A population-based temporal logic gate for timing and recording chemical events

Victoria Hsiao[1,*], Yutaka Hori[2], Paul W. K. Rothemund[3], Richard M. Murray[1,2]

October 26, 2015

1. Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125.
2. Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125.
3. Computation & Neural Systems, California Institute of Technology, Pasadena, CA 91125.

* Corresponding Author: Victoria Hsiao, vhsiao at caltech dot edu.

Running Title: Two-input integrase event detector

Character count (including spaces): 40,557 (not including fig. legends) / 47,637 (including fig. legends)

Abstract

Single-cell bacterial sensors have numerous applications in human health monitoring, environmental chemical detection, and materials biosynthesis. Such bacterial devices need not only the capability to differentiate between combinations of inputs, but also the ability to process signal timing and duration. In this work, we present a two-input temporal logic gate that can sense and record the order of the inputs, the timing between inputs, and the duration of input pulses. The temporal logic gate design relies on unidirectional DNA recombination with bacteriophage integrases to detect and encode sequences of input events. When implemented in a chromosomally-modified *E. coli* strain, we can utilize stochastic single cell responses to predict overall heterogeneous population behavior. We show that a stochastic model can be used to predict final population distributions of this *E. coli* strain, and thus that final differentiated sub-populations can be used to deduce the timing and duration of transient chemical events.

Introduction

Engineered bacteria could one day be powerful self-replicating single-cell sensors with environmental, health, and industrial applications. Synthetic biology has made important strides in identifying and optimizing genetic components for building such devices. In particular, much work has focused on Boolean logic gates which detect the presence or absence of static chemical signals (Gardner *et al*, 2000; Anderson *et al*, 2007; Wang *et al*, 2011; Moon *et al*, 2013; Shis *et al*, 2014) and compute a digital response.

Temporal logic gates, which process time-varying chemical chemical signals, have been much less explored. Pioneering work by Friedland *et al.* used serine integrase-based recombination for the counting and detection of sequential pulses of inducers (Friedland *et al*, 2009). But so far, no work has studied the potential for temporal logic gates to provide information about the duration of a signal, or the time between two chemical events. Here, we present a temporal logic gate which allows us to infer analog signal timing and duration information about the sequential application of two inducer molecules to a population of bacterial cells.

Similar to previous temporal logic gates, our design takes advantage of the irreversibility of serine integrase recombination. While bistable switches have been successfully deployed as memory modules in genetic circuits (Kotula *et al*, 2014), such switches require constant protein production to maintain state, and are sensitive to cell division rates and growth phase. The large serine integrases, on the other hand, reliably and irreversibly flip or excise unique fragments of DNA (Yuan *et al*, 2008). Thus logic circuits built from integrases intrinsically include DNA-level memory that requires virtually no cellular resources to maintain state, thus enabling permanent and low-cost genetic differentiation of individual bacterial cells based on transient integrase induction. Further advantages of the serine integrates include the short length (40-50 bp) and directionality of their attachment sites. Serine integrases recognize flanking DNA binding domains (attB, attP) and subsequently digest, flip or excise, and re-ligate the DNA between the attachment sites. Flipping or excision activity is determined by the relative orientation of the sites, which allows complex orientation-dependent behavior to be programmed into integrase circuits. Well-known serine integrases include Bxb1, TP901-1, and ϕC31, all of which have been used to demonstrate static-input logic gates (Siuti *et al*, 2013; Bonnet *et al*, 2013), and some have cofactors that can reverse directionality (Bonnet *et al*, 2012; Khaleel *et al*, 2011). Recently, an entirely new set of 11 orthogonal integrases was characterized, greatly expanding the set of circuits that can be built (Yang *et al*, 2014).

In contrast to previous studies of temporal logic gates, our work leverages the stochastic nature of single-cell switching to create a robust population-level response to a time-varying chemical signal. By traditional engineering standards, synthetic circuits would ideally perform identically in every cell in a population. When this ideal is applied to biology, the stochastic nature of molecular processes, particularly at low copy numbers, presents a significant barrier to reliable outputs from engineered cells. Thus while natural cellular dynamics and differentiation take advantage of noisy gene expression (Elowitz *et al*, 2002; Süel *et al*, 2007) synthetic circuits often require noise reduction for proper function (Dunlop *et al*, 2008). We designed a two-input temporal logic gate using strategically interleaved and oriented integrase (Bxb1, TP901-1) DNA recombination sites and used this gate to engineer an E. coli strain with four possible genetically-differentiated end states. This strain contained single genomic copies of the temporal logic gate, ensuring digital-yet-stochastic responses from individual cells. We then utilized the heterogeneity of individual cellular responses to encode sequences of chemical inputs into the overall population response, and use a stochastic model of single cell trajectories to predict the population response. By analyzing the distributions of final cell states, we deduce the timing and pulse duration of transient chemical pulses. Furthermore, because the states are genetically encoded, we can recover details of a chemical event long after its occurrence.
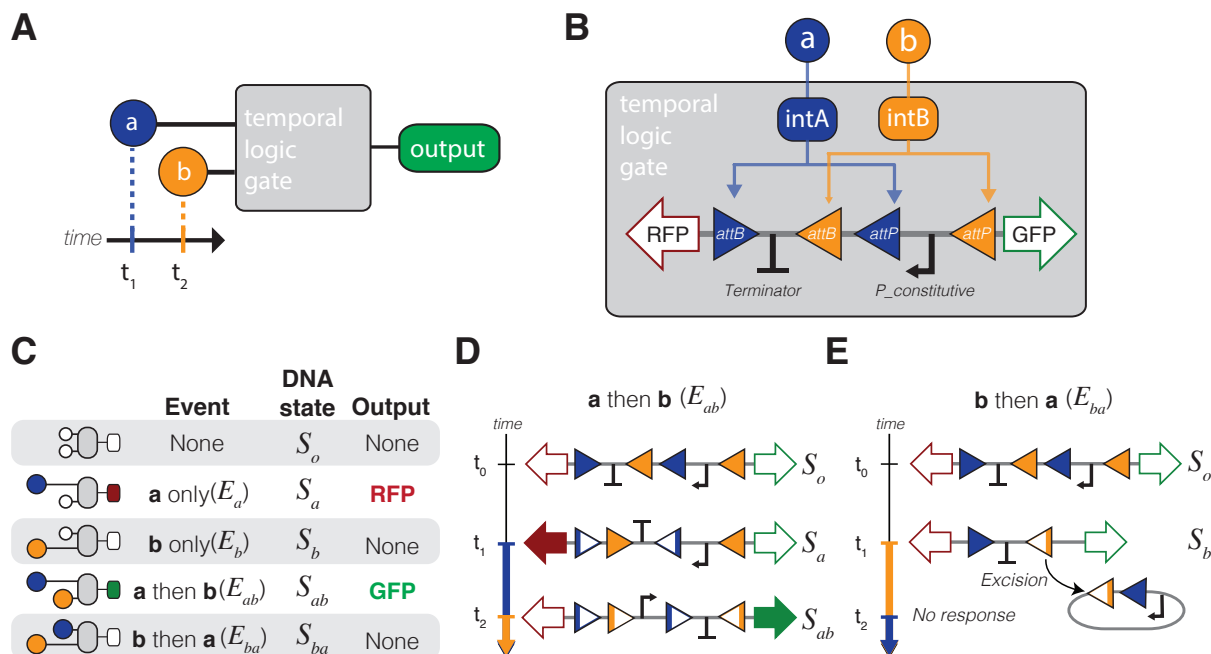
Results

Figure 1: Design overview of a temporal logic gate. A) A temporal logic gate distinguishes between two chemical inputs (**a**,**b**) with different start times. B) Implementation of the temporal logic gate using a set of two integrases with overlapping attachment sites. Chemical inputs **a** and **b** activate production of integrases intA and intB, which act upon a chromosomal DNA cassette. C) Table with all possible inputs and outcomes to the event detector. D) Sequence of DNA flipping following inputs with inducer **a** before inducer **b** (event E_{ab}). E) Sequence of DNA flipping following inducer inputs with **b** first (event E_{ba}). In any events in which **b** precedes **a**, the uni-directionality of the intB attachment sites results in excision.

