

Analysis of a Digital Clock for Molecular Computing

Johan Ugander, Mary J. Dunlop, and Richard M. Murray

Abstract—The control of synthetic genetic regulatory networks is an emerging engineering challenge. In this study, we propose a new synthetic genetic network that behaves as a digital clock, producing square waveform oscillations. We analyze two models of the network: a deterministic model based on Michaelis-Menten kinetics, as well as a stochastic model based on the Gillespie algorithm. Both models predict regions of oscillatory behavior; the deterministic model provides insight into the conditions required to produce the oscillating clock-like behavior, while the stochastic model is truer to natural dynamics. Intracellular stochasticity is seen to contribute phase noise to the oscillator, and we propose improvements for the network and discuss the conceptual foundations of these improvements.

I. INTRODUCTION

Gene expression is often controlled by natural genetic regulatory networks that govern the rates at which genes are transcribed. Recent work has shown that synthetic versions of genetic networks can be designed and built in living cells. Mathematical modeling can provide accessible insight into the dynamics behind the biology [1]. Applications for these synthetic regulatory networks include intracellular decision-making and computation.

Recent work in synthetic biology has produced biological examples of a toggle switch and a ring oscillator. These two components are the basis of our design for a digital clock. The toggle switch memory unit [2] is a genetic network where two genes repress each other (Fig. 1a) to achieve bistability between two protein concentrations; when one protein concentration is high, it holds the other low, and vice versa. The repressilator [3] is another genetic network, where three genes repress each other in a ring (Fig. 1b), which causes the three constituent protein concentrations to oscillate in time, phase-shifted 120 degrees from each other. These two genetic networks are examples of basic building blocks for molecular computing at the protein level.

The network we have designed consists of two toggle switches that are cross-wired to repress each other (Fig. 1c), and we call the network the togglator. Each of the four genes in the network is repressed by two other genes. This is achieved through a combinatorial (NOR gate) promoter site [4], a region upstream of the gene featuring two binding sites that proteins can occupy to inhibit transcription.

Research supported in part by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office.

J. Ugander is a student at the Faculty of Engineering, Lund University, Lund, Sweden. jugander@cds.caltech.edu

M. J. Dunlop and R. M. Murray are in the Division of Engineering and Applied Science, California Institute of Technology, CA 91125, USA.

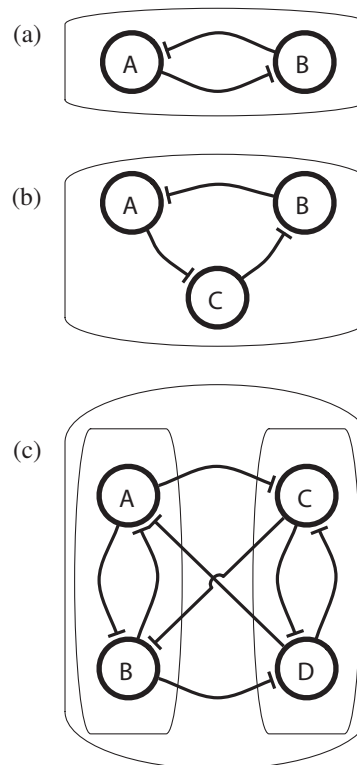


Fig. 1. Genetic regulatory network schematics. (a) The toggle switch memory unit, (b) the repressilator, and (c) our proposed digital clock, the togglator. It consists of two toggle switches modules, cross-wired to repress each other.

The concentrations of the constituent biomolecules in the network form data streams analogous to the voltage levels in an electronic circuit. For our network, the periodic oscillating concentrations of the four output proteins act as our proposed digital clock. The aim of molecular computing is to perform logical operations using these concentrations as inputs and outputs. In electrical computing, clocks are used to coordinate the actions of two or more circuits, and our clock aims to perform this task for molecular circuits.

