An Allee-based distributed algorithm for microbial whole-cell sensors

Check for updates

Fabricio Cravo[1,2], Matthias Függer [1]✉ & Thomas Nowak [1,3]✉

Reliable detection of substances present at potentially low concentrations is a problem common to many biomedical applications. Complementary to well-established enzyme-, antibody-antigen-, and sequencing-based approaches, so-called microbial whole-cell sensors, i.e., synthetically engineered microbial cells that sense and report substances, have been proposed as alternatives. Typically these cells operate independently: a cell reports an analyte upon local detection.

In this work, we analyze a distributed algorithm for microbial whole-cell sensors, where cells communicate to coordinate if an analyte has been detected. The algorithm, inspired by the Allee effect in biological populations, causes cells to alternate between a logical 0 and 1 state in response to reacting with the particle of interest. When the cells in the logical 1 state exceed a threshold, the algorithm converts the remaining cells to the logical 1 state, representing an easily-detectable output signal. We validate the algorithm through mathematical analysis and simulations, demonstrating that it works correctly even in noisy cellular environments.

Numerous disease indicators are based on detecting that the abundance of a particular substance exceeds a threshold concentration[1–3]. Widely-adapted techniques are sequencing for genetic information and antibody-based detection for proteins[4]. Recently microbial whole-cell sensors (MWCS), i.e., cells engineered to sense and report substances, emerge as an easy-to-use and cost-effective alternative to these classical detection methods[5]. MWCS have been demonstrated to successfully sense pollutants[6], detect inflammation in mice models[7], and provide means for environmental monitoring[8–10].

To detect analytes at low concentrations, MWCS use techniques such as optimizing the cellular sensing circuitry[7,11] and sensing multiple, correlated analytes[7]. These approaches are examples for *local sensor designs*, where engineered cells locally sense and report analytes. In this case, the population-level readout is obtained as the cumulative single-cell responses, which inherently limits the population-level response: Assume a population C of n cells, a small fraction $\alpha \in [0, 1]$ of which detect the analyte of interest, and an ideal local cell response $\text{out}_c(\text{in}_c)$ that maps the presence ($\text{in}_c = 1$) or absence ($\text{in}_c = 0$) of a detection event at cell c to a local cell output. In presence of an ideal local cell response that outputs a (normalized) 1 if $\text{in}_c = 1$ and 0 otherwise, the population-level response out_{pop} is given by

$$\text{out}_{\text{pop}} = \tfrac{1}{n} \sum_{c \in C} \text{out}_c(\text{in}_c) = \alpha, \tag{1}$$

i.e., remains linear in the fraction α of cells that detect the analyte.

By contrast, *population-based designs* use communication between cells to potentially achieve improved threshold-like population responses. In terms of the example before this is achieved by allowing out_c to depend not only on in_c, but also on the other cells' (communicated) inputs: the population-level response

$$\text{out}_{\text{pop}} = \tfrac{1}{n} \sum_{c \in C} \text{out}_c(\text{in}_1, \dots, \text{in}_n) \tag{2}$$

is not necessarily proportional to α in this case.

A commonly used mechanism to communicate is via quorum sensing (QS) molecules[12,13], which diffuse through the cell membrane and, if present in sufficiently high concentrations, allow cells to trigger a population-level response[8–10]. For example, the circuit by Hsu, Chen, Hu, and Chen[14] uses QS for population-level signal amplification: when cells sense metal ions, they start to secrete the QS molecule. This molecule, then diffuses into the surrounding medium and inside the population's cells. If a cell's internal concentration exceeds a certain threshold, a reporting pathway is triggered. Further examples for population-based designs based on QS are the detection of mercury[15] and phenolic compounds[11].

In this article, we analyze a simple distributed algorithm that acts as a distributed amplification circuit and, together with a local sensory and reporting circuit, yields a population-based MWCS design. In our

¹LMF, Université Paris-Saclay, CNRS, ENS Paris-Saclay, Gif-sur-Yvette, France. ²LISN, Université Paris-Saclay, CNRS, Gif-sur-Yvette, France. ³Institut Universitaire de France, Paris, France. ✉e-mail: mfuegger@lmf.cnrs.fr; thomas@thomasnowak.net

algorithm, cells transition from a low (L) state to a high (H) state upon reacting with rare-event substances of interest. When a specific number of cells successfully alter their state within a predetermined time frame, the algorithm guarantees the production of an amplified reporter signal, indicating the presence of rare events. Conversely, if this threshold is not met, the reporter signal is guaranteed to remain low and is not amplified at population-level. Given the high noise levels in biological circuits[16], the threshold is designed to mitigate the number of false positives compared to naive broadcasting methods, offering noise protection.

The algorithm is inspired by the Allee effect[17] as observed in biological populations: While in populations that compete for a shared resource, lower densities are supposedly more likely to thrive[18], the Allee effect describes the phenomenon that the fitness of small populations often decreases, e.g., due to the reliance on cooperation strategies within the population[19-21]. The algorithm is designed such that the population of H cells shows an Allee-like behavior: for low H densities, the "birth" rate, i.e., the rate by which L cells are transformed into H cells, is compensated by their "death" rate, i.e., the rate by which H cells are transformed into L cells. Above a certain threshold cell density, the situation is reversed, and the birth rate outweighs the death rate.

Further, differently than other QS amplification circuits[11,14,15], our algorithm autonomously maintains the amplified state indefinitely. This is achieved through a positive feedback loop, where the signal responsible for amplification promotes itself.
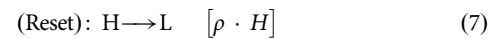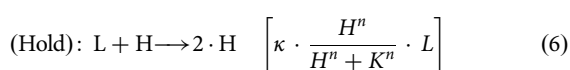
Results

The concentration of the analyte of interest is denoted by $A(t) \in \mathbb{R}_+$. Cells of the detection algorithm are in either of two states: low (L), voting for the absence of the analyte of interest, or high (H), voting for its presence. We denote the density of cells in L at time t by $L(t) \in \mathbb{R}_+$, and the density of cells in H by $H(t) \in \mathbb{R}_+$. We write $P(t) = H(t) + L(t)$ for the total population size. Assuming that the population size is within a steady state, we neglect replication and cell-death reactions.

A naive algorithm to obtain a non-linear population-level response would be to broadcast any detection of the analyte by a cell to all other cells. Such an algorithm comprises of two reactions

$$L + A \longrightarrow H + A \tag{3}$$

$$L + H \longrightarrow 2 \cdot H \tag{4}$$

the first of which models detection of an analyte by a cell and the second the broadcast of such an event to all other cells. However, as we show later, this algorithm does not tolerate incorrect detections, i.e., cells that incorrectly transition from state L to state H in absence of an analyte; a problem any biological implementation will necessarily have.