Design of a two-integrase temporal logic gate

We have designed a two-input temporal logic gate that differentiates between the start times of two chemical inputs and produces unique outputs accordingly (Figure 1A). The design relies on a system of two-integrases with nested integrase attachments sites (Figure 1B). The use of integrases irreversibly invert segments of DNA, resulting in a memory feature that can be maintained for multiple generations (Bonnet *et al*, 2012).

The design of the integrase temporal logic gate hinges on interleaving the attB attachment site of integrase B (intB) with the attP site of integrase A (intA), thus ensuring that the possible DNA flipping outcomes are mutually exclusive (Figure 1B). The serine integrases used in this design are TP901-1 (intA) and Bxb1 (intB). The fluorescent proteins mKate2-RFP (RFP) and superfolder-GFP (GFP) are used as placeholders for future downstream gene activation as well as real-time readouts of the logic gate. The design also features a terminator (Bba-B0015) and a strong constitutive promoter (P7). In the case where there are no inputs, the terminator prevents expression of RFP from the constitutive promoter.

There are five possible basic events that could occur in a two-input system (Figure 1C): no input, inducer **a** only (E_a), inducer **b** only (E_b), inducer **a** followed by **b** at a later time (E_{ab}), and inducer **b** followed by **a** at a later time (E_{ba}). Consequently, in a perfectly resolved temporal logic gate there should be five unique DNA states corresponding to the five types of events: S_o (the initial state), S_a, S_b, S_{ab}, and S_{ba}. This design is limited to only four DNA states due to excision when E_b occurs ($S_b = S_{ba}$). The two fluorescent outputs

4

correspond to the two states that occur when inducer **a** is detected first – RFP is produced when the cell is in state S_a, and GFP is produced when the cell is in state S_{ab}.

Figure 1D illustrates the sequence of recombination that occurs during an event E_{ab} that results in DNA state S_{ab} and the production of GFP. Upon addition of inducer **a** at time t_1, TP901-1 flips the DNA between its attachment sites, reversing the directionality of the terminator and the Bxb1 attB recognition site (state S_a). Then, when inducer **b** is added at some time t_2 that is greater than t_1, the directionality of the Bxb1 sites is such that the DNA is flipped to reverse the directionality of the P7 constitutive promoter (state S_{ab}). If inducer **b** is added first (Figure 1E), the Bxb1 attachment sites are uni-directional, a configuration that results not in recombination, but in excision of the DNA between the sites (state S_b).

Once a DNA recombination has occurred, it is irreversible. The unique attB and attP attachment sites are recombined into attL and attR sites, respectively, and no longer recognized by the integrases. The nesting of the integrase attachment sites is the key design feature that produces the temporal ***a* then *b*** logic, and the irreversibility of the recombination records the event in DNA memory. The result is a genetic record that can both be sequenced later and immediately read by via constitutive production of fluorescent outputs.
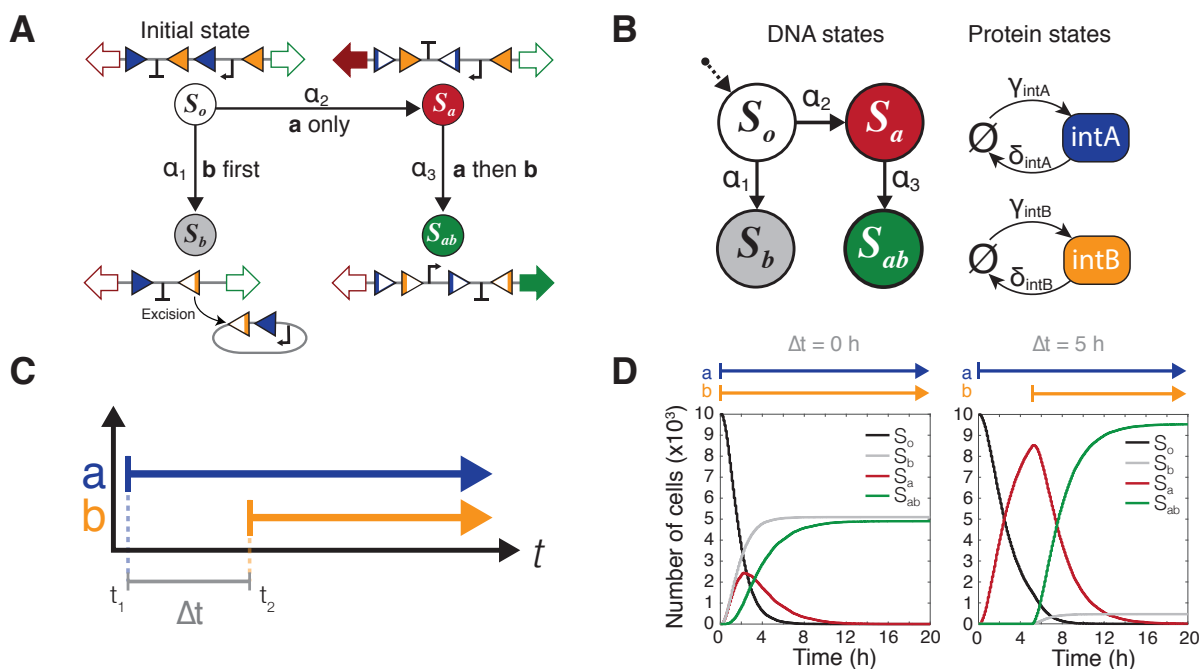

Figure 2: A Markov model of integrase-mediated DNA flipping. A) The four possible DNA states, illustrated with DNA state diagrams. All DNA begins in the initial state S_o and there are no reverse processes. The propensity functions α_1, α_2, and α_3 are dependent on the concentration of the two integrases and correspond to the events ***b* first** (E_b), ***a* only** (E_a), and ***a* then *b*** (E_{ab}), respectively. B) Representation of the same model as a Markov chain. Integrases are represented simply as protein states with production (γ_A, γ_B) and degradation (δ_A, δ_B) rates. C) Graphical representation of inducer step functions. Δt is defined as difference between the start time of the first inducer and start time of the second. D) Simulation results for inducer separation times of 0 and 5 hours. There are four possible DNA states, but all cells end up in either the S_b or S_{ab} final states. Individual trajectories are simulated for 10,000 cells and the number of cells in each DNA state are summed for each time point (Figure S1).

A Markov model for integrase recombination

We created a model of integrase-mediated DNA flipping and then used a stochastic simulation algorithm (Gillespie, 1977) to simulate individual cell trajectories (Figure 2A). All of the four possible DNA states are

represented in the model: the original state (S_o), the intB excision state (S_b), the intA single flip state (S_a), and the *a then b* double flip state (S_{ab}). We have implemented the system experimentally by chromosomally integrating the target DNA into the genome of the *E. coli* cell. This allows us to assume that each cell only has one copy of the temporal logic gate (Haldimann and Wanner, 2001), and that each cell can be characterized by the tuple $(\text{DNA}, \text{IntA}, \text{IntB})$ (Figure 2B). The DNA term is S_o, S_a, S_b, or S_{ab}, and IntA and IntB are non-negative integers representing the molecular copy number of each integrase. Once a DNA cassette has flipped into any of the states other than the original state S_o, there is no reverse process. The logic gate is designed such that if integrase B is expressed prior to integrase A, the DNA cassette is excised and the chain reaches the dead-end S_b state. In order for a cell to successfully detect E_{ab}, it first needs to switch into state S_a then transition into state S_{ab} upon addition of inducer **b**.

Since each cell contains only a single copy of the temporal logic gate DNA, we can expect each cell to behave differently, and to be highly susceptible to internal and external noise. This stochastic behavior will create a heterogeneous population response that can be analyzed for a more complex profile of event than if all the cells behaved uniformly. In order to capture the heterogeneity of cell population, we model the temporal logic gate using a stochastic model. Specifically, the stochastic transitions between the DNA states and the production/degradation of integrases are mathematically modeled by a continuous-time Markov chain over the state space $(\text{DNA}, \text{IntA}, \text{IntB})$ as illustrated in Figure 2B. Definitions of transition rates can be found in the Supplementary Information (Table S4).