For our network, mathematical modeling predicts that the concentrations of each of the four proteins in the network will oscillate in a square fashion. These square oscillations are generated such that each constituent switch module provides two clocks, which are phase-shifted 180 degrees from each other, serving as logical inverses. The two switch modules are phase-shifted 90 degrees from each other, effectively offering four square clocks spaced evenly across the phase space.

The period of an imperfect square waveform can be divided into four components, two stationary and two transition

periods. For perfect square waves, the transition periods have length zero. The speed of the inter-toggle dynamics govern the length of the stationary periods of the waveform, while the intra-toggle dynamics decide the length of the transition periods. Variation in the length of the stationary periods can be viewed as phase noise for the oscillations, and the cause of this is discussed. Our new design offers the possibility of adjusting the period of the oscillation through a choice of proteins used in a physical implementation. The possibility of further design control is discussed in connection with other regulatory mechanisms.

II. DETERMINISTIC MODELING

Genetic networks can be modeled deterministically as a system of non-linear ordinary differential equations using Hill functions, a modified version of Michaelis-Menten kinetics [5]. This deterministic model allows for analysis using mathematical tools from traditional non-linear dynamics. It is well known that intracellular stochasticity plays an important role in genetic network dynamics [6], and a stochastic model of the network will be discussed in a following section.

Transcriptional regulation of a gene Z by proteins X and Y is modeled as:

$$\dot{Z} = \alpha_{0z} + \alpha_z G(X, Y) - \beta Z, \quad (1)$$

where α_z is the regulatable transcription rate, the rate of protein production controlled by transcriptional regulation, while α_{0z} is the basal transcription rate, the leakage that cannot be controlled. β is the protein decay rate.

The function $G(X, Y)$ is the Hill OR-gate function describing how the transcription factors, here proteins X and Y , bind competitively to the DNA (for deterministic modeling, non-competitive and competitive binding show the same qualitative results [4]). The Hill OR-gate function is formed as the sum of two regulation functions: $G(X, Y) = f(X, Y) + f(Y, X)$. Here, both regulation functions model repression, which is expressed as:

$$f(X, Y) = 1/(1 + (X/K_{xz})^2 + (Y/K_{yz})^2). \quad (2)$$

The Hill function describes the ratios of transcription factor to threshold parameters (K_{ij}) as a multivariate sigmoidal function with values between 0 and 1. The Hill function can be derived from equations describing the chemistry of transcription factor binding. For simplicity, the state variables X and Y can be scaled so that the threshold parameters are 1.

With these methods, the toggator genetic network is now modeled as:

$$\dot{A} = \alpha_0 + \alpha(2/(1 + B^2 + D^2)) - \beta A \quad (3)$$

$$\dot{B} = \alpha_0 + \alpha(2/(1 + A^2 + C^2)) - \beta B \quad (4)$$

$$\dot{C} = \alpha_0 + \alpha(2/(1 + A^2 + D^2)) - \beta C \quad (5)$$

$$\dot{D} = \alpha_0 + \alpha(2/(1 + B^2 + C^2)) - \beta D, \quad (6)$$

where A , B , C , and D are the concentrations of the four proteins of the network, recalling that A - B and C - D form