To address the problem of the broadcasting algorithm to deal with faulty state transitions, we propose an algorithm that tolerates erroneous detection of the analyte up to a certain rate. Our algorithm comprises of three reactions that determine when a cell switches state: (i) Reaction (Detect): A cell in state L changes to state H upon local detection of the analyte. We assume that this happens with a rate $\sigma_A \in \mathbb{R}_+$. (ii) Reaction (Hold): A cell in state L also switches to state H with a rate that depends on the density H according to a Hill function with parameters $\kappa \in \mathbb{R}_{>0}$, $K \in \mathbb{R}_{>0}$, and $n \in \mathbb{R}$ with $n > 1$. The Hill function models the fact that this reaction is triggered by a QS molecule which is secreted by cells in the H state; see, e.g., ref. 22 for a Hill-function model of a QS circuit. (iii) Reaction (Reset): A cell in state H switches back to state L with a certain *reset rate* $\rho \in \mathbb{R}_{>0}$. Intuitively, this is to prevent accumulation of incorrect detection events in the system.

$$(\text{Detect}): L + A \longrightarrow H + A \quad [\sigma_A \cdot A \cdot L] \tag{5}$$

$$(\text{Hold}): L + H \longrightarrow 2 \cdot H \quad \left[\kappa \cdot \frac{H^n}{H^n + K^n} \cdot L\right] \tag{6}$$

$$(\text{Reset}): H \longrightarrow L \quad [\rho \cdot H] \tag{7}$$

Cells may incorrectly detect the analyte with a rate $\sigma_{err} \in \mathbb{R}_{>0}$, accounted for in the additionally reaction1

$$(\text{Error}): L \longrightarrow H \left[\sigma_{err} \cdot L\right] \tag{8}$$
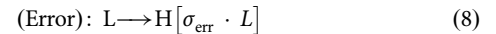
For the purpose of analysis, unless stated otherwise, we will subsume both (5) and (8) into the single reaction

$$(\text{Set}): L \longrightarrow H \quad [\sigma \cdot L] \tag{9}$$
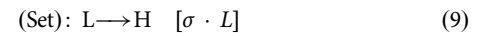
calling $\sigma = \sigma_A \cdot A + \sigma_{err}$ the *rare-event detection rate*.

Figure 1 illustrates the algorithm's reactions (Fig. 1a) and three modes of operation: in absence of the analyte (Fig. 1b), in its presence (Fig. 1c), and after detection of the analyte with the analyte potentially being absent (Fig. 1d). Observe that the high density of H cells is maintained in the latter case.

Figure 2 visualizes the Allee-like behavior of H cells: Fig. 2a, b show the "birth" rate of H cells, i.e., the sum of the rates in (6) and in (9), as well as the "death" rate of H cells, i.e., the rate in (7), over the density of H cells. In the absence of the analyte (Fig. 2a), the rare-event detection rate σ is low, and stable steady states for the H density are either low (close to $0\,\mathrm{mL}^{-1}$) or high (about $1.5 \cdot 10^8\,\mathrm{mL}^{-1}$). We show later that the low-density steady state is reached if the initial H density was below a threshold, and the high-density steady state if it was above this threshold, thus guaranteeing the memory effect for a once detected analyte. Conversely, in presence of the analyte, and a consequently high rare event detection rate σ, the H cell density converges to a high value (Fig. 2b).

We now argue correctness of the algorithm. We can write an ordinary differential equation (ODE) for the cell densities H and P from reactions (9), (7), and (6), obtaining

$$\frac{dH}{dt} = \sigma \cdot (P - H) + \kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H. \tag{10}$$

Using our proof strategy, we can establish the correctness of the Allee-based algorithm by showing that under certain conditions of its parameters κ, K, n, and the total population size P, the algorithm guarantees: (i) convergence to a low density of H cells if the rare event detection rate σ is below a critical rate σ_c and the initial density of H cells is low (no memorized detection happened), and (ii) to a high density of H cells either if the initial population of H cells was high (a detection was memorized), or the rare event rate σ exceeds the critical threshold rate σ_c (the analyte is being detected).

Theorem 1. If $\max_{H \in [0,P]}(\kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H) > 0$, then there exist α_i, α_f in $]0,1[$ with $\alpha_i < \alpha_f$ and $\sigma_c > 0$ such that:
- If $\sigma < \sigma_c$, there exists a critical point $H_{c,\sigma}$ such that:
  1. If $H(0) < H_{c,\sigma}$, then $H(t)$ converges to a value in $[0, \alpha_i P]$.
  2. If $H(0) > H_{c,\sigma}$, then $H(t)$ converges to a value in $[\alpha_f P, P]$.
- If $\sigma > \sigma_c$, then $H(t)$ converges to a value in $]\alpha_f P, P[$.
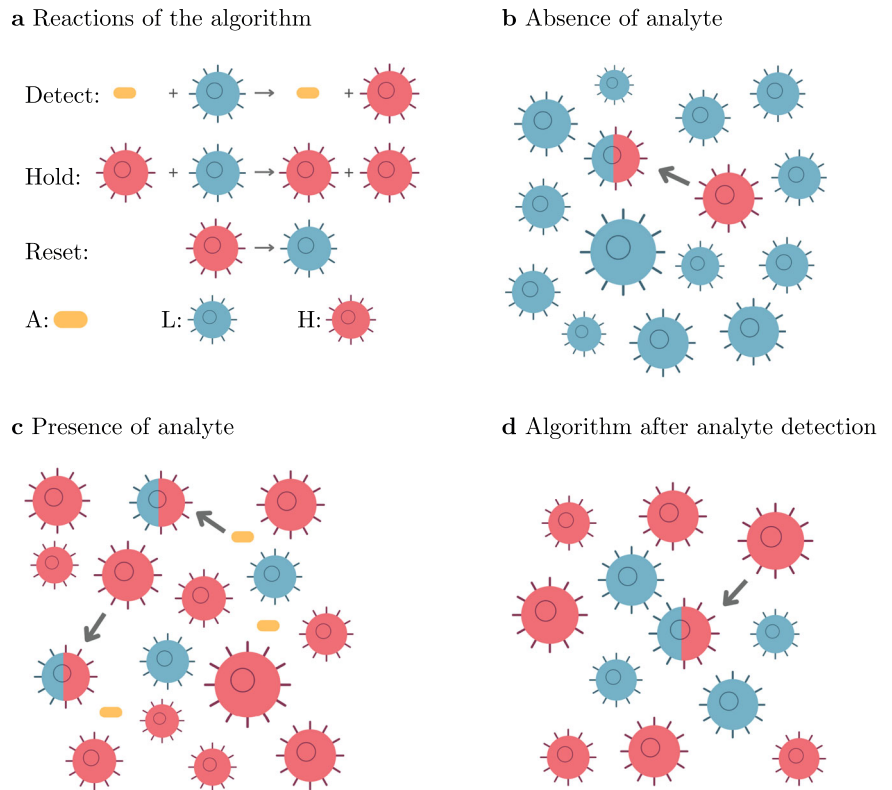
The following two corollaries immediately follow from Theorem 1 and establish the correctness of detection (Corollary 1) and memorization of a previously detected analyte (Corollary 2).