In silico, the dynamics of a single cell translates to each stochastic simulation of the Markov model starting with $(\text{DNA} = S_o, \text{IntA} = 0, \text{IntB} = 0)$ state. We define $\mathbb{P}_t(S_o), \mathbb{P}_t(S_a), \mathbb{P}_t(S_b)$ and $\mathbb{P}_t(S_{ab})$ as the probability that the DNA state of a single cell is S_o, S_a, S_b and S_{ab} at time t, respectively. The temporal dynamics of the probability can be modeled by the following ordinary differential equation (ODE) (see SI for derivation).

$$\frac{d}{dt}\begin{bmatrix} \mathbb{P}_t(S_o) \\ \mathbb{P}_t(S_a) \\ \mathbb{P}_t(S_b) \\ \mathbb{P}_t(S_{ab}) \end{bmatrix} = \begin{bmatrix} -k_{\text{flipB}}\mathbb{E}_t[\text{IntB}|S_o] - k_{\text{flipA}}\mathbb{E}_t[\text{IntA}|S_o] & 0 & 0 & 0 \\ k_{\text{flipA}}\mathbb{E}_t[\text{IntA}|S_o] & -k_{\text{flipB}}\mathbb{E}_t[\text{IntB}|S_a] & 0 & 0 \\ k_{\text{flipB}}\mathbb{E}_t[\text{IntB}|S_o] & 0 & 0 & 0 \\ 0 & k_{\text{flipB}}\mathbb{E}_t[\text{IntB}|S_a] & 0 & 0 \end{bmatrix}\begin{bmatrix} \mathbb{P}_t(S_o) \\ \mathbb{P}_t(S_a) \\ \mathbb{P}_t(S_b) \\ \mathbb{P}_t(S_{ab}) \end{bmatrix}, \qquad (1)$$

where the notation $\mathbb{E}_t[\cdot|\cdot]$ stands for the conditional expected value at time t.

We define the time between the introduction of the first inducer (t_1) and the arrival of the second inducer (t_2) as the *inducer separation time* (Δt), such that

$$\Delta t = t_2 - t_1, \qquad (2)$$

as shown in Figure 2C.

In the following set of simulations and experiments, we will consider cases with step inputs (Figure 2C), where the inducers are either present or not present. Concentrations of the inducers when they are "on" will be held constant. Also, it is important to note that inducer **a** is still present during and after time Δt when inducer **b** is introduced.

Simulations of the Markov model were done with biologically plausible parameters in order to predict qualitative circuit behavior (Table S4). We limited the parameters to only the basic processes (integrase production, degradation, and DNA flipping), and parameter values were chosen to be within biological orders of magnitude. The single production rate constants, k_{prodA} and k_{prodB}, combine the transcription and translation rates of each integrase. When an integrase in the model is induced, its production rate, γ_*, is the sum of $k_{\text{prod}*}$ and any leaky transcriptional expression, $k_{\text{leak}*}$ ($* = $ intA or intB). Any individual integrase, once produced, will need to dimerize, search for the DNA binding site, bind to the DNA, form a dimer of dimers (tetramer), digest, flip, and ligate the DNA (Yuan *et al*, 2008). We have combined all of those rates into the rate constant, k_{flipA} and k_{flipB}. Parameter values for preliminary simulations were $k_{\text{prodA}} = k_{\text{prodB}} = 0.5(\mu\text{m}^3\cdot\text{hr})^{-1}$, $k_{\text{deg}} = 0.01\text{hr}^{-1}$ (69 min half-life), $k_{\text{flipA}} = k_{\text{flipB}} = 1\text{hr}^{-1}$, and $k_{\text{leakA}} = k_{\text{leakB}} = 0(\mu\text{m}^3\cdot\text{hr})^{-1}$.

Our analysis of initial numerical simulation results underscored the significant role that the inducer separation time, Δt, plays in setting the final population distributions (Figure 2D). For each Δt, individual cell trajectories (N = 10,000) were generated with the assumption that each cell only has one copy of the target DNA. Then, at every time point, the total number of cells in each DNA state was counted (Figure

6

S1). Figure 2D shows the contrast between adding both inducers simultaneously ($\Delta t = 0$h) and adding inducer **b** after a 5 hour delay ($\Delta t = 5$h). Since both inducers are present by the end of simulation, all of the cells must have a final state that is either the S_{ab} state or the S_b state. No cells remain in the original S_o configuration. S_a is a transient state that builds up prior to the addition of inducer **b** and begins to convert to S_{ab} immediately after the introduction of **b**. These initial simulation results suggest that Δt may be a way to reliably tune the final population fractions of S_{ab} versus S_b state cells.
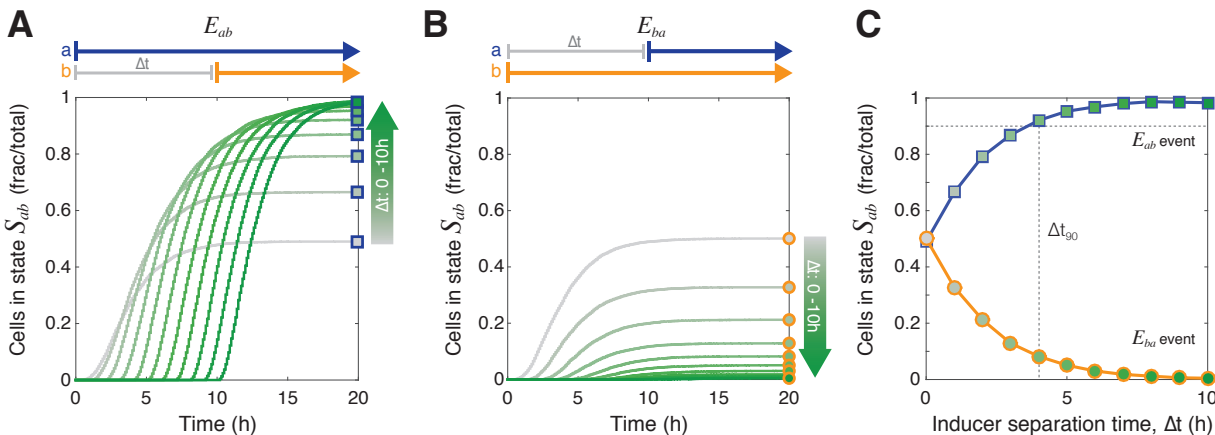

Figure 3: Simulation results for inducer separation time for $\Delta t = 0 - 10$h. A) The population fraction (N/10,000 cells) that switches into state S_{ab} following an E_{ab} event is dependent on the inducer separation time, Δt. The gray to dark green color gradient represents increasing Δt values. Square markers indicate final population fractions for specific values of Δt. B) In the case of the inverse E_{ba} event, the fraction of cells in state S_{ab} decreases proportionally to Δt. Circular markers indicate final population fractions for specific values of Δt. C) Final S_{ab} cell fractions from Figure 3A, B are plotted as a function of Δt. Blue line with square markers are endpoint population fractions from an E_{ab} event. Yellow line with circular markers are final endpoint population fractions from an E_{ba} event. The gradient inside the markers corresponds to increasing Δt value. The dotted gray line corresponds to the Δt_{90}, the value of Δt at which 90% of the cells are in state S_{ab}. All simulations were done with a population of N = 10,000 cells.

Characterization of inducer separation time

We can further investigate the effects of varying both inducer order and separation time *in silico* (Figure 3). Populations of cells have again been exposed to a sequence of overlapping step functions (N = 10,000). In the case of an E_{ab} event, the proportion of cells that successfully detect *a then b* and switch to state S_{ab} is a function of the inducer separation time, Δt (Figure 3A). Exposing cells to the inverse sequence of events, E_{ba}, results in a decrease of S_{ab} cells proportional to increasing Δt (Figure 3B). If we plot the final number of S_{ab} cells from both E_{ab} and E_{ba} as a function of Δt (Figure 3C), we see that the two curves do not overlap. This indicates that just the proportion of S_{ab} cells alone is enough to uniquely distinguish which inducer appeared first. Simulation results tracing the other three DNA states can be found in Figure S2.