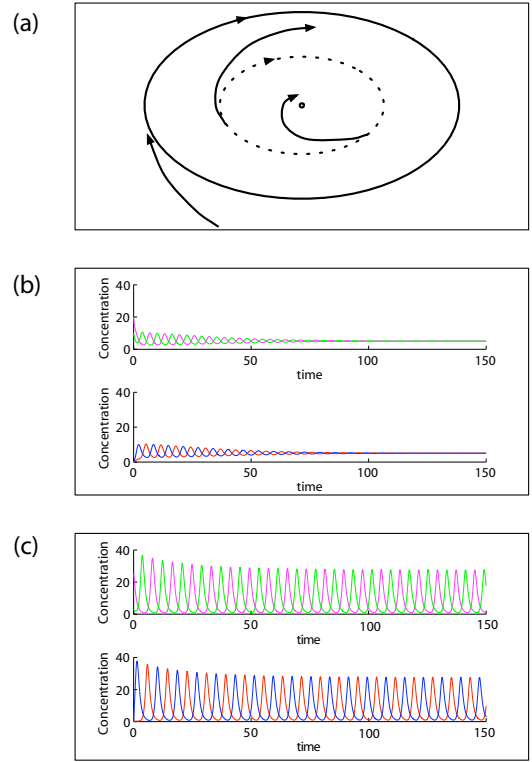


Fig. 2. Results for the deterministic model of the network. (a) A 2D schematic illustrating the qualitative behavior of the 4D system, featuring a stable limit cycle and equilibrium point, separated by an unstable limit cycle. (b) Realization leading to the equilibrium point, and (c) the limit cycle. Both figures show all four protein concentrations, where $A(t)$ is green, $B(t)$ pink, $C(t)$ blue, and $D(t)$ red. The parameters values ($\alpha = 130$, $\alpha_0 = 130 \cdot 10^{-3}$) are taken from Elowitz's model of the repressilator [3], while $\beta = 1$ for simplicity.

switch modules. Here, symmetry of the parameters α_0 and α across switch components has been assumed.

The dynamic behavior of the 4-dimensional system can be qualitatively understood with a two-dimensional schematic representation, seen in Fig. 2a. For most realistic parameter values, the system has two basins of attraction, one leading to a stable limit cycle describing the expected oscillations, and one leading to a stable equilibrium point. The separatrix between the two basins is an unstable limit cycle. In order to reach the oscillating basin, the protein concentrations must be initially strongly imbalanced; the equilibrium basin is reached as a result of an initially gridlocked regulatory network. Oscillatory and equilibrium realizations are shown in Fig. 2 b and c.

We will now show that there is exactly one equilibrium concentration at which the proteins stabilize for a given set of parameter values α and α_0 , and that this equilibrium point is a stable fixed point for the system. We find this point by setting $\dot{A} = \dot{B} = \dot{C} = \dot{D} = 0$ and $A = B = C = D = x$, for which the system reduces to the equation:

$$0 = \alpha_0 + \alpha(2/(1 + 2x^2)) - x,$$

Transcription	Translation & Decay	Promoter region	Parameters
$D_{a+00} \xrightarrow{k_1} D_{a+00} + M_a$ $D_{a+b0} \xrightarrow{\gamma \cdot k_1} D_{a+b0} + M_a$ $D_{a+0d} \xrightarrow{\gamma \cdot k_1} D_{a+0d} + M_a$ $D_{a+bd} \xrightarrow{\gamma \cdot k_1} D_{a+bd} + M_a$	$M_a \xrightarrow{k_2} M_a + P_a$ $P_a \xrightarrow{k_3} \emptyset$ $M_a \xrightarrow{k_4} \emptyset$	$D_{a+00} + P_b \xrightarrow{k_5} D_{a+b0}$ $D_{a+00} + P_d \xrightarrow{k_5} D_{a+0d}$ $D_{a+0d} + P_b \xrightarrow{k_5} D_{a+bd}$ $D_{a+b0} + P_d \xrightarrow{k_5} D_{a+bd}$ $D_{a+b0} \xrightarrow{k_6} P_b + D_{a+00}$ $D_{a+0d} \xrightarrow{k_6} P_d + D_{a+00}$ $D_{a+bd} \xrightarrow{k_7} P_b + D_{a+0d}$ $D_{a+bd} \xrightarrow{k_7} P_d + D_{a+b0}$	$k_1 = 0.5$ $k_2 = 0.167$ $k_3 = \ln(2)/120$ $k_4 = \ln(2)/600$ $k_5 = 1$ $k_6 = 9$ $k_7 = 224$ $\gamma = 0.001$