Corollary 1. (Detection). If the conditions for Theorem 1 hold, and with α_i, α_f and σ_c as defined in Theorem 1, if $H(0) = 0$ then $H(t)$ converges to a value in $[0, \alpha_i P[$ if $\sigma < \sigma_c$, and to a value in $[\alpha_f P, P[$ if $\sigma > \sigma_c$.
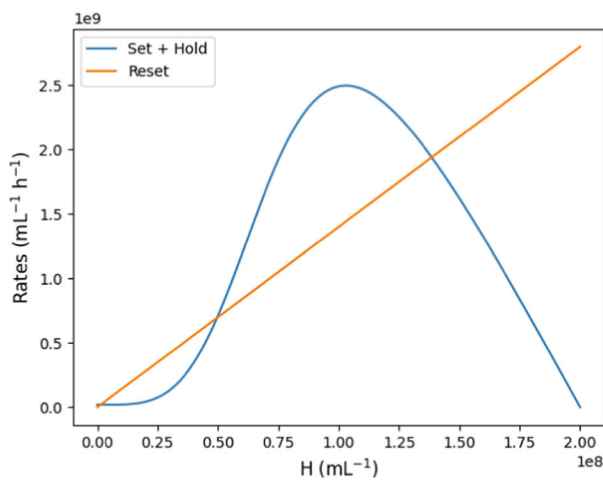
Corollary 2. (Memory). If the conditions for Theorem 1 hold, and under the notation for Theorem 1, if $H(0) > \alpha_i P$, then $H(t)$ converges to a value in $[\alpha_f P, P[$.

We finally establish an upper bound on the time the algorithm needs to converge to a high density of H cells in presence of an analyte. The proof is
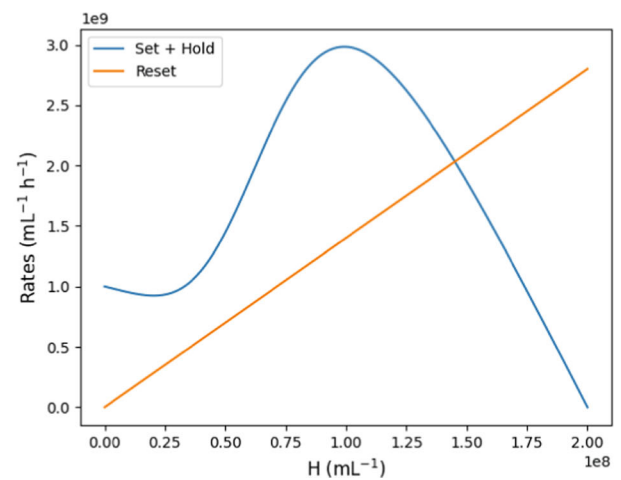
Fig. 1 | Illustration of the algorithm. a The three reactions (5), (6), and (7). **b** Cell population in absence of the analyte. A small amount of cells are incorrectly converted from L to H, but the population remains mostly in the L state. **c** Cell population in presence of a sufficiently high concentration of the analyte. Interaction with the analyte converts L cells into H cells. Additionally H cells convert other L cells into H cells. **d** State of the population after the analyte has been detected and was potentially removed thereafter. The high density of H cells is maintained by H cells continuously converting L cells into H cells.

a Reactions of the algorithm

Detect:
Hold:
Reset:
A: L: H:

b Absence of analyte

c Presence of analyte

d Algorithm after analyte detection

a Analyte absent ($\sigma = 0.1\ \mathrm{h}^{-1}$)

b Analyte present ($\sigma = 5\ \mathrm{h}^{-1}$)

Fig. 2 | Allee-like behavior of the H cells. a, b Birth rate (9) + (6) of H cells (orange) and death rate (5) of H cells (blue) versus H density. The characteristic s-shape of a birth rate in a population with Allee effect is visible. **a** Rates in absence of the analyte ($\sigma = 0.1\ \mathrm{h}^{-1}$). **b** Rates in presence of the analyte ($\sigma = 5\ \mathrm{h}^{-1}$). Parameters used: $\kappa = 35\ \mathrm{h}^{-1}$, $\rho = 14\ \mathrm{h}^{-1}$, $P = 2 \cdot 10^8\ \mathrm{mL}^{-1}$, $n = 4$.

given in the Supplementary Material, for which the convergence time is established.

Theorem 2. (Convergence time). Let $\sigma > \sigma_c$. Let $\max_{H \in [0,P]}(\kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H) > 0$. Let $H_{c,0}$ be the second lowest non-negative root solution of $\kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H = 0$. If $\max_{H \in [0,P]}(\kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H) > 0$, there exists a time $t_l = \beta_{\sigma_1} \ln(1 - R_{\sigma_1})$ with β_σ and R_σ functions of σ, such that $H(t) > H_{c,0}$ for any $t > t_l$.

To validate that the algorithm performs well within realistic parameter ranges, we estimated parameters and ran simulations from a potential genetic circuit implementation (Fig. 3). Following previous QS circuit

designs in synthetic biology[22,23], we use an N-acyl homoserine lactone (AHL) as the QS molecule. The AHL molecule is synthesized by LuxI under the control of a promoter (p1 in Fig. 3) that is activated by the binding of an LuxR- AHL complex. LuxR is consituently expressed by the circuit (not shown in the figure). Additionally, the detection of the analyte by the cell is assumed to activate promoter p1.

We next outline how this circuit implements the Allee-based algorithm: The algorithm's cell states L and H model cells with low internal LuxI, respectively, high internal LuxI concentrations. Reaction (9) models the fact that an anlyte leads to an increasing internal concentration of LuxI, thus converting an L cell to an H cell. Also, LuxI is degraded and diluted within
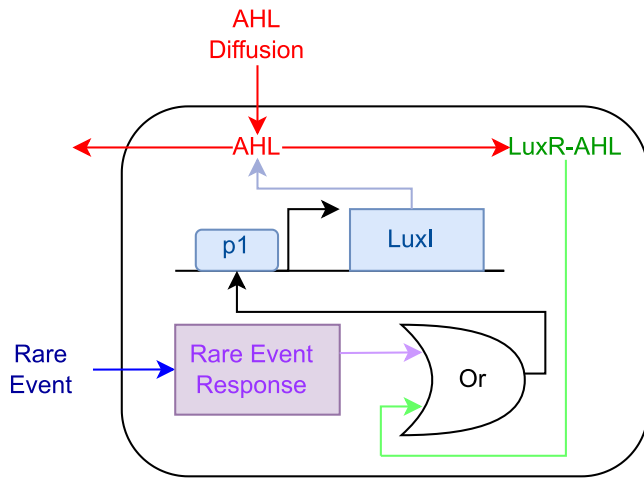
Fig. 3 | Genetic circuit implementation of the Allee-based algorithm. Both the AHL and the rare event can transform the cell from the state L to H by triggering the expression of LuxI. LuxR is constitutively expressed (circuit not shown).

Table 1 | Reaction and population parameters used in simulations

Parameter	Value	Source
κ	35 h^{-1}	Din et al. [22]
ρ	14 h^{-1}	Din et al. [22]
n	4	Din et al. [22]
K	$8 \cdot 10^7$ mL^{-1}	Smith et al. [24]
P	$1.5 \cdot 10^8$ mL^{-1}	assumed $> 2K$

the cell, accounting for reaction (7). Finally, H cells synthesize AHL that diffuses into the medium and from there into surrounding cells. The so-formed LuxR-AHL complex consequently activates promoter p1 that shows a Hill-type activation profile. The promoter's activation again leads to expression of LuxI, making an L cell switch to an H cell, as required by reaction (6).

