RESEARCH

An Oscillating Reaction Network With an Exact Closed Form Solution in the Time Domain

Joseph Hellerstein

Correspondence: jlheller@uw.edu
eScience Institute, University of
Washington, Seattle, USA
Full list of author information is
available at the end of the article

Abstract

Background: Oscillatory behavior is critical to many life sustaining processes such as cell cycles, circadian rhythms, and notch signaling. Important biological functions depend on the characteristics of these oscillations (hereafter, oscillation characteristics or OCs): frequency (e.g., event timings), amplitude (e.g., signal strength), and phase (e.g., event sequencing).

Results: We conduct a theoretical study of an oscillating reaction network to quantify the relationships between oscillation characteristics and the structure and parameters of the reaction network. We develop a two species (S_1, S_2) reaction network whose dynamics can be described by a system of linear differential equations. We solve this system to obtain exact, closed-form formulas for our network's OCs. These formulas are used to develop the `designOscillator` algorithm that designs oscillators by finding values of parameters of our reaction network that achieve desired OCs.

Conclusions: The OC formulas are employed to analyze the roles of the reactions in the network, and to comment on other studies of oscillatory reaction networks. For example, others have stated that nonlinear dynamics are required to create oscillations in reaction networks. We refute this, at least in theory, since our oscillatory reaction network has linear dynamics. Our formulas also show that the rate of negative feedback must exceed the rate of positive feedback in our reaction network.

Keywords: Systems biology

Background

Oscillatory behavior is critical to many life sustaining processes. Examples include: cell cycles [?], circadian rhythms [?], notch signaling in the development of the nervous system [?], tissue development [?], and gene transcription [?]. Biological oscillators are also important elements in building applications in synthetic biology [?, ?, ?].

The characteristics of biological oscillations often have critical biological functions. Frequency is used to control the times at which events are initiated, such as circadian cycles and chromatin modifications [?]. Amplitude controls the strength of signaling [?]. Phase plays a role in the sequencing of processes within the cell cycle [?]. Since biological oscillators typically cause changes in the concentration of chemical species, these oscillators must have a DC offset so that values are non-negative. Collectively, we refer to frequency, amplitude, phase, and DC offset as **oscillation characteristics (OCs)**.

One way to understand the relationship between an oscillating reaction network and its OCs is to construct a closed-form, time-domain solution of the network's behavior in terms of parameters such as kinetic constants and initial concentrations of chemical species. From these mathematical expressions, we obtain insights such as: (a) if one or more reactions are unnecessary to achieve oscillations; (b) relationships between the kinetic constants of reactions; and (c) how to assign values to network parameters so as to achieve desired OCs.

Many researchers have investigated structural aspects of oscillating reaction networks. These structures include: positive feedback, negative feedback, balancing reaction rates, and ultrasensitivity [?, ?, ?, ?, ?]. Others have built biological oscillators [?, ?, ?, ?, ?, ?, ?]. But neither kind of investigation addresses our interest in a closed-form, time-domain solution that relates parameters of a reaction network to its OCs.

More relevant to our work are quantitative models of biological oscillators. For the most part, existing models are systems of nonlinear ordinary differential equations [?, ?, ?, ?, ?, ?]. The complexity of these models prohibits the construction of a closed-form, time-domain solution.

We are aware of two approaches that circumvent the limitations of nonlinear ODEs. The first uses an empirical approach, system identification, to construct a linear model that approximates the nonlinear system (e.g., [?]). Models constructed in this way provide accurate predictions near the operating point at which system identification is done. The second approach constructs a linear approximation to a nonlinear ODE (e.g., [?]). Typical approximations make assumptions about relative reaction rates and/or magnitudes of species concentrations. In both cases, the construction of a linear models greatly reduces the complexity of the mathematical expressions, and this in turn makes it possible to obtain a closed-form, time-domain solution. However, the approximations limit the extent to which the resulting mathematical expressions provide useful interpretations of how the parameters of the reaction network affect OCs.

The present work is a theoretical study in that we do not build a biological oscillator. Rather, we develop an oscillating reaction network whose dynamics are described by a system of linear ordinary differential equations (ODEs). The network is consistent with biological networks (hereafter, **biologically feasible**) in that: (a) species concentrations are not negative; and (b) reactions have rate laws that are used in models of biological systems (e.g., BioModels [?]). We solve the ODEs to obtain an *exact* closed-form, time-domain solution, and then construct formulas that relate OCs to the parameters of our reaction network. These formulas provide considerable insight into how parameters of the reaction network affect OCs, something that likely cannot be determined from the network structure alone. Further, we use the formulas to develop an algorithm for designing oscillators that achieve desired oscillation characteristics.

