL: https://doi.org/10.1038/nature12051, doi:10.1038/nature12051.

25 Amir Pandi, Mathilde Koch, Peter L Voyvodic, Paul Soudier, Jérôme Bonnet, Manish Kushwaha, and Jean-Loup Faulon. Metabolic perceptrons for neural computing in biological systems. *Nature Communications*, 10(1):1–13, 2019.

26 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.

27 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, December 2010. URL: https://doi.org/10.1038/nature09679, doi:10.1038/nature09679.

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

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

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

31 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.

32 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.

33 David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, February 2008. URL: https://doi.org/10.1007/s11047-008-9067-y, doi:10.1007/s11047-008-9067-y.

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

35 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.

36 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. URL: https://doi.org/10.1038/nature09565, doi:10.1038/nature09565.

37 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.

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