For simulations we parametrized the algorithm's reactions with rate parameters from literature (Table 1). For the previously discussed implementation, κ corresponds to the expression of LuxI controlled by p1. From the model by ref. 22 [supplementary information], we choose $\kappa = 35$ h^{-1}. The reaction rate constant ρ corresponds to the degradation rate constant of LuxI and was set to 14 h$^{[-122]}$. The Hill coefficient n for activation of p1 via LuxR-AHL was set to 4[22]. The threshold parameter K for the activation via LuxR-AHL was set to a relatively high value of $8 \cdot 10^7$ mL^{-1}, reported in a circuit by Smith and Schuster[24]. The total cell density was set to a value larger than 2-times the threshold K, and well in the range of reachable E. coli cell densities; we used $P = 1.5 \cdot 10^8$ mL^{-1}. The parameters in Table 1 fulfill the condition of Theorem 1.

The critical threshold rate σ_c was determined through binary search. We start from the interval $I = [0, \sigma_{utr}]$, where $\sigma_{utr} \cdot (P - H) + \kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H = 0$ has only one solution for H. From Theorem 1, the equation $\kappa \frac{H^n}{H^n + K^n}(P - H) - \rho H = 0$ necessarily has three solutions for H. We then repeatedly determine the midpoint of the interval, and verify if the above equation has three or one solution with the midpoint as σ. In case of three solutions, we replace the left bound of I by the midpoint, and in case of one solution we replace the right bound by the midpoint. In case two solutions or a sufficient precision is reached, the search terminates. We refer the reader to the Supplementary Material for more details. For our setting, we find that $\sigma_c \approx 3.04$ h^{-1}, about 22% of the system's lowest rate constant, which is ρ.

The parameters in Table 1 were used to obtain the transient and steady-state simulation results. Figure 4a shows $\frac{dH}{dt}$, i.e., the net birth rate of H cells, versus H for different rare event detection rates σ. While larger values of σ lead to a single equilibrium points, lower values (0 h^{-1} and 1.5 h^{-1} in the figure) result in three equilibrium points. Of the three points, the smallest and the largest are stable and are seen to have different H densities for different σ: the smallest equilibrium point corresponds to a negative detection result and the largest to a positive detection result. Further, choosing $\sigma = 3.04$ h^{-1} close to the critical σ_c, results in a net birth rate function that barely touches the x-axis.

We next ran transient simulations for the same σ values as in Fig. 4a over a time range of 0.6 h simulated time (Fig. 4b). One observes, the convergence of $H(t)$ to a high density of H cells for a σ of 4 h^{-1} and 5 h^{-1}, and to a low density of H cells for a σ of 0 h^{-1} and 1.5 h^{-1}; in agreement with Theorem 1. The transient simulation for $\sigma = 3.04$ h^{-1} close to the critical σ_c does not visibly converge in the simulated time. Increasing the simulation time, however, shows that it converges to a high H density.

he memorization of the presence of an analyte that has been removed thereafter. In Fig. 4c we varied the exposure time of the cells to the analyte, while in Fig. 4d the concentration of the analyte was varied. The memorization above a certain critical rate $\sigma_c(t) = \sigma_A \cdot A(t)$ is observed in both cases, which is consistent with Corollary 2.

We finally ran simulations for an extended duration to determine steady-state values for different settings of rare event detection rates σ (Table 2). The results are consistent with Theorem 1. When σ is above the critical rate σ_c, larger values of σ lead to faster convergence of $H(t)$ to its steady state. As shown in Table 2, the highest convergence time is 0.61 h, which occurs for $\sigma = 3.04$ h^{-1}, a value close to σ_c. For a slightly higher $\sigma = 4$ h^{-1}, the convergence time is already reduced to 0.44 h. For a mathematical analysis of the convergence times, we refer the reader to the Supplementary Material (Lemmas 7 and 19).
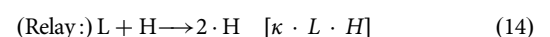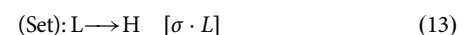
To demonstrate the effectiveness of the Allee-based algorithm, we compare its performance to other algorithms, including two natural adaptations and one presented by ref. 14. In the comparison we focus on the thresholding behavior of the algorithms: ideally the detection algorithm shows a strong amplification of its detection output around a threshold concentration of the analyte, below of which the output is strong negative, and above of which it is strong positive.

A natural simplification of the Allee-based algorithm is to remove the (6) reaction, and only keep the reactions that transform cells to H cells in presence of the analyte, as well as reset H cells to L cells with a certain reset rate ρ:

$$\text{(Set): } L \longrightarrow H \quad [\sigma \cdot L] \tag{11}$$

$$\text{(Reset): } H \longrightarrow L \quad [\rho \cdot H] \tag{12}$$
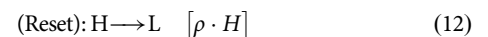
Steady-state analysis of H cells via setting $\frac{dH}{dt} = 0$ and subsequent algebraic manipulation yields, $H = \frac{\sigma}{\rho} P / (1 + \frac{\sigma}{\rho}) = P \frac{\sigma}{\rho + \sigma}$ as the unique steady-state. Since it is unique, the algorithm lacks the possibility to memorize previous presence of the analyte. Further, for most applications we expect σ to be small compared to the other rates (and in particular ρ), implying a low amplification from the analyte concentration A to the output H.

A natural distributed algorithm that solves the problem of detecting an analyte is to broadcast any detection of the analyte to all other cells that relay this broadcast. Here, relay is obtained by a cell in state H that had been informed of the presence of the analyte, to pass this information to any L cell it interacts with. In terms of reactions, and referring to $\kappa > 0$ as the broadcasting rate, this algorithm can be written as
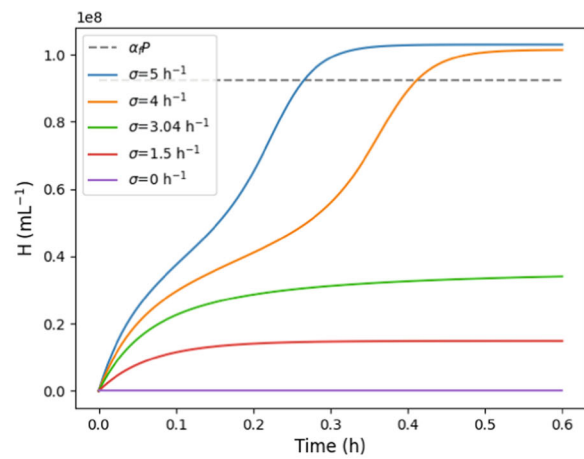
$$\text{(Set): } L \longrightarrow H \quad [\sigma \cdot L] \tag{13}$$

$$\text{(Relay :) } L + H \longrightarrow 2 \cdot H \quad [\kappa \cdot L \cdot H] \tag{14}$$

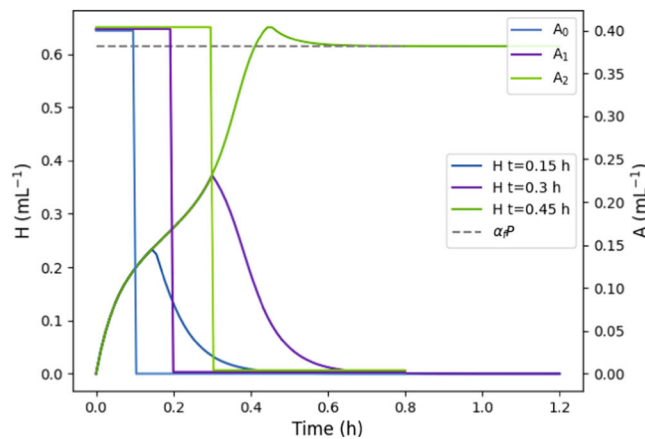and its dynamics are $\frac{dH}{dt} = \sigma \cdot (P - H) + \kappa H(P - H)$.