Methods

Our method is to propose a reaction network whose kinetics are linear and biologically feasible. We solve this system of linear ODEs (an initial value problem) to obtain formulas that relate OCs to parameters of the reaction network. Finally,

we develop an algorithm for designing an oscillator with desired OCs by properly assigning values to parameters of the reaction network.

Reaction Network

This section develops a biologically feasible reaction network whose kinetics can be described as a system of linear ODEs. A system of linear ODEs can be expressed in matrix form as:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}, \tag{1}$$

where $\mathbf{x}(t) = \{x_n(t)\}$ is an N dimension vector of time varying of species concentrations; $\dot{\mathbf{x}}(t)$ is the time derivative of $\mathbf{x}(t)$; $\mathbf{A} = \{a_{ij}\}$ is an $N \times N$ Jacobian matrix of constants; and $\mathbf{u}$ is an N dimensional vector of constants that are forced inputs. We want to construct a reaction network that has a sustained oscillation. Since this is a linear system, the oscillations will be sinusoids.

From the foregoing, we have the following constraints:

- **C1**: Rate laws in the reaction network are a linear function of the concentrations of $x_n(\mathrm{t})$.
- **C2**: $x_n(t) \geq 0$ so that the reaction network is biologically feasible.

We simplify the problem by having $N = 2$, since this is sufficient to obtain oscillations. This means that the eigenvalues of $\mathbf{A}$ must be pure imaginary numbers. Let $\tau = a_{11} + a_{22}$ be the trace of $\mathbf{A}$, and $\Delta = a_{11}a_{22} - a_{12}a_{21}$ be the determinant of $\mathbf{A}$. The eigenvalues are complex conjugates λ_1, λ_2 such that

$$\lambda_n = \frac{\tau}{2} \pm \frac{\sqrt{\tau^2 - 4\Delta}}{2}$$

Clearly, we obtain pure imaginary eigenvalues only if $\tau = 0$ and $\Delta > 0$. Thus, we add the constraints

- **C3**: $\tau = 0$, where τ is the trace of $\mathbf{A}$.
- **C4**: $\Delta > 0$, where Δ is the determinant of $\mathbf{A}$.

With these constraints, $\lambda_n = \pm\theta i$, where $\theta = \sqrt{\Delta}$.

We use $k_i \geq 0$ to specify kinetics constants, where i indexes the reaction in the network. Let S_n be a chemical species whose time varying concentration is the state variable $x_n(t)$ for $n = \{1, 2\}$. We start by having two-way interactions between the species. Rate laws are specified above the reaction arrow.

- $R_1 : S_1 \xrightarrow{k_1 S_1} S_2$
- $R_2 : S_2 \xrightarrow{k_2 S_2} S_1$

The foregoing reactions have mass action kinetics, which is widely used in models of chemical systems. With just these reactions, $\mathbf{A}$ is

$$\begin{pmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix}$$

Clearly, C3 does not hold: $\tau < 0$ since both terms in the diagonal are negative. To address this, we make a_{11} positive by adding an autocatalysis reaction, a kind of reaction that arises in many biological oscillators [**?**].

- $R_3 : S_1 \xrightarrow{k_3 S_1} 2S_1$

and so $\mathbf{A}$ becomes

$$\begin{pmatrix} k_3 - k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix}$$

which is positive with appropriate choices of the k_i.

Since R_3 synthesizes S_1, we need a reaction that degrades S_1 in order for the system to be stable. If we use mass action kinetics for this reaction, kS_1, it changes a_{11} to $k_1 - k_2 - k$, which makes it more difficult to satisfy C3. An alternative is a fixed degradation rate of $k_4 > 0$, where $k_4 = u_1$ is the first element of the vector $\mathbf{u}$ in Eq. (1).

- R_4: $S_1 \xrightarrow{k_4} \emptyset$

Examples of fixed rate degradation reactions in BioModels are: reaction `reaction_0` in `BIOMD0000000112`, reaction `ATP_Jerp` in `BIOMD0000000059`, and reaction `inhibition_parameter2` in `BIOMD0000000224`.