Additionally, we can define a detection limit, Δt_{90}, for which the inducer separation time results in 90% of population switching into the S_{ab} state (Figure 3C). This Δt_{90} limit provides a way to capture the two response regimes of the population. If the inducer separation time is less than the detection limit ($\Delta t < \Delta t_{90}$), then the rate of population switching is fast enough such that the number of S_{ab} cells will correspond uniquely to some Δt value. If $\Delta t > \Delta t_{90}$, then most cells have already switched to a final state, and the differences in S_{ab} cell count are too small to uniquely determine Δt.

In vivo experimental results showed that population fractions of S_{ab} cells could be tuned using Δt, and verified model predictions (Figure 4). DH5α-Z1 cells were chromosomally integrated with one copy of the integrase target DNA and then transformed with a high copy plasmid containing Ptet-Bxb1 and PBAD-TP901-1. When Δt is varied from 0 to 8 hours, we observed results qualitatively similar to model
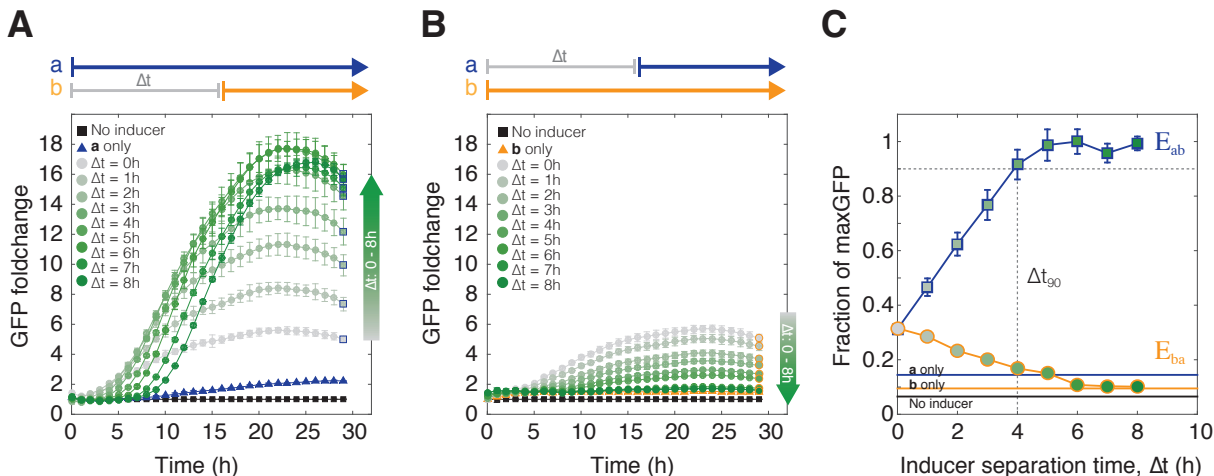
Figure 4: *In vivo* results for varying inducer separation time from Δt = 0 – 8h. A) Populations of cells exposed to an E_{ab} event sequence. Cell-switching to state S_{ab} (indicated by GFP fluorescence) begins when inducer **b** (aTc) is added. Maximum normalized GFP fluorescence increases as a function of the inducer separation time Δt. Gray to dark green gradient represents increasing Δt values. Square markers are final endpoint measurements. Error bars represent standard error of the mean. B) Cells exposed to the inverse E_{ba} sequence of events. GFP fluorescence is inversely proportional to the inducer separation time between **b** and **a**. Circular markers are final endpoint measurements. C) The final fluorescence values for Figure 4 A, B at 30 hours are plotted as a function of Δt. The final fluorescence values have been normalized to the maximum GFP expression. Dotted line marks Δt_{90} detection limit.

predictions (Figure 4). In Figure 4A, the cells have been exposed to an E_{ab} event, where inducer **a** is present from time $t = 0$ h to t_{end}, and **b** is present from $t = \Delta t$ h to t_{end}. GFP expression is a proxy for S_{ab} state cells, and total fluorescence has been normalized by a non-induced control sample. The number of cells in the GFP-expressing S_{ab} state increases proportionally with increasing Δt, and continue to be responsive even when the two inducers are separated by 8 hours. There is some expression of GFP in the presence of only inducer **a** (E_a), indicating some basal levels of intB. RFP expression, a proxy for the number of cells in state S_a, begins to increase at $t = 0$ h and drops at time $t = \Delta t$ when inducer **b** is added (Figure S3A). Aligning all of the GFP expression curves by Δt (Figure S4) shows that lower values of Δt not only have lower final GFP expression values, but also have slower rates of GFP production. This is consistent with modeling results because if we assume inducer **b** has an equal probability of entering any one cell, then in case of small Δt ($\Delta t \leq 4$ hours) there is a much larger number of S_o cells and so the rate of $S_a \rightarrow S_{ab}$ state conversion will be lower. In the case of $\Delta t > 4$ hours, the majority of cells in the population are already in the S_a state configuration, and so the rate of cell state conversion to S_{ab} will be much higher.

When cells are exposed to E_{ba}, the number of S_{ab} cells decreases proportionally to Δt (Figure 4B), and there is no RFP expression above background (Figure S3B). In both types of events, the cells maintained their state for up to 30 hours. Raw data for this set of experiments can be found in Figure S5.

We can differentiate between populations that have been exposed to E_{ab} versus E_{ba} within one hour of separation time between inducers (Figure 4C). Endpoint GFP fluorescence measurements at 30 hours from Figure 4A,B were normalized by the maximum fluorescence and plotted against Δt. As Δt increases, the GFP-expressing S_{ab} sub-population increases. The populations that encountered E_{ba} show decreasing GFP expression as Δt increases, and at $\Delta t = 6$ h, the expression is equal to the baseline expression of a **b** *only* population, indicating that the addition of inducer **a** after a 6 hour exposure to only inducer **b** has no effect at all. Based on where the GFP fraction exceeds 90% of the max expression, the Δt_{90} detection limit for the experimental system is ~ 4 hours. Finally, the baseline population split when both **a** and **b** are added simultaneously ($\Delta t = 0$ h) is not 50% of the maximum GFP expression as expected. This suggests that the integrase flipping rates, k_{flipA} and k_{flipB}, may not be equal and that the basal expression rates, k_{leakA} and k_{leakB}, are non-zero.
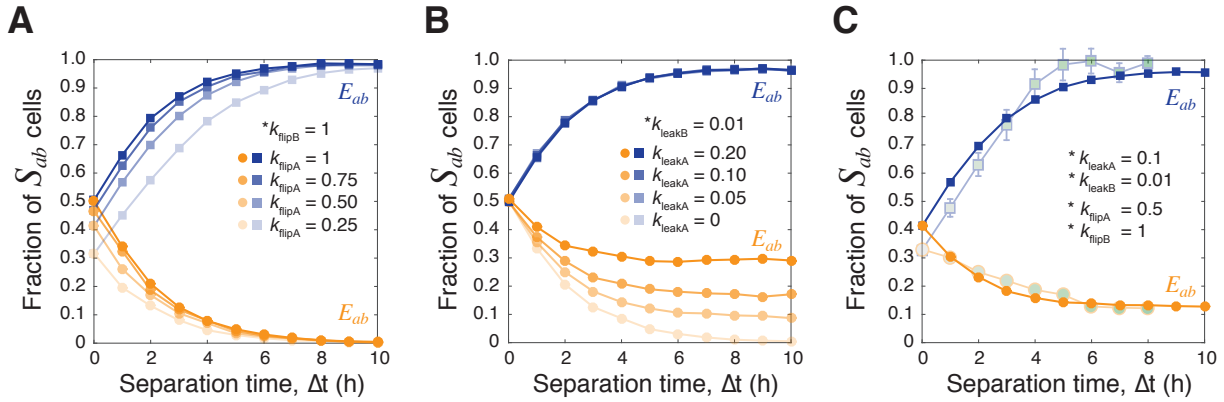
8

Figure 5: Varying model parameters for integrase flipping and leaky expression. A) As DNA flipping rates of intA (k_{flipA}) are decreased relative to k_{flipB}, the population of S_{ab} cells at $\Delta t = 0$ h has a downward shift. Simulations are done with N = 10,000 cells. B) Increasing the leaky expression of intA (k_{leakA}) changes the minimum number of cells that erroneously switch in to the S_{ab} state when exposed to E_{ba}. C) The model was revised to more closely match the experimental data (opaque). The revised parameters are $k_{\text{flipA}} = 0.5\text{hr}^{-1}$, $k_{\text{flipB}} = 1\text{hr}^{-1}$, $k_{\text{leakA}} = 0.1\text{hr}^{-1}$, and $k_{\text{leakB}} = 0.01\text{hr}^{-1}$.