TABLE I

THE SYSTEM OF CHEMICAL REACTIONS UNDERLYING THE STOCHASTIC MODEL OF THE NETWORK. P_z AND M_z ARE THE PROTEIN AND MRNA FOR GENE Z . D_{z+xy} IS THE DNA FOR GENE Z , WITH PROTEINS X AND Y BOUND, WHILE D_{z+0y} IS DNA FOR GENE Z WITH ONLY PROTEIN Y BOUND. FOR THIS STUDY, PARAMETER VALUES WERE AGAIN CHOSEN FROM SIMULATIONS OF THE ELOWITZ REPRESSILATOR [3], CHOSEN TO AGREE WITH THE PARAMETER VALUES FOR THE DETERMINISTIC MODEL. NOTICE HOW COOPERATIVITY IS OBTAINED THROUGH THE HEIGHTENED REACTION RATE FOR THE UNBINDING OF THE SECOND-OCCUPIED (k_7 COMPARED TO k_6) OPERATOR SITE.

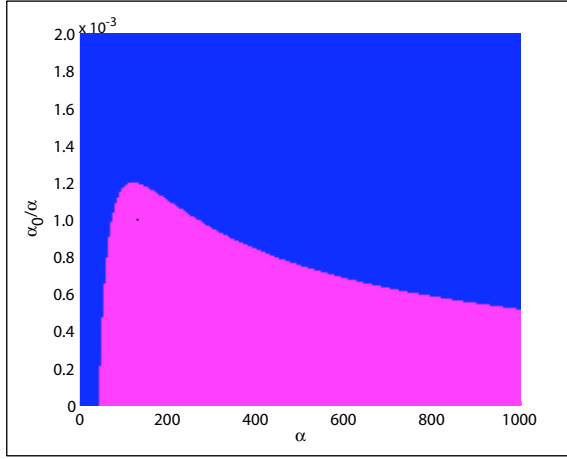


Fig. 3. Parametric study of the deterministic model. The parameter space ($\alpha \times \alpha_0/\alpha$) illustrates how the oscillatory basin exists only in certain regions of parameter space. Blue indicates the presence of the equilibrium basin, pink indicates the presence of both the oscillatory and equilibrium basin. The black point indicates the parameter values chosen from the Elowitz model and used in Fig. 2c.

which can be rewritten as the third order polynomial:

$$0 = x^3 - \alpha_0 x^2 + (1/2)x - (2\alpha + \alpha_0)/2. \quad (7)$$

Using a computer based analytic solver such as Maple [7], this polynomial can be shown to have only one real root for positive real parameters, corresponding to the one equilibrium concentration. The fixed point is proven to be stable by examining the eigenvalues of the Jacobian linearized around the fixed point. The eigenvalues are found to be:

$$\lambda_k = \{-1, 2y - 1, -y - 1 + iy, -y - 1 - iy\},$$

where $y = ((-2\alpha x)/(1 + 2x^2))$. Because $\Re(\lambda_k) < 0$ for all $\alpha \in \mathbb{R}^+$, the fixed point is stable.

We will now explore the behavior of the system when we vary the two parameters α , the regulatable transcription rate, and α_0 , the basal transcription rate. Intuitively, a high basal transcription rate α_0 would make the regulatory mechanisms ineffective. Likewise, very low values of α could disable the regulatory mechanism. Fig. 3a maps the parameter space

($\alpha \times \alpha_0/\alpha$) to the possible behaviors of the system, showing that a bifurcation occurs.

In agreement with our intuitive understanding from biology, high basal transcription rates cause the stable and unstable limit cycles to annihilate each other, leaving only the equilibrium behavior. Moving in the other direction, low basal transcription rates result in the unstable limit cycle contracting in on the stable fixed point, and for exceptionally low values of α_0 the equilibrium basin becomes miniscule.

Before presenting the stochastic model, the expected consequences of adding noise to the system will be discussed. One expected consequence is that small equilibrium basins will become highly unstable; trajectories in the equilibrium basin will be easily ejected by noise, crossing the separatrix and entering the oscillatory basin. Trajectories in the oscillatory basin may cross into the equilibrium basin, but this will be difficult for small equilibrium basins, and can be expected to occur only for short periods before being returned. As will be discussed later, this equilibrium basin can be identified as the source of phase noise in the stochastic model of the network.