We must still address constraint C4, that the determinant is positive. This requires that $a_{12}a_{21} < 0$, which mean that one of a_{12}, a_{21} is positive and the other is negative. We make $a_{21} < 0$ by degrading S_2 at a rate controlled by S_1. That is,

- R_5: $S_2 \xrightarrow{k_5 S_1} \emptyset$

Having the degradation of one chemical species controlled by another chemical species is common in BioModels. Some examples are: the `fast` reaction in model `BIOMD0000000108`, reaction `r10` in `BIOMD0000000145`, and reaction `RuBisCO_5_EOP` in `BIOMD0000000392`. With the addition of reaction R_5, $\mathbf{A}$ becomes

$$\begin{pmatrix} k_3 - k_1 & k_2 \\ k_1 - k_5 & -k_2 \end{pmatrix}.$$

The final reaction in our network compensates for degrading S_2 by synthesizing this species at a fixed rate. That is,

- R_6: $\emptyset \xrightarrow{k_6} S_2$

(We do not explicitly cite examples of similar synthesis reactions since they are widely used in BioModels.) From this, we observe that

$$\mathbf{u} = \begin{pmatrix} -k_4 \\ k_6 \end{pmatrix}. \tag{2}$$

The full network is depicted in Fig. 1.

C3 and C4 constrain the values of the kinetic constants. From C3, we know that

$$k_3 = k_1 + k_2 \tag{3}$$

From C4, we know that $(k_3 - k_1)(-k_2) - k_2(k_1 - k_5) > 0$, or $k_3 < k_5$. We define $k_d > 0$ such that

$$k_5 = k_3 + k_d \tag{4}$$

And so, $k_5 = k_1 + k_2 + k_d$. This gives us

$$\mathbf{A} = \begin{pmatrix} k_2 & k_2 \\ -k_2 - k_d & -k_2 \end{pmatrix}. \tag{5}$$

And from this we calculate the determinant of $\mathbf{A}$:

$$\Delta = k_2 k_d. \tag{6}$$

And hence, the frequency θ is

$$\theta = \sqrt{k_2 k_d}. \tag{7}$$

Further, $k_2, k_d > 0$ implies that

$$\Delta > 0. \tag{8}$$

A brief technical note on Eq. (7). There are actually two solutions for θ, $\pm\sqrt{k_2 k_d}$. These solutions result in oscillations at the same frequency but with different phases. Our approach is to treat phase as a separate oscillation characteristic, and so we just use θ for the frequency.

We can now satisfy 3 of our 4 constraints. C1 is satisfied since all rate laws are linear in $x_n(t)$, the time varying concentrations of S_n. C2 is satisfied since $\tau = k_2 + (-k_2) = 0$. And, C4 is satisfied by Eq. (8). To address C2, that $x_n(t) \geq 0$, we must find the time domain solution of the reaction network.

Time Domain Solution

Solving Eq. (1) is an initial value problem, where the initial values of S_1, S_2 are $x_1(0), x_2(0)$. We proceed as follows: (a) solve the homogeneous equation $\dot{\mathbf{x}}^H = \mathbf{A}\mathbf{x}_h(t)$; (b) find a particular solution such that $\dot{\mathbf{x}}^P(t) = \mathbf{A}\mathbf{x}^P(t) + \mathbf{u}$; and (c) properly structure the complete solution $\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \mathbf{x}^H(t) + \mathbf{x}^P(t)$ so that we isolate terms for amplitude, frequency, phase, and DC offset. The derivation is a bit long, and so full details are reported in the supplemental material.

The solutions for $x_n(t)$ have the form

$$x_n(t) = \alpha_n cos(t\theta + \phi_n) + \omega_n, \tag{9}$$

where α_n is the amplitude of oscillation for $x_n(t)$, θ is the frequency in radians, ϕ_n is the phase in radians, and ω_n is the DC offset. $\alpha_n, \theta_n, \phi_n, \omega_n$ are functions of the k_i and $x_n(0)$. Table 1 displays the formulas for the OCs. Because of technical details related to the inverse tangent function, ϕ_n depends on the term π_n. (See the supplemental material for more on these technical details.) These terms are:

$$\begin{aligned} cond_1 &= \frac{k_2^2 x_1(0)}{k_2\theta + k_d\theta} + \frac{k_2^2 x_2(0)}{k_2\theta + k_d\theta} + \frac{k_2 k_4 \theta}{k_2\theta^2 + k_d\theta^2} - \frac{2k_2 k_4}{k_2\theta + k_d\theta} - \frac{k_2 k_6 \theta}{k_2\theta^2 + k_d\theta^2} \\ &+ \frac{k_2 k_6}{k_2\theta + k_d\theta} + \frac{k_2 k_d x_1(0)}{k_2\theta + k_d\theta} + \frac{k_4 k_d \theta}{k_2\theta^2 + k_d\theta^2} - \frac{2k_4 k_d}{k_2\theta + k_d\theta} + \frac{\theta x_2(0)}{k_2 + k_d} \end{aligned}$$

$$\pi_1 \quad = \quad \pi \text{ if } cond_1 < 0$$
$$= \quad 0 \text{ otherwise}$$