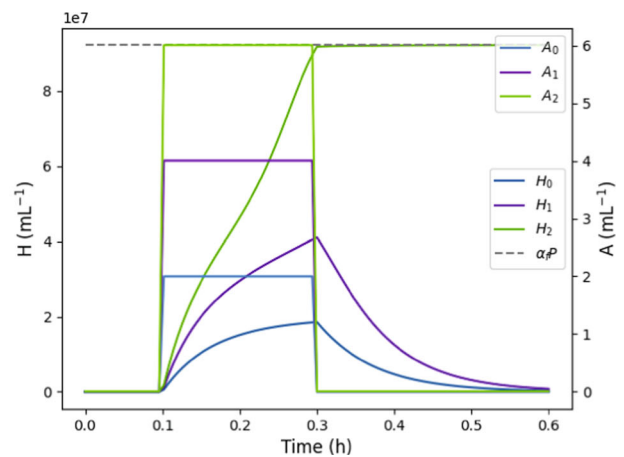**a** Net birth rate of H cells over H for different σ
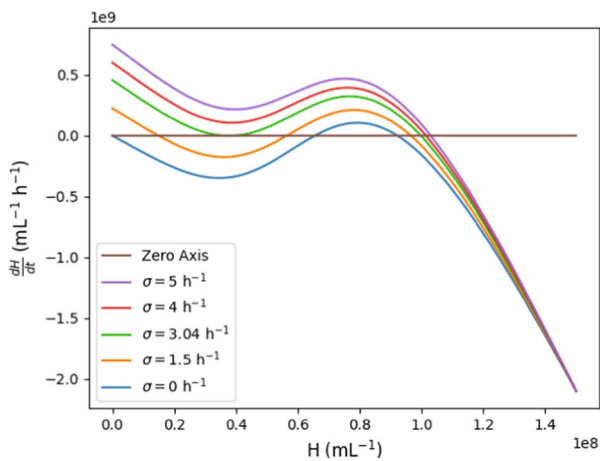
b Transient behavior of $H(t)$ for different σ

c $H(t)$ for different analyte presence times

d $H(t)$ for different analyte concentrations

Fig. 4 | Simulation results. a Plot of the net birth rate H cells (dH/dt) versus H for different σ. Equilibrium points of H are the roots of dH/dt. **b** Plot of $H(t)$ over time t. For values of σ bellow the critical σ_c, $H(t)$ converges to H cell densities below $\alpha_i P$ (consistent with Corollary 1). For values of σ above σ_c, $H(t)$ converges to H cell densities above $\alpha_f P$ (again, consistent with Corollary 1). For $\sigma = 3.05\ h^{-1}$, the plotted time range is not enough to show convergence of $H(t)$. **c** Plot of $H(t)$ and the analyte concentration $A(t)$ over time t. The analyte concentration was chosen as a negative step function with different durations until the step. One observes that below a certain exposure time of the cells to the analyte, the analyte is not memorized by the algorithm. Above this exposure time, the detection of the analyte is memorized even if it removed thereafter (consistent with Corollary 2). **d** Plot of $H(t)$ and the analyte concentration $A(t)$ over time t. The analyte concentration was chosen as a pulse of varying amplitude. One observes that below a certain amplitude, the detection is not memorized, while above it is memorized.

Since $0 \leq H \leq P$, one has that H is bounded. We now perform a case distinction between the cases where $\sigma > 0$ and $\sigma = 0$. **Case** $\sigma = 0$: For any $H \in [0, P[$, one has $\frac{dH}{dt} > 0$. From the monotonicity of H and its boundedness, it follows that H converges to a finite steady-state. Since $\frac{dH}{dt} = 0$

when $H = P$, one has that $H(t)$ converges to P. **Case** $\sigma > 0$: For any $H \in]0, P[$, one has $\frac{dH}{dt} > 0$. If $H(0)$ is in $]0, P]$, from the monotonicity of H in $]0, P]$ and its boundedness, it follows that H converges to a finite steady-state. Since $\frac{dH}{dt} = 0$ when $H = P$, one has that $H(t)$ converges to P. If $H(0) = 0$, since, one has that $H(t) = 0$ for all t. Indeed, two steady states, obtained by setting $\frac{dH}{dt} = 0$, are possible: (i) $\sigma = 0$ and $H = 0$, and (ii) $\sigma > 0$ and $H = P$.
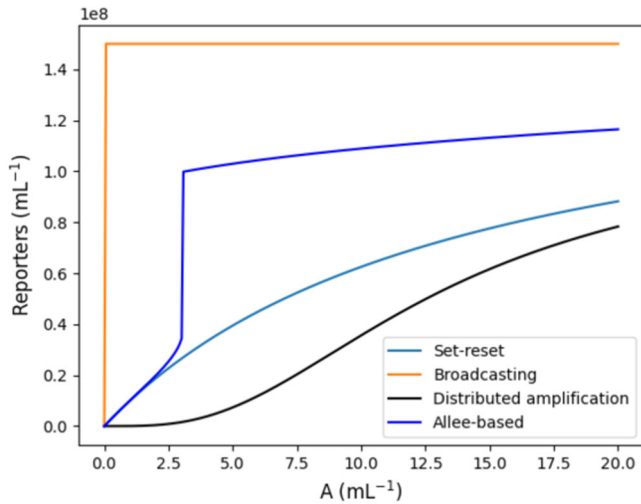
While the presence of the two steady-states shows that the algorithm can memorize previously detected analytes, the steady-states also show that the algorithm cannot tolerate incorrect transitions of L cells to H cells: for an arbitrarily small σ, all cells switch to state H.

Ref. 14 proposed a circuit that uses distributed amplification via a QS pathway: Cells that detect the analyte synthesize the QS molecule. Similar to the model for the Allee-based algorithm, we abstract this via two cell types: L cells with low internal concentrations of LuxI and H cells with high concentrations of LuxI. Any cell whose QS threshold is triggered, expresses a reporter molecule S (e.g., YTP). In terms of reactions, this algorithm is

Table 2 | Steady-state density of $H(t)$ and convergence times for different rare event detection rates

Rate σ	Steady-state of H	Convergence time
$0.00\ h^{-1}$	$0\ mL^{-1}$	$0\ h$
$1.50\ h^{-1}$	$0.15 \cdot 10^8\ mL^{-1}$	$0.21\ h$
$3.04\ h^{-1}$	$0.36 \cdot 10^8\ mL^{-1}$	$0.61\ h$
$4.00\ h^{-1}$	$1.01 \cdot 10^8\ mL^{-1}$	$0.44\ h$
$5.00\ h^{-1}$	$1.03 \cdot 10^8\ mL^{-1}$	$0.3\ h$

σ Convergence times are given as the time until 95% of the steady-state density is reached.

