Temporal pattern recognition through analog molecular computation

Jackson O’Brien and Arvind Murugan
The James Franck Institute and Department of Physics, University of Chicago, Chicago, IL 60637

Living cells communicate information about physiological conditions by producing signaling molecules in a specific timed manner. Different conditions can result in the same total amount of a signaling molecule, differing only in the pattern of the molecular concentration over time. Such temporally coded information can be completely invisible to even state-of-the-art molecular sensors with high chemical specificity that respond only to the total amount of the signaling molecule. Here, we demonstrate design principles for circuits with temporal specificity, that is, molecular circuits that respond to specific temporal patterns in a molecular concentration. We consider pulsatile patterns in a molecular concentration characterized by three fundamental temporal features - time period, duty fraction and number of pulses. We develop circuits that respond to each one of these features while being insensitive to the others. We demonstrate our design principles using abstract Chemical Reaction Networks and with explicit simulations of DNA strand displacement reactions. In this way, our work develops building blocks for temporal pattern recognition through molecular computation.

Recent breakthroughs in synthetic biology have led to molecular sensors that report on the local environment in cells by detecting signaling molecules with high chemical specificity [1-4]. However, by themselves, such sensors can be completely blind to temporally coded information in cells and tissues. In fact, living cells often communicate information about physiological conditions by producing a signaling molecule in a specific timed manner [5-7]. For example, many rapid pulses of nuclear p53 in mammalian cells indicates γ radiation damage and leads to cell cycle arrest, while a longer sustained single pulse of nuclear p53 indicates UV damage and leads to programmed cell death [8]. Thus different biological conditions can result in the same total amount of a signaling molecule, differing only in the pattern of the molecule’s concentration over time [8]. Such biological conditions cannot be distinguished by a sensor that responds to the total amount (or exposure) to a target molecule, even if the sensor has high chemical specificity.

In this article, we demonstrate design principles for molecular circuits with temporal specificity, i.e., molecular circuits that respond to specific temporal features in the concentration of an input molecule, instead of the total exposure to that input molecule. We show chemical reaction networks satisfying these principles and derive constraints on their rate constants. We then find explicit implementations of these abstract reaction networks using DNA strand displacement reactions and verify the performance of such DNA circuits using simulation software explicitly designed for this purpose [9]. We anticipate that this work will also be of use for other synthetic biology constructs based on enzymes and transcriptional gates [10, 11] and for analyzing naturally occurring temporal decoding mechanisms in cells [5].

While time-varying signals are high dimensional and can vary in endless ways, we restrict our study here to pulsatile patterns as shown in Figure 1. We focus on three fundamental temporal features that characterize such signals - time period $T$ (i.e., time between onset of pulses), duty fraction $\delta$ (i.e., width of each pulse relative to duration between pulses) and the number of pulses $n$ (i.e., length of the pulse train). We construct circuits that respond to each one of these three independent features while being insensitive to other features.

**Analog computing with transients**

Recognizing temporal patterns can be interpreted as an analog operation that is naturally suited for molecular computation. While molecular circuits of digital gates have solved remarkable problems [12-14] and even mimicked deep neural networks [15], several recent papers [16-25] have shown that analog computations are naturally suited for molecular circuits. Analog operations represent information directly in continuous physical variables like concentration or time intervals [26, 27]. Such direct representations can make analog devices smaller and simpler than corresponding digital computation with binary encoding. The temporal pattern recognition problem studied here is naturally suited to analog computation since the input itself is temporally encoded [17, 18]. Indeed, some of the earliest applications of analog computing were to similar temporal decoding problems in electrical circuits, such as AM or FM decoding in radios [27]. While there are theoretical limitations on chemical timers [28], several recent works on DNA-based and other molecular computing have addressed related temporal questions. Even a simple boolean AND-gate can be considered a coincidence detector between two signals [29] and a temporal sequence detector was recently demonstrated [30], though these constructs do not detect specific temporal intervals. An AND gate and a time-delay mechanism can be combined in a Coherent Feed-Forward Loop that serves as a pulse width discriminator [31]. Recent work has shown how to convert oscillatory signals