Varying model parameters for integrase activity and basal expression

The parameters for integrase flipping and leaky basal expression were tuned to account for the asymmetrical population responses to E_{ab} versus E_{ba} events (Figure 4C). We hypothesized that this asymmetry arises from a combination of unequal integrase activity when searching for and flipping the DNA, and unevenness in protein expression as well as leaky background expression of the integrases (Figure 5). We varied these parameters for intA in the model relative to constant intB parameters. When the relative flipping efficiency parameters of intA ($k_{\text{flip}}A$) was decreased from 100% of intB efficiency to 0.25-0.75k_{flipB}, we observed a bias in the baseline population split when $\Delta t = 0$ h and both inducers are introduced simultaneously (Figure 5A). Previously in the preliminary model (Figure 3C), the two integrases were assigned equal flipping rates, and the population split was expected to be 50/50. As the flipping rate of intA decreases relative to that of intB, that baseline shifts downwards to favor the more active integrase, intB.

The experimental data show a small population of S_{ab} cells even when only one of the inducers is present (Figure 4C, **b** only, **a** only). When the flipping rates were held equal and the leaky expression of intA (k_{leakA}) was varied from 0 to 0.2 hr^{-1} (Figure 5B), this increased the baseline minimum number of cells that incorrectly end up in the S_{ab} state during an E_{ba} event. The fraction of S_{ab} cells for $\Delta t >6$ h have reached this threshold of minimum leaky expression.

When the combined effects of unequal integrase flipping activity and leaky expression are added to the model (Figure 5C), the separation time Δt versus S_{ab} population fraction graph generated by the model more closely matches the *in vivo* results. Experimental results (Figure 4C) have been overlaid with reduced opacity to show fit. Based on this qualitative fitting, intA appears to be less active and more leaky than intB, with parameters changed to $k_{\text{flipA}} = 0.5\text{hr}^{-1}$ and $k_{\text{leakA}} = 0.1\text{hr}^{-1}$. This suggests that to the high leaky expression of intA, around 10% of the population will "detect" E_{ab} and be in state S_{ab} even when no inducer **a** has been introduced. Although intB also has some basal level of expression, it is more difficult to ascertain the percentage of cells that erroneously switch to state S_b since S_b does not have a fluorescent output.

Finally, the Δt_{90} detection limit can be tuned by increasing or decreasing the overall production rate $k_{\text{prod}*}$ ($* = \text{A or B}$) (Figure S6). In this particular implementation of the temporal logic gate, Δt_{90} is ~ 4 hours. Within this 0 to 4 hour window, the S_{ab} population fraction can be used to uniquely determine Δt. Outside of this time window, the only assertion that can be made is that $\Delta t > 5$ hours. If $k_{\text{prod}*}$ ($* = \text{A or B}$) were higher, the integrases would accumulate faster in each cell and increase the probability of DNA flipping. Such a system would have a lower Δt_{90} but also higher resolution of events within the

narrower $t = 0$ to $t = \Delta t_{90}$ hour time window. This is because with higher protein production rates, cells commit to a final state faster, thus shrinking the effective window of time in which events could be resolved. However, within that smaller time window, S_{ab} populations fractions would also be measurably different at much smaller intervals, and so Δt could be resolved with much higher resolution. On the other end of the spectrum, if protein production were slow, the stochastic DNA recombination events happen less frequently, resulting in a population that is more sensitive to inputs for a longer period of time, but would have lower resolution overall since the population fractions are not changing as quickly.
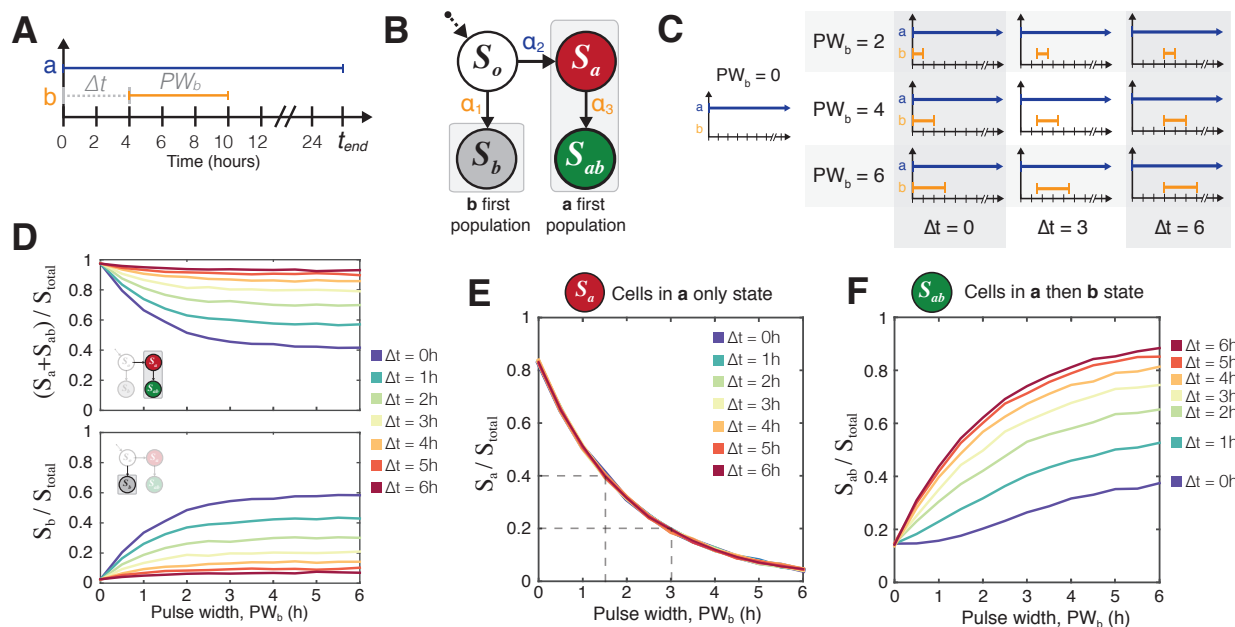

Figure 6: Simulation results for pulse width modulation. Simulations were done with revised parameters found in Figure 5C. A) Inducer **a** can be used as a reference signal against which to measure the time and duration of the inducer **b** pulse. B) The population eventually divides into one of two partitions: those that see inducer **a** first and those that see inducer **b** first. Only if a cell has entered the *a first* pathway does it have the possibility to express RFP or GFP. Furthermore, S_a can be thought of as a necessary precursor to S_{ab}. C) A matrix illustrating a subset of the Δt and PW_b values to be tested. D) Simulation results show that for any given Δt, the number of cells in S_b = total number of cells $- (S_a + S_{ab})$ E) The fraction of the population in the S_a state is totally independent of Δt and depends only on the pulse duration of inducer **b**. F) Once PW_b is known, then the fraction of the population in S_{ab} state can be used to find the time at which the pulse of inducer **b** began.

Deducing inducer pulse width

Using the fraction of GFP-expressing S_{ab} cells alone, we can determine Δt values up to a Δt_{90} limit for any given sequence of two step inputs. Now consider a pulse type of event, in which inducer **a** begins at time $t = 0$ h and remains constant throughout and inducer **b** is introduced as a finite pulse at time $t = \Delta t$ h (Figure 6A) . The start time of inducer **a** then becomes a reference for when the entire system is activated and ready to detect inducer **b**. Cell states are measured via fluorescence at time t_{end}, where $t_{\text{end}} > 24$ hours. Modeling results presented in this section are using the refined set of parameters defined in Figure 5C and Table S5.