Fig. 5 | Comparison of the four algorithms: Allee-based, Set-reset, broadcasting, and distributed amplification algorithm. The plot shows steady-state reporter densities (S respectively H) over analyte concentrations A obtained via simulation. Simulation parameters: $P = 1.5 \cdot 10^8$ mL^{-1}, $\rho = 14$ h, $\kappa = 35$ h, $K = 8 \cdot 10^7$ mL^{-1}, and $n = 4$.
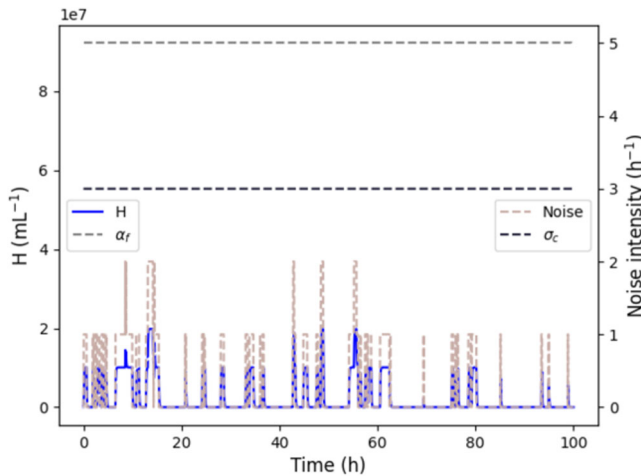
Fig. 6 | Robustness of the Allee-based algorithm to an incorrect detection of the analyte. The plot shows $H(t)$ and $\sigma(t)$ over time t for 4 h simulated time with a low, stochastic rate σ in absence of an analyte. The algorithm is seen to not incorrectly amplify the detection of the analyte: the H cell density H remains low throughout the simulation. Simulation parameters: $P = 1.5 \cdot 10^8$ mL^{-1}, $\rho = 14$ h, $\kappa = 35$ h, $K = 8 \cdot 10^7$ mL^{-1}, $n = 4$, $\beta = 0.5$, and $\gamma = 2$. The parameters α_f (see Theorem 1) and σ_c are shown as horizontal lines.

expressed as

$$(\text{Set}): \text{L} \longrightarrow \text{H} \quad [\sigma \cdot L] \tag{15}$$

$$(\text{Reset}): \text{H} \longrightarrow \text{L} \quad [\rho \cdot H] \tag{16}$$

$$(\text{Signal-Secretion}): \forall C \in \{L, H\}: \text{C} + \text{H} \longrightarrow \text{C} + \text{H} + \text{S} \quad \left[\kappa \cdot \frac{H^n}{H^n + K^n} \cdot C\right] \tag{17}$$

$$(\text{Signal-Decay}): \text{S} \longrightarrow \emptyset \quad [\rho \cdot S] \tag{18}$$

where S is the reporter molecule, and the reaction (18) accounts for the decay of S.

To compare the thresholding behavior of the distributed amplification, the Set-reset, and the broadcasting algorithm to the Allee-based algorithm, we ran simulations in presence of the same analyte concentration A for all four algorithms (Fig. 5). Simulation parameters where chosen identically for similar reactions. One observes the strong amplification of the Allee-based algorithm around a non-zero critical concentration of A. The distributed amplification algorithm shows a threshold behavior, but with a weaker amplification. The broadcasting algorithm has a strong threshold at 0, and the Set-reset algorithm shows no thresholding behavior.

The non-zero thresholding behavior around a critical threshold σ_c as shown in Theorem 1 and demonstrated by simulations in Fig. 5 suggests that the Allee-based algorithm is robust to incorrectly detected analytes by L cells ($\sigma_{err} > 0$ in our model). To demonstrate that this is the case, also for stochastically varying incorrect detections as expected in a real genetic circuit implementation, we ran simulations where we stochastically varied σ over time, simulating the effect of a stochastic σ_{err} in absence of an anlyte. For any such simulation, an ideal algorithm is expected to keep H cell densities low, and thus not wrongly signal the detection of an analyte.

Simulations were carried out in MobsPy[25] with a simulated time of 1000 h. For the stochastic model of wrongly detected analytes we chose a stochastic birth-death process of H cells with a birth rate β of 0.5 h^{-1} and a death rate γ of 2 h^{-1}. The parameters have been set to the leaky expression rate and the decay rate of LacI[22] to reflect a realistic parameter range for leaky expression resulting in incorrect L to H transitions of a cell. The so-obtained stochastic rates were then fed into a deterministic transient-time simulation of the Allee-based algorithm. Figure 6 shows the resulting densities $H(t)$ as well as the rate $\sigma(t)$ over time t for a simulated time of 100 h. We can observe that while the stochastic event detection rate is capable of increasing the density $H(t)$, the algorithm does not amplify the H cells further, thus preventing the cells from incorrectly detecting the analyte.

To examine the impact of parameters like the population density P, as well as the rate parameter κ from (6) and ρ from (7), on the algorithm's critical threshold σ_c, we determined σ_c for parameter sweeps (Fig. 7). The parameter ranges that violate the condition of Theorem 1 are marked with setting $\sigma_c = 0$ in the heat map.
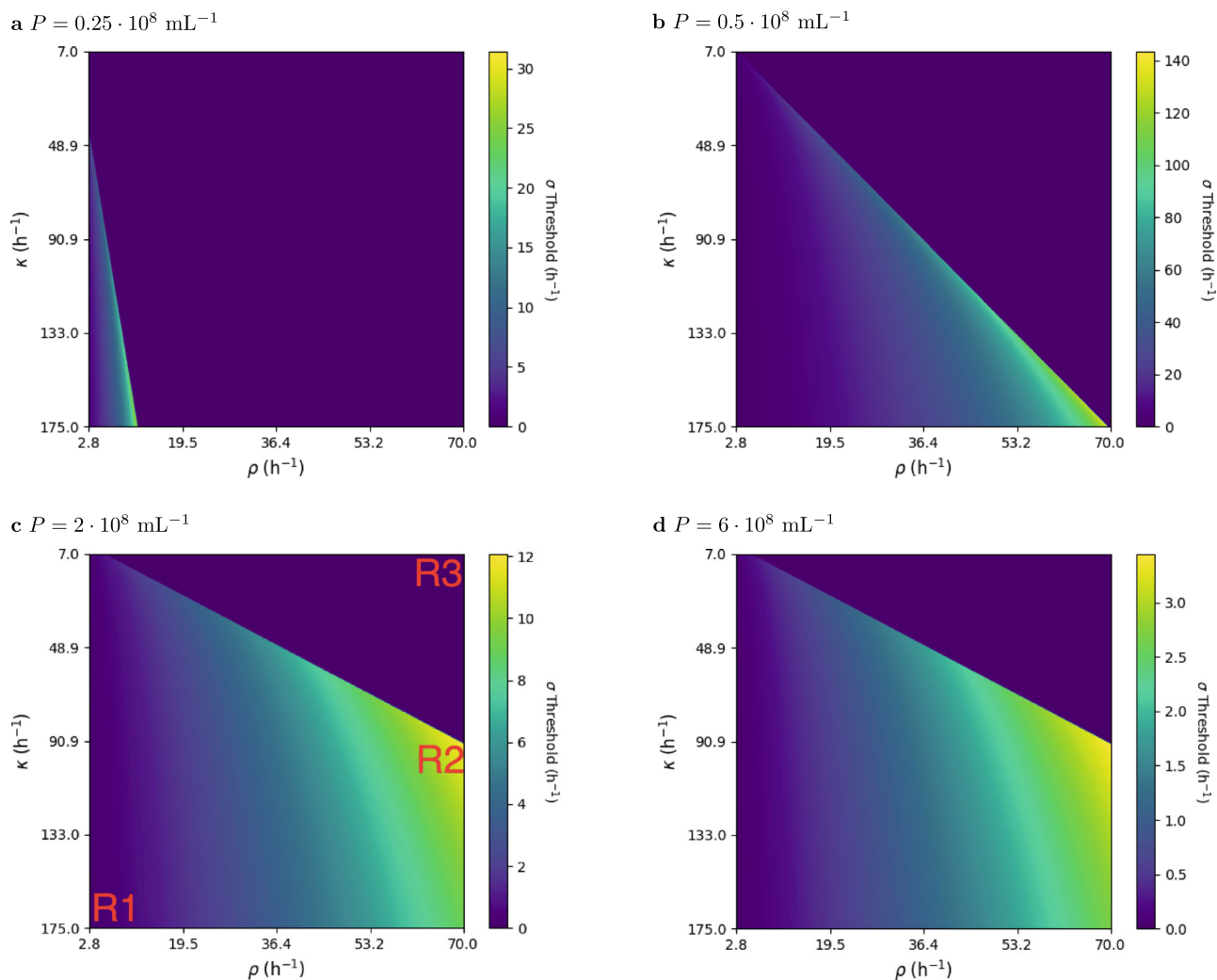
In all the heat maps, we can observe a distinct linear boundary separating a region where $\sigma_c = 0$ indicating a violation of the condition of 1 and a valid region. For instance, in Fig. 7c, the region marked as $R3$ falls in this category where the reset rate constant ρ is significantly larger than the hold rate constant κ, leading to the inability of the amplification process to trigger. Consequently, the region possesses only a stable equilibrium state with $H = 0$ and no stable equilibrium state for H with a high cell density. This is expected as the (7) reaction dominates the system behavior in this case.