*amurugan@uchicago.edu*
FIG. 1. Schematic of temporal feature extraction. The three time-varying patterns of a molecular concentration \( c(t) \) differ in the number of pulses \( n \), the time period \( T \) (i.e., time between onset of pulses), and duty fraction \( \delta \) (i.e., pulse width relative to time period) but have the same total integrated exposure \( nT\delta \). (a) Circuits that respond to the total integrated exposure are thus unable to distinguish the three signals. (b) We seek chemical reaction networks that can extract independent temporal features \( n, T, \delta \) and report each of them quantitatively. For example, such a sensor must be able to count pulses without regard to pulse width and report duty fraction without regard to the time period.

As in these recent analog circuits, our approach here seeks to turn a weakness inherent to molecular devices in the context of digital computation, into a strength for analog computation. Digital computation through molecules requires all internal transients to rapidly die out, so the system can be approximately described by discrete steady state concentrations. In contrast, even on a digital computer, temporal pattern recognition problems are best solved using recurrent networks with internal transients\[33\], unlike, say, feed-forward networks used for image recognition. Such finite timescale transients serve as natural rulers to measure and process information encoded as time intervals. Thus our work exploits the inevitable transients in real molecular systems to perform useful computation.

Natural temporal patterns can be more complex than the family studied here and some applications might require detecting more complex features or combinations of features presented here. Such complex temporal processing can be performed by combining the basic motifs and design principles introduced here. In this way, our work lays the foundation for temporal pattern recognition through analog molecular computation.

I. RESULTS AND DISCUSSION

We consider time varying patterns \( c(t) \) in the concentration of an input molecule \( I \) composed of pulses like those shown in Fig[1]. The patterns shown can be described by three independent numbers: the number of pulses \( (n) \), the time period \( T \), and the duty fraction \( \delta \in [0, 1] \) (i.e., pulse width relative to time period \( T \)).

As shown in Fig[1], the same total amount of input molecule \( I \) can be spread out over time in many different ways. The different patterns shown have distinct \( n, T, \delta \) but have the same total area = \( nT\delta \). Hence a naive sensor that is only sensitive to the total amount of \( I \) would respond in the same way to all patterns shown.

Here, we seek decoders that can report on each of these temporal features independently - e.g., the duty fraction readout should be independent of the time period and the number of pulses while the number of pulses readout should not depend on the width or separation of the pulses. In what follows, we demonstrate the decoder for each of these temporal features - \( n, T, \delta \) - one at a time. In each case, we first demonstrate a high-level Chemical Reaction Network of the architecture and then show the DNA strand displacement-based implementations. The three independent numbers \( n, T, \delta \) completely characterize the family of patterns shown here, i.e., they form a complete independent 'basis' for the family shown here. Other combinations of these features might be relevant for specific applications (e.g., pulse width \( T\delta \) instead of duty fraction \( \delta \) but the principles behind the basis set of decoders should allow development of sensors for such features as well. We will assume that the amplitude of the input signal is fixed and will not consider amplitude fluctuations. A novel mechanism to convert oscillatory signals with such amplitude noise into a stereotyped signal was proposed recently[21]; such an amplitude buffering mechanism can be used upstream of the temporal decoders proposed here.

A. Pulse number decoder

We begin with a molecular circuit that can count the number of pulses \( n \) seen, without regard to the width of each pulse or the separation between pulses.