If either of the two inducers is present in the media to some limit t_{end}, we would expect all of the S_o cells will end up in one of two populations (Figure 6B). Cells that encounter inducer **b** first will be in the S_b state, while cells that encounter **a** first will either be in the S_a or S_{ab} states. In the previous sections, once an inducer was added to the population, it was not removed, and the assumption was made that at times

greater than 24 hours, only a negligible number of S_o cells remained. This type of step function induction also meant that only the number of S_{ab} cells (GFP) was needed to uniquely determine the separation time Δt because *any and all* cells that had switched to S_a would eventually become S_{ab}.

However, in the case of a transient pulse, some cells that are in the S_a state (RFP) will not ever encounter inducer **b**. Assuming the k_{leakB} is small, these cells will remain in the S_a state. Therefore, the population of *a first* cells equals $S_a + S_{ab}$. We simulated a matrix of populations exposed to varying inducer separation times (Δt) and inducer **b** pulse widths (PW_b) to measure the resolution of detectable events (Figure 6C). In simulation (Figure 6D), we can see that the two populations mirror each other to add up to 100% of the total cells (N = 10,000 cells).

Since the step induction of **b** is equivalent as a pulse of infinite length ($PW_b = \infty$), and no cells remain in state S_a when $PW_b = \infty$, then perhaps the number of S_a cells can be used to deduce information about pulse width. In *silico*, we can test this hypothesis by running a matrix of simulations with varying Δt and PW_b. In Figure 6E, we see that the fraction of S_a cells over the total number of cells decreases monotonically with increasing PW_b. Due to some non-zero k_{leakB}, the curves produced by different Δt values do not completely overlap but the different is indistinguishably small. The maximum number of S_a cells does not go to 1 at $PW_b = 0$h because of leaky intB expression ($k_{\text{leakB}} = 0.01$).

If the fraction of S_a cells depends only on the duration of PW_b, the fraction of RFP cells relative to the total number of cells can be used to uniquely determine the pulse length of inducer **b**, PW_b (Figure 6E). Once PW_b is known, the fraction of GFP-expressing cells (S_{ab}) can be used to uniquely determine the time between inducers, Δt (Figure 6F). Furthermore, the genetically encoded state means that these population fractions should be maintained and measurable at a time, t_{end}, that is much later than the time of the events.
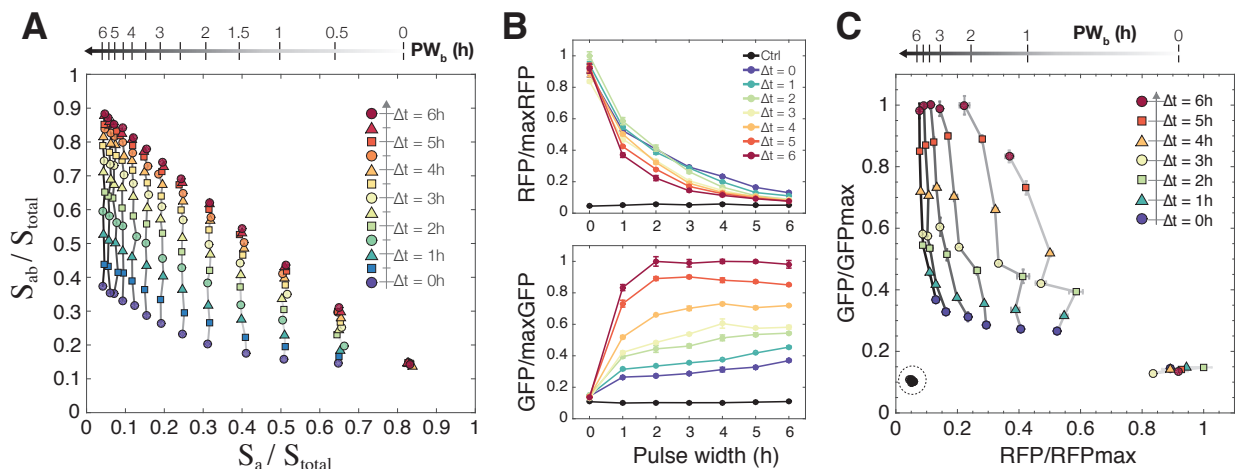

Figure 7: Determining arrival time and pulse duration of inducer **b** with population fractions. A) Simulation results from testing an 11×11 matrix of parameters with Δt and PW_b varying from $0-6$ hours in increments of 0.5 hours. Each point represents a population of 10,000 cells. Increasing PW_b goes from right to left, and increasing Δt goes from bottom to top. B) Experimental results showing RFP and GFP expression as a function of increasing Δt and PW_b. Fluorescence values have been normalized to the highest GFP and RFP fluorescence in the sample set. Experimental results from exposing temporal logic gate *E. coli* populations to varying PW_b and Δt values ($0-6$ hours, 0.01%/vol L-ara, 200 ng/ml aTc, measurements taken at 48 hours). C) A scatterplot of each population using their RFP and GFP expression as coordinates. The non-induced control samples are circled on the bottom left, and the samples with $PW_b = 0$h are on the bottom right. Samples with the same PW_b are connected with a solid line, and line darkness represents increasing PW_b duration. Samples with the same Δt are shown with the same colored shape marker and increasing Δt goes from bottom to top.

These conclusions can be extended in simulation to create a scatterplot of S_a cells versus S_{ab} cells in a population (Figure 7A) over an 11×11 parameter matrix varying Δt and PW_b from $0-6$ hours in increments of 0.5 hours (Additional plots in Fig. S7). Each point on the chart in Figure 7A represents a simulated

population (N = 10,000) exposed to a unique combination of Δt and PW_b values. Vertical lines represent the same PW_b value, and points with the same shape and color have the same Δt value. (See also Figure S7, Table S2). The simulation results suggest sufficient resolution of events as long as PW_b and Δt values are between 0 to 4 hours. For any single value of PW_b, we can follow the increasing Δt values vertically and see that the population response saturates after 4.5 hours resulting in overlapping between populations with $4.5 < \Delta t < 6$ hours. We can trace any individual Δt value horizontally from right to left, and observe that the points begin to cluster and overlap when $4.5 < PW_b < 6$ hours. These simulation data suggest that there should be some defined detection range of Δt and PW_b where each possible combination of the two is uniquely identifiable.

Experimentally, we tested a 7×7 matrix of varying Δt and PW_b (0 – 6 hours, 1 hour increments) on independent populations of the temporal logic gate *E. coli* strain (Figure 7B). All populations, except for the control, were exposed to inducer **a** (L-ara 0.01%/vol) at time t_0 to t_{end}. Pulses of inducer **b** (aTc, 200ng/ml) were achieved by sampling 5ul of the population and diluting 1:100 into fresh media with only inducer **a** (M9CA + 0.01%/vol L-ara). Fluorescence measurements were taken at 24 and 48 hours. For all values of Δt, the number of S_a cells (RFP) is highest when there is no exposure to inducer **b** ($PW_b = 0$h) and decreases monotonically as a function of PW_b (Figure 7B, top). RFP and GFP fluorescence expression has been normalized by the highest RPF and GFP expression values in the sample set. Our numerical simulations predicted a complete overlap of the S_a curves (Figure 6E), and the experimental results are consistent with those predictions though there is some downwards drift with increasing Δt. We see a more pronounced separation of the Δt curves when we look at GFP expression, a representation of the number of cells in the S_{ab} state. The number of S_{ab} cells is dependent on both Δt and PW_b and increases proportionally with both increasing **b** pulse duration and inducer separation time.