The heat maps also reveal that when the (6) rate parameter κ is higher than the (7) rate constant ρ, the critical threshold σ_c is often close to zero. This is evident in the region marked as $R1$ in Fig. 7c. Due to the higher hold rate, the algorithm can produces wrong positives for a lower σ.

Discussion

We presented and discussed a distributed algorithm to detect rare events, such as the presence of a rare analyte, by a population of engineered cells. The algorithm is intended to be used in combination with sensory and reporting circuits within MWCS. The algorithm is inspired by the Allee effect observed in natural systems: it uses the fact that a certain critical threshold cell density that signal the presence of an analyte is hard to reach initially, but once it is reached, the fact that an analyte has been detected it quickly propagated to the whole cell population.

We have established conditions under which the algorithm provably works as intended (Theorem 1). Numerical simulations of a proof-of-concept circuit demonstrate that the algorithm shows strong amplification of near a critical threshold concentration of the analyte (Fig. 5). This is in contrast to three other natural algorithms that have been discussed in this work (Fig. 5). Additionally, hybrid stochastic–deterministic simulations (Fig. 6) were carried out to demonstrate the robustness of the algorithm to cells that wrongly detect the analyte.

a $P = 0.25 \cdot 10^8 \text{ mL}^{-1}$

b $P = 0.5 \cdot 10^8 \text{ mL}^{-1}$

c $P = 2 \cdot 10^8 \text{ mL}^{-1}$

d $P = 6 \cdot 10^8 \text{ mL}^{-1}$

Fig. 7 | Parameter variations showing σ_c for different reaction rate parameters κ and ρ. Subplots (**a**–**d**) differ in the total population density P.

Since the detection thresholds and the targeted total cell populations may differ significantly per application, one may need to adapt the reaction parameters for these cases. We speculate that the simplicity of the detection algorithm as well as the mechanistic understanding of all parameters greatly simplifies this adaption, e.g., via plasmid copy number manipulation to affect the reaction rates[26] and gene removal to alter the parameter K[24]. We leave the impact of inter-cellular differences and changing population sizes to future work.

Finally, as has been show by the robustness simulations, the choice of the critical threshold rate σ_c balances the capability of sensing rare events and robustness: while a low threshold favors early detection, a high threshold tolerates a larger concentration of wrongly detected analytes.

Methods
Model construction
The model was constructed using the ODE CRN formulation[27]. In this formulation, considering no inflow and outflow of matter into the system, the rate of change over time for the concentration/density $X(t)$ for any species X is expressed as:

$$\frac{dX(t)}{dt} = \text{Generative reaction rates of X} - \text{Consuming reaction rates of X} \quad (19)$$

Here, the "generative reaction rates of X" refer to the rates of reactions where X is produced times its stoichiometry in each reaction, while the

"consuming reaction rates of X" refer to the rates of reactions where X is consumed times its stoichiometry in each reaction.

For this model, reaction rates from Reactions (5) and (9) are defined using mass-action kinetics[28], which assumes that the reaction rate is the result of the product of the reactants' concentration and a constant reaction rate. Conversely, Reaction (6) uses the Hill formulation[29], which accounts for cooperativity among multiple ligand binding sites.

Taking X as H, writing the ODE associated with the proposed chemical reaction network by inserting the proposed rate expressions, and assuming a constant steady-state population value such that $L = P - H$ yields Equation (10).

Proof strategy
Theorem 1 is proven with the following strategy: From Equation (10), algebraic manipulation of $\frac{dH}{dt} = 0$ yields up to three potential equilibrium points for H, the smallest of which is stable, the middle one unstable, and the largest one being stable (if they exist). Using the monotonicity of $\frac{dH}{dt}$ within certain subdomains of H, one can show that for the interval I between the first, stable, and the second, unstable, equilibrium point, $\frac{dH}{dt}$ is negative for $H \in I$, and $I = \emptyset$ for rare event detection rates σ larger than a critical rate σ_c. The proof we give in the Supplementary Material is based on the intermediate value theorem. For sufficiently large values of σ, only one stable equilibrium point remains and its value can be bounded away from 0. The proof in the Supplementary Material uses a quadratic Lyapunov function to show convergence to this fixed point.

Theorem 2 is proven with the following strategy: We begin by establishing a lower bound ordinary differential equation (ODE), $\frac{dH^*}{dt} < \frac{dH}{dt}$, known for its exponential convergence to a fixed point. In parallel, we define an interval I starting at zero. If, at any point in time t_c, the value of $H(t_c)$ falls outside this interval, it indicates that $H(t)$ will converge to the amplified state. Furthermore, leveraging the ODE's lower bound and the initial conditions $H^*(0) = H(0) = 0$, if there exists a specific time instance t_l where $H^*(t_l)$ exits the I interval, it implies that $H(t_l)$ must also leave this interval.

Simulation model

We used the Python simulation framework MobsPy[25] to obtain transient and steady-state simulation results for the parameters in Table 2. The core of the simulation code is shown in Listing 1.

The robustness simulation was separated into two parts. First, we generated a birth and death process to simulate noisy event rates introduced to the system, as shown in code Listing 2.