III. STOCHASTIC MODELING

Genetic networks can be modeled more realistically as a biochemical system using Gillespie's stochastic simulation algorithm [8]. The Gillespie algorithm can be derived from the chemical Master equation, and simulates the behavior of a system of chemical species probabilistically on a reaction-by-reaction basis. Essentially, molecular collisions are modeled to occur at times drawn from an exponential distribution characterized by the total concentration of molecules, and a second random number determines what reaction occurs based on the reaction rates of the system. For more information, reference [8] provides an excellent introduction with many examples. For genetic networks, abstract chemical species such as 'protein' and 'DNA' can be used to avoid modeling the system on a molecular level.

For the Gillespie algorithm, the toggator network is modeled with the system of chemical reactions seen in Table 1. The table presents only those reactions directly governing one of the four genes. The reactions for the remaining three genes are constructed symmetrically, for a total of 60 reactions.

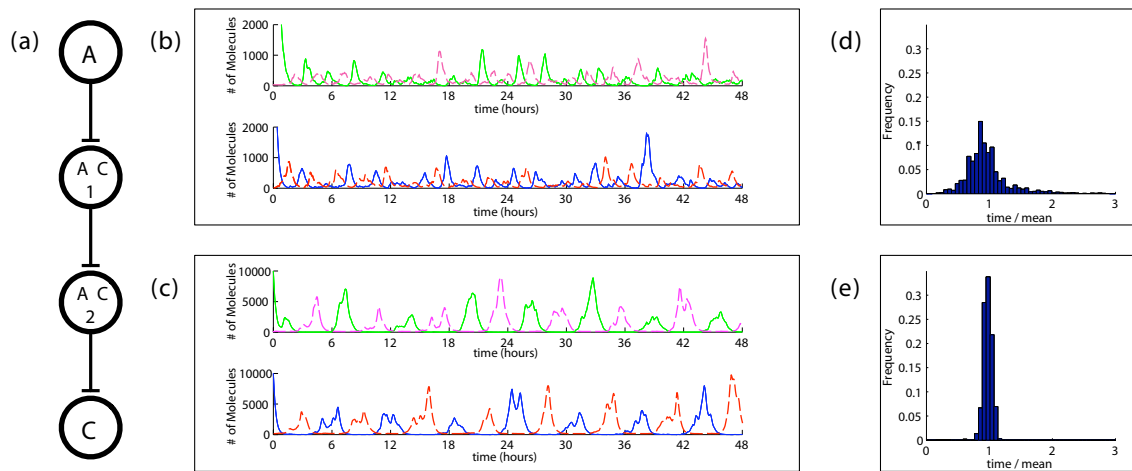


Fig. 4. Results for the stochastic model of the network. (a) A repression cascade with length 3, used to delay the inter-toggle communication. (b) Realization of the non-delayed system, and (c) the delayed system. The top half of each figure shows the concentrations of proteins A (green solid line) and B (pink dashed line), while the bottom half shows the concentrations of proteins C (blue solid line) and D (red dashed line). (d) A histogram of the period length of each oscillation in the non-delayed system, and (e) delayed system. The periods are identified using the Viterbi algorithm [10] [11], where the most likely sequence of transitions is identified for a Hidden Markov model with 2 states corresponding to high and low Gaussian behavior.

With symmetries, the system of reactions contain 7 parameters. The γ parameter for our stochastic model is the ratio of basal to regulatable transcription, corresponding to α_0/α from our deterministic model. The dynamics for the deterministic model are summarized by the parameters α and α_0 , where the stochastic model is parameterized in greater detail by the reaction rates k_{1-6} . See reference [3] for a more detailed explanation. No extensive parametric study was conducted for the stochastic model, though we confirmed that the chosen values were not anomalous.

As was discussed previously, the performance of the network is improved dramatically by separating the time scales of the inter-switch and intra-switch repression pathways. This separation of times scales is difficult to achieve with only transcriptional regulation. The method used in this study has been to introduce a delay to the inter-switch pathways by replacing the direct repression pathways with repression cascades (Fig. 4a), a chain of genes repressing each other [9]. These cascades effectively serve as a delay function by routing the signal in a circumlocutory fashion.