Even with these bulk fluorescence measurements, we can resolve the different populations that result from varying Δt and PW_b values (Figure 7C). As with Figure 7A, each point on the graph represents an independent population of cells (OD ~ 0.7). For any one value of Δt, increasing PW_b is inversely proportional to RFP expression, or S_a state cells. For any one value of PW_b, the RFP expression remains relatively constant with increasing Δt, while GFP, or the number of S_{ab} state cells, increases 3-fold from $\Delta t = 0$ to 6 hours. In the case where there is **b** pulse ($PW_b = 0$h), the amount of RFP fluorescence increases with sampling time Δt, but there is negligible expression of GFP. Populations with different PW_b exposures are well separated up to 4 hours though there is some drift in RFP as Δt increases. All of the populations exposed to either or both of the inducers are out of the range of the no inducer controls (indicated by dotted circle). An additional table of this data sorted by fluorescence expression can be found in Table S2.

This method of profiling is only valid if the fraction of S_a state cells can be used as a measure of PW_b that is independent of Δt. To ensure that this is not an artifact of simulation or experimental systems, we also mathematically analyzed the equation (1) for $\mathbb{P}_t(S_a)$, which represents the fraction of cells with DNA state S_a at time t.

If inducer **a** is used as a constant reference signal, all cells transition into either of either of S_a, S_b or S_{ab} state, thus $\mathbb{P}_\infty(S_a) = 1 - (\mathbb{P}_\infty(S_b) + \mathbb{P}_\infty(S_{ab}))$. If we assume that the basal leaky expression of intB is zero ($k_{\text{leak}B} = 0$), $\mathbb{P}_t(S_b) + \mathbb{P}_t(S_{ab}) = 0$ holds for $t \leq \Delta t$, since there is no intB that turns DNA state into S_b or S_{ab}. Then, we can show that $\mathbb{P}_t(S_b) + \mathbb{P}_t(S_{ab})$ is dependent only on PW_b, the duration of the pulse width of inducer B, for $t > \Delta t$ (see SI for details). Thus we can conclude that $\mathbb{P}_\infty(S_a) = 1 - (\mathbb{P}_\infty(S_b) + \mathbb{P}_\infty(S_{ab}))$ is dependent on PW_b but not on Δt, implying that the number of cells in the S_a state is a function of PW_b as $t \rightarrow \infty$ (See SI for derivation).

Practical use and calibration

Curve-fitting methods were used to automatically convert experimentally measured RFP and GFP fluorescence into PW_b and Δt values and to evaluate the resolution with which population fluorescence ratios can be used to determine inducer separation time and pulse duration. Using the experimental data from Figure 7B,C, we generated fitting curves for PW_b as a function of RFP/maxRFP(R), and for Δt as a function of both GFP/maxGFP(G) and PW_b (Figures S8 – S9, Table S1). We will denote these functions with $PW_b(R)$ and $\Delta t(G, PW_b)$, respectively. The functions $PW_b(R)$ and $\Delta t(G, PW_b)$ can then be used to generate a mesh of predicted PW_b and Δt values for any given normalized fluorescence values (Figure 8A, equations in SI). The predicted values were compared against the actual values to determine the approximate time win-
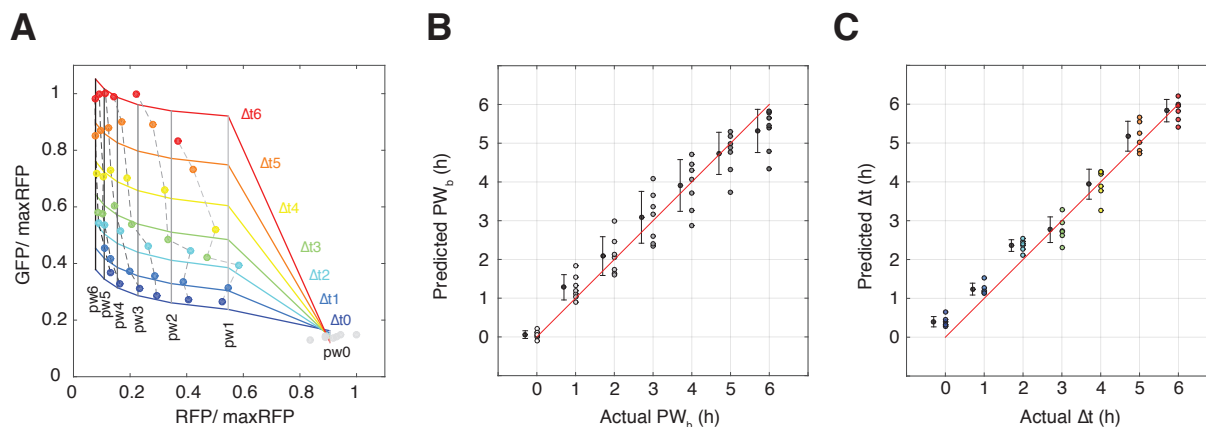
Figure 8: Determining prediction resolution for PW_b and Δt from fluorescence data. A) A mesh generated from fitted curves for PW_b as a function of RFP/maxRFP(R) and Δt as a function of pulse width and GFP/maxGFP(G). Experimental data is overlaid. B) Comparison of actual versus predicted PW_b values generated by fitted function $PW_b(R)$. For each actual PW_b value, the average of the predicted PW_b values with ± 1 standard deviation (Slightly offset on the x-axis for better comparison). C) Comparison of actual versus predicted Δt generated by the fitted function $\Delta t(G, PW_b)$. For each actual Δt values, the average of the predicted Δt with ± 1 standard deviation (Slightly offset on the x-axis for better comparison).

dow with which a specific PW_b or Δt can be resolved. For each actual value of PW_b and Δt, we calculated the average and standard deviation for the set of predicted values. The standard deviation allows us to visualize the range for which the majority of predictions will fall for any given actual value. For instance, a PW_b of 0 hours can be detected ± 0.25 hours, but as PW_b increases, this prediction window widens and for $PW_b \geq 3$ hours, the resolution of detection is closer to ± 1 hour (Figure 8B). Similarly, predicted values of Δt fall within ± 0.25 hours for $0 < \Delta t < 3$ hours and increase to ± 0.5 hours when $\Delta t \geq 3$ hours (Figure 8C). Using these fitting functions, we can also pre-generate a reference table that converts normalized RFP and GFP fluorescence data into predicted PW_b and Δt values (Table S3).

Discussion

We have designed and implemented a temporal logic gate that takes advantage of the population dynamics to collectively sense and record sequences of transient chemical inputs. As with all engineered systems, proper calibration of these temporal logic gate populations will be required prior to deployment in the "field." We envision a process similar to the one described in this report. First, experimental populations are exposed to a matrix of PW_b and Δt values. This will set the maximum and minimum fluorescence for RFP and GFP and provide necessary data for determining the Δt_{90} limit and producing the fitting functions $PW_b(R)$ and $\Delta t(G, PW_b)$. Once the fitting functions have been determined, values for PW_b and Δt for experimental samples can be estimated within ± 0.25 to 1 hour of the actual values. A calibrated table could also be generated and used for as a reference for samples that have been exposed to unknown conditions.

The stochastic nature of molecular processes often presents a significant barrier to homogenous outputs from an engineered population of cells. This implementation of event detection via population fractions takes advantage of stochastic and heterogenous individual responses to environmental conditions in order to map final population fractions back to unique sequences and durations of chemical events. The sensitivity of the system and the Δt_{90} detection limit could be modulated by increasing or decreasing protein production rates via tuning of plasmid copy numbers, signal concentration, or transcription/translation sequences. The use of digital cellular outputs combined with the analog population response creates event detection systems that are more robust to stochasticity and can be tuned more easily.