Subsequently, the generated data was incorporated into another model using simulation events, which represent changes in species values during simulation. Its core code is shown in Listing 3.

```
# s = sigma, k = kappa, and p = rho

L, H = BaseSpecies(2)
L >> H [s] # Set
H >> L [p] # Reset
L + H >> 2 * H [lambda l, h: f'{k}*{l}*1/(1 + ({K}/{h})
^{n})'] # Hold
H(0), L(P) # initially, H(0) = 0 and L(0) = P

MySim = Simulation(L|H)
```

Listing 1: MobsPy simulation code for the Alle-effect-based algorithm. Initialization of reaction and population parameters is according to Table 1. Parameter σ (s in the code) was varied in the simulations.

```
A = BaseSpecies(1)
Zero >> A [pr]
A >> Zero [dr]

MySim = Simulation(E)
MySim.simulation_method = 'stochastic'
```

Listing 2: MobsPy simulation code for the birth and death noise generation. The simulation parameters were set to pr = 0.5 and dr = 2.

```
with open('noise.pkl', 'rb') as file:
        noise = pickle.load(file)
L, H, A = BaseSpecies(3)

L + A >> H + A [1]
H >> L [p]
L + H >> 2 * H[lambda l, h: f'{k}*{l}*1/(1 + ({K}/
{h})^{n})']

L(P)
S1 = Simulation(H|L|A)
for time, data in zip(noise['Time'], noise['Data']):
    with S1.event_time(time):
        A(data)
S1.simulation_method = 'stochastic'
```

Listing 3: MobsPy simulation code for the robustness test. Initialization of reaction and population parameters is according to Table 1. The noise.pkl file contains the results from Listing 2.

All simulations were run in Python version 3.10 and MobsPy version 2.2.0. The hardware used was a MacBook Air with 1,1 GHz Quad-Core Intel Core i5, Intel Iris Plus Graphics 1536 MB, and 8 GB 3733 MHz LPDDR4X. The operating system was Mac OS.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated from the simulations was generated by the code available at https://github.com/BioDisCo/alle_effect_for_rare_event.

References

1. Anker, P., Mulcahy, H., Qi Chen, X. & Stroun, M. Detection of circulating tumour dna in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev.* **18**, 65–73 (1999).
2. Vestergaard, J. et al. Hedgehog signaling in small-cell lung cancer: frequent in vivo but a rare event in vitro. *Lung Cancer* **52**, 281–290 (2006).
3. Zimmerlin, L., Donnenberg, V. S. & Donnenberg, A. D. Rare event detection and analysis in flow cytometry: bone marrow mesenchymal stem cells, breast cancer stem/progenitor cells in malignant effusions, and pericytes in disaggregated adipose tissue. *Flow Cytometry Protoc.* 251–273 https://link.springer.com/protocol/10.1007/978-1-61737-950-5_12 (2011).
4. Sidransky, D. Emerging molecular markers of cancer. *Nat. Rev. Cancer* **2**, 210–219 (2002).
5. Andreescu, S. & Sadik, O. A. Trends and challenges in biochemical sensors for clinical and environmental monitoring. *Pure Appl. Chem.* **76**, 861–878 (2004).
6. Belkin, S. Microbial whole-cell sensing systems of environmental pollutants. *Curr. Opin. Microbiol.* **6**, 206–212 (2003).
7. Woo, S.-G. et al. A designed whole-cell biosensor for live diagnosis of gut inflammation through nitrate sensing. *Biosens. Bioelectron.* **168**, 112523 (2020).
8. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
9. Dong, Y.-H. & Zhang, L.-H. Quorum sensing and quorum-quenching enzymes. *J. Microbiol.* **43**, 101–109 (2005).
10. Pu, L., Yang, S., Xia, A. & Jin, F. Optogenetics manipulation enables prevention of biofilm formation of engineered pseudomonas aeruginosa on surfaces. *ACS Synth. Biol.* **7**, 200–208 (2018).
11. He, J., Zhang, X., Qian, Y., Wang, Q. & Bai, Y. An engineered quorum-sensing-based whole-cell biosensor for active degradation of organophosphates. *Biosens. Bioelectron.* **206**, 114085 (2022).
12. Wu, Y., Wang, C.-W., Wang, D. & Wei, N. A whole-cell biosensor for point-of-care detection of waterborne bacterial pathogens. *ACS Synth. Biol.* **10**, 333–344 (2021).
13. Moraskie, M. et al. Microbial whole-cell biosensors: current applications, challenges, and future perspectives. *Biosens. Bioelectron.* **191**, 113359 (2021).
14. Hsu, C.-Y., Chen, B.-K., Hu, R.-H. & Chen, B.-S. Systematic design of a quorum sensing-based biosensor for enhanced detection of metal ion in escherichia coli. *IEEE Transact. Biomed. Circ. Syst.* **10**, 593–601 (2016).
15. Cai, S. et al. Engineering highly sensitive whole-cell mercury biosensors based on positive feedback loops from quorum-sensing systems. *Analyst* **143**, 630–634 (2018).
16. Macía, J., Posas, F. & Solé, R. V. Distributed computation: the new wave of synthetic biology devices. *Trends Biotechnol.* **30**, 342–349 (2012).

17. Allee, W. & Bowen, E. S. Studies in animal aggregations: mass protection against colloidal silver among goldfishes. *J. Exp. Zool.* **61**, 185–207 (1932).

18. Courchamp, F., Berec, L. & Gascoigne, J. *Allee effects in ecology and conservation* (OUP Oxford, 2008).

19. Berec, L., Angulo, E. & Courchamp, F. Multiple allee effects and population management. *Trends Ecol. Evol.* **22**, 185–191 (2007).

20. Mooring, M. S., Fitzpatrick, T. A., Nishihira, T. T. & Reisig, D. D. Vigilance, predation risk, and the allee effect in desert bighorn sheep. *J. Wildl. Manag.* **68**, 519–532 (2004).

21. Clutton-Brock, T. et al. Predation, group size and mortality in a cooperative mongoose, suricata suricatta. *J. Animal Ecol.* **68**, 672–683 (1999).

22. Din, M. O. et al. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85 (2016).

23. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).

24. Smith, P. & Schuster, M. Antiactivators prevent self-sensing in pseudomonas aeruginosa quorum sensing. *Proc. Natl. Acad. Sci.* **119**, e2201242119 (2022).

25. Cravo, F., Függer, M., Nowak, T. & Prakash, G. Mobspy: a meta-species language for chemical reaction networks. In *International Conference on Computational Methods in Systems Biology*, 277–285 (Springer, 2022).

26. Tsao, K.-L. & Waugh, D. S. Balancing the production of two recombinant proteins inescherichia coliby manipulating plasmid copy number: high-level expression of heterodimeric ras farnesyltransferase. *Protein Expr. Purif.* **11**, 233–240 (1997).

27. Feinberg, M. *Foundations of chemical reaction network theory* (Springer, 2019).

28. Horn, F. & Jackson, R. General mass action kinetics. *Arch. Ration. Mech. Anal.* **47**, 81–116 (1972).

29. Weiss, J. N. The hill equation revisited: uses and misuses. *FASEB J.* **11**, 835–841 (1997).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41540-024-00363-3.

Correspondence and requests for materials should be addressed to Matthias Függer or Thomas Nowak.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.