To count pulses \( n \) in this manner, we first seek a circuit that produces a stereotyped response to each pulse that is independent of pulse width and separation. To do so, we take inspiration from 'biochemical adaptation' mechanisms used e.g., in bacterial chemotaxis [34]. Such molec-
FIG. 2. A pulse count decoder implemented by an Incoherent Feed-Forward Loop that responds only to step ups in input (but not step downs). (a) Reaction network to decode pulse number. Species $G$ is created only when $I$ and $A$ are present. (b) DNA Strand Displacement (DSD) implementation of mechanism in (a) with waste products suppressed (see SI for full circuit). Species $\bar{I}$ is the negation of input $I$ as in dual-rail logic, i.e., $\bar{I}$ is high when $I$ is low and vice-versa. (c) $G$ shows a stereotyped response to pulses of $I$ that is independent of the width and separation of such pulses in $I$. The stereotyped response is due to the ‘incoherent’ regulation of $G$ by $I$; $I$ exerts a direct fast positive effect on $G$, causing a rapid rise, but also exerts an indirect delayed negative effect on $G$ by suppressing $A$ (green). (d) The output $P$ integrates and reports the number of stereotyped responses of $G$. Consequently, changing the (e) duty fraction or (g) time period of pulses in $I$ has no impact on the output $P$ which does change with (f) pulse count.

ular circuits show a transient response to step changes in an input signal but return to their prior state and are insensitive to the steady state value of the input. Incoherent Feed-Forward Loops (IFFL) and Negative Feedback loops are two common molecular circuits that carry out adaptation in biology [35]. IFFLs have been recently implemented with DNA strand displacement reactions [36]. Recent work on buffers with finite response time [37] also resemble adaptation.

However, adaptive circuits in the biological literature often respond in an equal and opposite manner to both the rising and falling edges of each input pulse [34, 38]. To count pulses, we desire an asymmetric response to step ups and step downs in the input. For example, a circuit that responds only to step ups but ignores step downs and steady values of the input would naturally count pulses. Such asymmetric adaptation [39, 40] can be naturally achieved in a CRN if the adaptive variable has a resting concentration at zero; then step downs can have little effect since concentrations cannot fall below zero.

Here, we present a simple CRN implementation of such an adaptive Incoherent Feed-Forward Loop circuit with an asymmetric response to step ups and downs:

\[
I + A \overset{k_1}{\rightarrow} I + G \quad (1)
\]
\[
A \overset{\lambda_1}{\rightarrow} \phi \quad (2)
\]
\[
\bar{I} \overset{k_2}{\rightarrow} A \quad (3)
\]
\[
G \overset{\lambda_2}{\rightarrow} \phi \quad (4)
\]
\[
G \overset{k_4}{\rightarrow} P + G \quad (5)
\]

Here $\bar{I}$ refers to a species that is the negation of the input $I$ as in dual-rail logic [13, 41]; i.e., we assume a second input species $\bar{I}$ that is high when the input $I$ is low and vice-versa. In principle, such dual rail $(I, \bar{I})$ input can be created from a single input $I$ by a fast reusable NOT-gate [41]. One functional fast NOT-gate is provided in
FIG. 3. A duty fraction decoder implemented by a timed decay mechanism that computes a moving average. (a) Reaction network to decode duty fraction. (b) DNA Strand Displacement (DSD) implementation of mechanism in (a). See Methods for kinetic parameters and SI for circuit that includes waste products. (c) Sample time traces for a typical input signal $[I] = c(t)$ from simulations of the DNA network. (d) Changing the duty fraction of input $[I]$ changes steady state value of output $[P]$. (e,f) Changing time period $T$ or pulse count $n$ do not affect the readout $[P]$.