As a proof-of-concept, we have used the common laboratory inducers L-arabinose and aTc as inputs, but we hope that our temporal logic gate system can be used modularly with any biosensors of choice. In particular, we believe there are possibilities for detection of miRNAs and biofilm formation. Stable populations of microRNAs (miRNAs) circulating in the blood have generated a lot of interest as biomarkers for human health (Cortez *et al*, 2011). These short ($\sim$ 20-30nt) regulatory RNAs have been shown to have sequential tissue-specific expression signatures that correlate with pregnancy, tumor formation, and other diseases (Gilad *et al*, 2008; Mitchell *et al*, 2008), and synthetic biology has developed many customizable RNA sensors (Friedland *et al*, 2009; Green *et al*, 2014). Another possible application of this would be detection of harmful biofilms. Biofilms are self-assembling, highly structured, multi-species consortia that develop in stages and have sophisticated networks of interaction and function (Stoodley *et al*, 2002; Flemming and Wingender, 2010; Elias and Banin, 2012). Unnatural biofilm development in environments such as industrial water sources or waste streams can be both harmful for both the natural environment and the industrial mechanisms. Detection of biomarkers for known strains of biofilm colonizers would provide early warning of changing ecosystems, and although we do not yet fully understand these networks, it is known that quorum-sensing plays a critical role in the process. Quorum-sensing molecules and receptors are available in the synthetic biology toolbox and so may provide an accessible way of detecting the sequential colonization of different microbes. Field deployment of engineered bacteria will likely involve transient signals, low-nutrient environments, and possibly even other microbial competitors (i.e. soil, flowing rivers, the digestive tract). We used minimal media in this study to better approximate low-nutrient environments, and anticipate further characterization in more customized 'local' environments (i.e. gut model or air model or soil model) and with hardier microbial chassis.

Finally, this study focused on the population outputs as indicators of past events, but we believe that this temporal logic gate could be used to reliably differentiate a single strain into controlled sub-populations via input pulse order, duration, and frequency. In recent years, it has been recognized that many natural systems modulate cellular behavior not only by changing the concentration of signaling molecules but also by regulating signal pulse frequency (Cai *et al*, 2008; Lin *et al*, 2015). If we consider the fluorescent proteins GFP and RFP in this circuit as simply placeholders for downstream genes, then this system could easily be applied as a top-down population differentiator. By modulating the sequence of inputs, one could systematically predict and create mixed populations of genetically differentiated cells. As the scientific community turns towards further understanding of microbiomes and multi-cellular consortia, engineered bacteria populations could be used not only as a tool for investigating the activities of natural communities but also as a way to build synthetic communities from the ground up.

Materials and methods

Cell strains and plasmids

All plasmids used in this study were designed in Geneious 7.1 (Biomatters, Ltd.) and made using standard Gibson isothermal cloning techniques. Integrases Bxb1 and TP901-1 are on a high-copy plasmid (pVHed05, plasmid map in Figure S9) with a ColE1 origin of replication (original template from the Dual Recombinase controller (Bonnet *et al*, 2013), Addgene Plasmid 44456). Integrase A (Bxb1) is behind a Ptet promoter and integrase B (TP901-1) is behind a PBAD promoter. The plasmid has been modified with an additional TetR gene. The temporal logic gate was integrated into the Phi80 site on the *E. coli* chromosome using CRIM integration (Haldimann and Wanner, 2001) and screened for single integrant colonies. The integration plasmid template and DH5α-Z1 strain were generously provided by J. Bonnet and D. Endy and modified to contain the temporal logic gate (pVHed07, plasmid map in Figure S9).

Additional DNA and oligonucleotides primers were ordered from Integrated DNA Technologies (IDT, Coralville, Iowa).

A custom formulation of M9CA media was used for all experiments. The media contained 1x M9 salts (Teknova, M1906) augmented with 100mM NH_4CL, 2mM $MGSO_4$, 0.01% casamino acids, 0.15 μg/mL biotin, 1.5 μM thiamine, and 0.2% glycerol, and then sterile filtered (0.2 μm).

Simulations of the model

The stochastic simulation algorithm by Gillespie (Gillespie, 1977) was implemented to generate the sample paths of individual cells using the Markov model (see Table S6 for the definitions of Markov transitions and transition rates). All simulation runs and their analyses were done with MATLAB (R2014b,The MathWorks, Inc.). All simulated populations were done with 10,000 individual cell trajectories.

Experimental methods

Prior to all experiments, cells were grown overnight from plate cultures in M9CA for two days, then diluted to OD 0.1 and recovered for 4-6 hours at 37^oC. L-arabinose and anhydrous tetracycline (aTc) were used as inducers **a** and **b**, respectively. L-ara was used a concentration of 0.01% by volume, and aTc was used a concentration of 200 ng/ml (450nM). All media contained the antibiotics chloramphenicol (source and concentration) and kanamycin (source and concentration). All experiments were performed with the aid of timed liquid handling by a Hamilton STARlet Liquid Handling Robot (Hamilton Company).

For step function experiments, the cells were diluted to OD 0.1 into a 96-well matriplate (Brooks Automation, Inc., MGB096-1-2-LG-L) with $500\mu l$ total volume in M9CA. Cultures were incubated at 37^oC in a BioTek Synergy H1F plate reader (BioTek Instruments, Inc.) and inducers were added at appropriate time by the Hamilton robot. OD and fluorescence measurements (superfolder-GFP ex488/em520, mKate2-RFP ex580/em610) were taken by the BioTek every 10 minutes.

For the pulse experiments, pulses were achieved through dilution of the culture into fresh M9CA media containing 0.01% L-arabinose. 5 μl of the culture was sampled and diluted into 500 μl of fresh M9CA + 0.01% L-ara to achieve pulsatile exposure to aTc. 96-well deep-well plates containing the diluted cultures were then incubated at 37^oC and fluorescence measurements were taken at 24 and 48 hours in the plate reader.

Analysis of experimental data was done using custom MATLAB scripts. All depicted error bars are standard error of the mean. Fitting of curves was done in MATLAB.

Acknowledgements

Author Contributions

V.H. conceived of the circuit design, constructed the necessary experimental strains, performed experimental work, and ran model simulations. Y.H. developed the stochastic model and derived the mathematical results. P.W.K.R provided feedback and guidance on data analysis and interpretation. R.M.M provided feedback and guidance on overall project vision, circuit design, and interpretation of results.

References

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. *Molecular Systems Biology* **3**: 133

Bonnet J, Subsoontorn P, Drew E (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *PNAS* **109**: 8884–8889

Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Drew E (2013) Amplifying genetic logic gates. *Science* **340**: 599–603

Cai L, Dalal CK, Elowitz MB (2008) Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**: 485–490

Cortez MA, Bueso-Ramos C, Ferdin J (2011) MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nature reviews Clinical*

Dunlop MJ, Cox RS, Levine JH, Murray RM, Elowitz MB (2008) Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics* **40**: 1493–1498

Elias S, Banin E (2012) Multi-species biofilms: living with friendly neighbors. *FEMS microbiology reviews* **36**: 990–1004

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297**: 1183–1186

Flemming HC, Wingender J (2010) The biofilm matrix. *Nature Reviews Microbiology*

Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ (2009) Synthetic gene networks that count. *Science* **324**: 1199–1202

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**: 339–342

Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, Chajut A (2008) Serum microRNAs are promising novel biomarkers. *PLoS ONE* **3**: e3148

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**: 2340–2361

Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell* **159**: 925–939

Haldimann A, Wanner BL (2001) Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. *Journal of Bacteriology* **183**: 6384–6393

Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MCM (2011) A phage protein that binds φC31 integrase to switch its directionality. *Molecular Microbiology* **80**: 1450–1463

Kotula JW, Kerns SJ, Shaket LA, Siraj L, Collins JJ, Way JC, Silver PA (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. *PNAS* **111**: 4838–4843

Lin Y, Sohn CH, Dalal CK, Cai L, Elowitz MB (2015) Combinatorial gene regulation by modulation of relative pulse timing. *Nature*

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 10513–10518

Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2013) Genetic programs constructed from layered logic gates in single cells. *Nature* **491**: 249–253

Shis DL, Hussain F, Meinhardt S, Swint-Kruse L, Bennett MR (2014) Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras. *ACS Synthetic Biology* : 140728104557004

Siuti P, Yazbek J, Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology* **31**: 448–452

Stoodley P, Sauer K, Davies DG, Costerton JW (2002) Biofilms as complex differentiated communities. *Annual Review of Microbiology* **56**: 187–209

Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* **315**: 1716–1719

Wang B, Kitney RI, Joly N, Buck M (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications* **2**: 508

Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA (2014) Permanent genetic memory with >1-byte capacity. *Nature Methods* **11**: 1261–1266

Yuan P, Gupta K, Van Duyne GD (2008) Tetrameric structure of a serine integrase catalytic domain. *Structure* **16**: 1275–1286