Stochastic simulations for both the delayed and non-delayed networks were conducted, and a realization from the non-delayed network can be seen in Fig. 4b, while realizations from the delayed network featuring cascade schemes can be seen in Fig 4c.

The cascades introduce a delay in the inter-switch communication, which determines the length of the stationary period of the square waveform. This delayed network has a much clearer clock, where the concentrations are distinctly high or low. The drastic improvements brought by the cascade delay schemes comes at the cost of an equally drastic increase in biological complexity. With length 3 cascades added for each inter-switch pathway, the system requirements jump from 4 to 12 proteins. Other methods of delay will be discussed in the following section.

A notable advantage of the cascade delayed network is that it exhibits reduced phase noise, the variation in how long each switch stays ON. Understanding this requires insight into both the stochastic and deterministic models presented in this study. When the time scales are comparable, random pulses in a given protein concentration are comparable to the total concentration of that protein, since it has not had the time to fully rise. This pushes the system into the equilibrium basin seen in the deterministic model, where it will remain until a new random pulse in concentration ejects it back into the oscillatory basin. Introducing delay allows the switch to stabilize at a significantly higher level, drowning out random pulses much more effectively.

The amount of time spent in the equilibrium basin can be viewed as a stochastic variable that adds to the period of the oscillations, an effect that can be summarized as phase noise. Histograms of the periods of oscillation for the regular and delayed networks can be seen in Fig. 4d-e. Comparing the two histograms, the period of the delayed network clearly has a narrower distribution.

IV. FUTURE WORK AND CONCLUSIONS

In this study, we present designs for an intracellular digital clock in the form of a genetic regulatory network, while providing analysis of both a deterministic and stochastic model of the network. Stochastic modeling reveals that the network is susceptible to large amounts of phase noise, and this problem is addressed by adding cascade delay mechanisms to drown out erroneous concentration spikes.

Transcriptional regulation is the only regulatory mechanism considered in this study, and it is important to note that further improvements can be expected if different regulatory mechanisms are harnessed to extend the separation of time scales. Two techniques worth exploring are phosphorylation cascades [12] and ubiquitin-dependent protein degradation [13].

By designing the inter- and intra-toggle pathways with different regulatory techniques, the potential for time scale separation is much larger. The cascade delay mechanism in this study should only be considered as a proof of concept. Other mechanisms may offer means of producing behavior in line with the delayed model in this study without the added complexity of the cascade. Once a solution is found that separates inter- and intra-toggle time scales and avoids introducing immense complexity, construction and verification *in vivo* would be the next obvious step.

REFERENCES

- [1] H. El Samad, D. Del Vecchio, and M. Khammash. Repressilators and Promotilators: Loop Dynamics in Gene Regulatory Networks. *American Control Conference*, 2005. (Invited Paper)
- [2] T.S. Gardner, C.R. Cantor, J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339-342, 2000.
- [3] M.B. Elowitz and S.A. Leibler. Synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335-338, 2000.
- [4] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci*, 100:11981-11985, 2003.
- [5] J.D. Murray. *Mathematical Biology I*. Springer Verlag, 3rd edition, 2002.
- [6] H.H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15:65-69, 1999.
- [7] Waterloo Maple Software. *Maple 10*, 2005, Waterloo, Ont. (Computer Program)
- [8] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340-2361, 1977.
- [9] S. Hooshangi, S. Thiberge, and R. Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci*, 102:3581-3586, 2005.
- [10] G.D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61:268278, 1973.
- [11] K. Murphy. *Hidden Markov Model Toolbox for Matlab*, <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html> (accessed September 19, 2006)
- [12] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall, 2007, pg. 106-108.
- [13] R.L. Bar, Ubiquitin-dependent protein degradation. *Annual Review of Genetics*, 30:405-439, 1996.