FIG. 4. A time period decoder implemented by a combining an Incoherent Feed-Forward Loop with a timed decay. (a) Reaction network to decode time period. (b) DNA Strand Displacement (DSD) implementation of mechanism in (a) with waste products suppressed. Species $\overline{I}$ is the negation of input $I$ as in dual-rail logic. (c) As with the pulse count decoder, $G$ shows a stereotyped response of fixed width $\tau_s$ to pulses of $I$. However, the output $P$ now computes the moving average of the stereotyped responses of $G$ over a fixed timescale set by its decay constant. Hence the output $P$ reflects the duty fraction of the stereotyped pattern in $G$ which is $\tau_s/T$. Consequently, changing the (d) duty fraction or (e) number of pulses in $I$ has no impact on the output $P$ which does change with (f) time period $T$. 
To understand the mechanism, consider the response to a single step up in $[I]$ shown in Fig 2c. Production of $G$ requires both $I$ and $A$ to be present; while $A$ is high in its resting state, $G$ starts being produced only when $I$ steps up. But meanwhile turning on $I$ also increases the degradation of $A$ on a timescale $1/(\lambda_2 + k_1[I])$, with $A$ eventually reaching a small steady state. As a result, $G$ stops being produced after a time

$$\tau_{up} = \frac{\log \lambda_2 - \lambda_1}{\lambda_2 - k_1[I] - \lambda_1}$$

and starts falling due to its own degradation timescale $1/\lambda_2$, thus producing a stereotyped response to the step up in $I$. This stereotyped response lasts a total time $\tau_a \sim \tau_{up} + 1/\lambda_2$. A subsequent step down in $I$ that occurs more than a time $\tau_a$ after the step up has minimal impact on $G$ because $G$ is already near zero. After the step down of $I$, $A$ is restored back to its resting value on a timescale $1/\lambda_1$. If an output species $P$ is produced in response to $G$, the total amount of $P$ will report the number of pulses $n$ without regard to time period $T$ or duty fraction $\delta$.

**Limits of operation:** There are two critical requirements for the mechanism above to work. First, the pulse width $T\delta$ must be larger than the length of the stereotyped adaptive response $\tau_a$, so that the stereotyped response is not interrupted by the step down in $I$. For our parameter choices, $\lambda_2 \approx k_1 \gg \lambda_1$, we can approximate $\tau_{up} \approx 1/\lambda_2$. Thus, $\tau_a \sim \tau_{up} + 1/\lambda_2 \sim 1/\lambda_2$ and we require that $T\delta > 1/\lambda_2$. Second, the pulse off time $T(1 - \delta)$ needs to be long enough so that $A$ can be restored to its resting state before the next pulse in $I$ comes, hence requiring $T(1 - \delta) > 1/\lambda_1$. Taken together, we require $T > T_{min} = \sup\left\{1/(\lambda_1, 1/\lambda_2)\right\}$. In our CRN with $\lambda_1 \ll \lambda_2, T_{min} = 1/\lambda_1$.

To verify that real chemical networks can operate in this kinetic regime, we designed a DNA strand displacement implementation of this CRN. As with all the DSD circuits we present here, our design process leaned on the reaction designs laid out in Soloveichik et al. [12]. A simplified representation of the reaction network with waste products suppressed is shown in Fig 2c; see SI for the full network. Simulating the DNA strand displacement network using Visual DSD software with realistic kinetic parameters, we find that the output is indeed sensitive to pulse number $n$ but insensitive to duty fraction $\delta$ and time period $T$ over a significant range.

### B. Duty fraction decoder

We now develop a circuit needed to decode duty fraction $\delta$. The Chemical Reaction Network capable of decoding $\delta$ is deceptively simple in topology,

1. **I** $\xrightarrow{k} \text{I} + \text{P}$
2. **P** $\xrightarrow{\lambda} \phi$

where $I$ is the time-varying input species whose concentration changes over time as $[I] = c(t)$. Species $P$ is created by every pulse of $I$ but decays with a time constant $1/\lambda$. For any $c(t)$, this simple linear system has an output $P$ given by,

$$P(t) = k \int_{-\infty}^{t} c(t') e^{-\lambda(t-t')} \, dt'.$$

Intuitively, this operation effectively takes the moving average of input species $I$ over a time window $1/\lambda$. If $\lambda$ is in the right range, the output $P$ is insensitive to $T, n$. For a pulse train $c(t)$ that starts at $t = 0$, we find that $A$ rises and oscillates about a steady state value; see Fig 3c-f. As desired, the output $A(t)$ is only sensitive to changes in duty fraction $\delta$ and insensitive to changes in $n, T$.

### C. Time period decoder

We can build a time period $T$ decoder by modifying the adaptive circuit motif introduced above; we simply add a decay process for species $P$ so that $[P]$ reflects the moving average of $G$ over a fixed timescale $1/\lambda_3$,

$$P \xrightarrow{\lambda_3} \phi.$$
FIG. 5. Output of DNA-based decoders with realistic kinetic parameters on a library of temporal patterns that systematically vary in pulse number, duty fraction and time period. We see that the output of each decoder circuit changes significantly in response to changes in one of the three features but is insensitive to the other two features. For example, the duty fraction decoder changes by a factor of 4× as duty fraction changes from 0.2 to 0.8 but only changes by a factor of 1.02 as time period changes from 5000 seconds to 15000 seconds. The breakdown (e.g., high duty fraction for time period decoder) is described by the limits of operation derived in the text. Thus, taken together, the decoders demonstrated here, with realistic DNA hybridization-based reaction rates, can distinguish temporal patterns accurately over a range of timescales.

All other reactions for I, A, G, P are as shown in Eqn.1 - 5. The analysis for this CRN looks identical to that done for our duty fraction sensor. However, now we are taking the moving average of \([G](t)\) which possesses the same periodicity \(T\) as the initial pulse train but has pulses with a duration set only by the kinetic parameters of the network \(\tau\). Thus, \([G](t)\) has an effective duty fraction \(\delta_{eff} = \tau/T\). If we tune kinetic parameters to the same regime from our duty fraction sensor, the output species \(P\) will then report this \(\delta_{eff}\) (up to geometric factor for the shape of the pulse) and is thus proportional to 1/T but independent of the \(\delta\) and \(n\) from the initial species I.

\textbf{Limits of operation:} This CRN directly inherits the kinetic constraints from both the pulse number decoder applied to the I, A, G part of the network and the duty fraction decoder applied to G, P part of the network. The former restricts \(T(1 - \delta) > 1/\lambda_1\) as discussed for the pulse number decoder. The latter constraint requires \(T < 1/\lambda_3\) and \(n > 1/(\lambda_3T)\) as discussed for the duty fraction decoder. In this kinetic regime, our output depends strongly on \(T\) and only weakly on \(n, \delta\).

The DNA implementation is shown in Fig 4b. By choosing kinetic constants listed in Methods, we simulated this network and see that the output is insensitive to changes in \(\delta, n\) but sensitive to changes in \(T\).

Finally, we systematically tested all three of the mechanisms proposed here against a library of temporal patterns that vary in all three features \((n, \delta, T)\). Each decoder was implemented with DNA strand displacement with fixed kinetic rates. See Fig 5. We see that each decoder shows a much larger response to changes in its relevant feature than to the other features over a substantial dynamic range. Thus, collectively, the three circuits can discriminate each member of the temporal pattern library.

\section{II. DISCUSSION}

The circuits introduced here exploit a weakness in the context of digital computation - internal transients - to naturally perform analog computation on temporally coded information. The performance of such analog ‘computation using transients’ is limited by the dynamic range over which timescales of transients can be tuned.
Hence DNA strand displacement reactions is particularly suitable for such computation since their kinetics can be tuned over a large dynamic range through toehold sequence design [43–46].

The methods developed here can be combined with other developments in the molecular technology field[47]. For example, a drug payload carried by a DNA origami pill [48] can be released only in those cells with a pulsatile pattern of the transcription factor NFkB that precedes a inflammatory response but not in cells with NFkB patterns that precede an adaptive immune response[49]. Similar in situ temporal computation using molecules can also help surveil complex ecosystems, such as the gut, where a future ecological collapse is often indicated by temporal precursors [49].

We have produced circuits that decode broadly relevant but predetermined temporal features. Going forward, it would be interesting to develop molecular circuits that can learn relevant temporal features dynamically [50–52] as in machine learning approaches. In the learning paradigm, for example, a molecular circuit could be exposed to two classes of time-varying patterns during a ‘training phase’ (the temporal equivalent of cat and dog images in static pattern recognition). The circuit would determine which temporal features can best distinguish those two classes.

ACKNOWLEDGMENTS

The authors thank Aaron Dinner, Anders Hansen, Kabir Husain, John Hopfield, Sidney Nagel and Michael Rust for useful discussions. We acknowledge nsf-mrsec 1420709 for funding and the University of Chicago Research Computing Center for computing resources. AM is grateful to the Simons Foundation’s Mathematical Modeling of Living Systems investigator program for support.

Appendix A: Methods

We first formulated abstract chemical networks with the desired feature detecting properties. Then, largely following the design principles laid out in [42], we implemented these CRNs as DNA strand displacement reactions. All strand displacement circuits are designed within the Visual DSD software described in [9] using default kinetic parameters (unless explicitly noted in the Supplemental Information) and concentrations $\in [0.5\text{ nM}, 1\text{ mM}]$. Then, by adapting the MATLAB code generated within this program, we exposed these circuits to the pulsatile inputs defined in the main text, defined by their duty fraction $\delta$, number of pulses $n$, and period $T$. All results shown are from deterministic simulations.

### Table I. Pulse Counter Initial Conditions

| Species | Initial Conc. (nM) |
|---------|--------------------|
| sp_0   | 100 (0)            |
| sp_6   | 10000              |
| sp_2   | 10                 |
| sp_10  | 10000              |
| sp_10  | 100 (0)            |
| sp_5   | 100                |
| sp_7   | 10000              |
| sp_8   | 10000              |
| sp_11  | 10000              |
| sp_12  | 100                |
| sp_13  | 10000              |

### Table II. Duty Fraction Sensor Initial Conditions

| Species   | Initial Conc. (nM) |
|-----------|--------------------|
| sp_0      | 1000 (0)          |
| sp_1      | 100                |
| sp_2      | 100                |
| sp_3      | 10                 |

### Table III. Period Sensor Initial Conditions

| Species   | Initial Conc. (nM) |
|-----------|--------------------|
| sp_0      | 100 (0)            |
| sp_6      | 10000              |
| sp_2      | 10                 |
| sp_10     | 10000              |
| sp_10     | 100 (0)            |
| sp_5      | 100                |
| sp_7      | 10000              |
| sp_8      | 10000              |
| sp_11     | 10000              |
| sp_12     | 100                |
| sp_13     | 10000              |
| sp_14     | 0.05               |

[1] S. Song, L. Wang, J. Li, C. Fan, and J. Zhao, Trends Anal. Chem. 27, 108 (2008).
[2] S. Modi, S. M G, D. Goswami, G. D. Gupta, S. Mayor, and Y. Krishnan, Nat. Nanotechnol. 4, 325 (2009).
FIG. 6. Pulse count decoder. Here we present the full DSD reaction network for our pulse counting circuit. See the main text for an analysis of its dynamics (Fig. 2) and performance (Fig. 5). The two species in red had their dynamics directly modulated to the parameters of the input series, with sp10 pulsing exactly out of phase with sp6. Graphs and labels are generated automatically within the Visual DSD software [9]. See Table 1 for a list of initial concentrations.
FIG. 7. Duty fraction decoder. This circuit effectively take the moving average of the dynamics of sp_0 and reports it in sp10. See main text Figure 3 for analysis and Figure 5 for performance. Initial concentrations of bolded species are listed in Table II.
FIG. 8. Time period decode. By taking the moving average of sp34, this circuit decodes the period of sp_0. See main text Figure 4 for analysis of dynamics and Figure 5 for analysis of performance. Initial concentrations of bolded species are reported in Table III.