Similarly,

$$cond_2 \quad = \quad \frac{k_2 x_1(0)}{\theta} + \frac{k_2 x_2(0)}{\theta} - \frac{k_6}{\theta} + \frac{k_d x_1(0)}{\theta}$$

$$\pi_2 \quad = \quad \pi \text{ if } cond_2 > 0$$
$$= \quad 0 \text{ otherwise}$$

Studying Table 1, we note that only of a subset of the k_i appear in the formulas: $k_2, k_d (= k_5 - k_3), k_4, k_6, x_1(0), x_2(0)$. We refer to these as the **independent parameters** since they can be any non-negative value (although constraint C_3 may not be satisfied). k_3, k_5 are the **dependent parameters** in that they are calculated from the independent parameters.

Designing Oscillators

This section considers the problem of designing a biologically feasible oscillator with the OCs $\theta^\star, \alpha^\star, \phi^\star, \omega^\star$. We address this problem by developing an algorithm that selects values of independent parameters so that at least one of S_1, S_2 has these OCs.

Since we have a closed form formulas for the OCs, why not just invert these formulas to find the independent parameters that produce the desired OCs? The issue here is that the formulas are sufficiently complicated that it is extremely difficult to find an inverse symbolically.

Since a symbolic solution is intractable, we instead develop an algorithm that finds the inverse numerically. Our approach is to formulate an optimization problem. The objective is related to the desired OCs, $x^\star(t) = \alpha^\star sin(\theta^\star t + \phi^\star) + \omega^\star$. This means searching the space of possible values of independent parameters. We denote elements of this earch space by $\mathcal{P}$. We want to find $\hat{\mathcal{P}}$ in this space such that $x_n(t; \hat{\mathcal{P}})$ is close to $x^\star(t)$ for one of $n \in \{1, 2\}$.

How can we do this? Given a candidate $\mathcal{P}$, we can apply the formulas in Table 1 to calculate $x(t; \mathcal{P})$. The quality of this candidate is assessed using the loss function

$$L_n(\mathcal{P}) = \sum_t \left(x_n^\star(t) - x_n(t; \mathcal{P}) \right)^2 \tag{10}$$

And, so the solution to the optimization is

$$\hat{\mathcal{P}}_n \quad = \quad argmin_{\mathcal{P}} L_n(\mathcal{P})$$
$$\text{such that } x_1(t; \hat{\mathcal{P}}_n), x_2(t; \hat{\mathcal{P}}_n) \geq 0$$

To simplify our notation in the sequel, $\hat{x}_n(t) = x_n(t; \hat{\mathcal{P}}_n)$.

Our algorithm constructs a single sinusoid. Does it matter whether we use S_1 or S_2? We refer to the chosen species as the **oscillating species**. It turns out that

depending on the desired OCs, sometimes it is better to choose S_1, and other times S_2 works better. The algorithm chooses the oscillating species based on which loss function is smaller, $L_1(\hat{\mathcal{P}}_1)$ or $L_2(\hat{\mathcal{P}}_2)$. The chosen species is $\hat{x}(t)$.

We note that the optimization can be simplified a bit since we are given $\theta^\star$. The simplicity of the formula for θ in Eq. (7) allows us to calculate the independent parameter k_d in terms of k_2 given $\theta^\star$. We define $\mathcal{P}(\theta^\star)$ to be the parameters $\mathcal{P}$ with $k_d = \frac{(\theta^\star)^2}{k_2}$. By using $\mathcal{P}(\theta^\star)$ instead of the full $\mathcal{P}$, we reduce the number of parameters to estimate from 6 to 5. This significantly reduces the computational complexity of our algorithm.

Fig. 2 displays `designOscillator`, our algorithm for designing an oscillator for the reaction network in Fig. 1. The python implementation of the algorithm is in the module `designer.py` in the `github` repository for this project. The implementation uses the python package `lmfit` to find $\hat{\mathcal{P}}_n(\theta)$ using gradient descent.

We note several details in the implementation of `designOscillator`. First, it is important to do multiple calls to `lmfit` using randomly chosen initial values for the independent parameters to more effectively scan the search space of independent parameters $\mathcal{P}(\theta^\star)$. Second, in calculating the loss functions, it is critical to adjust the density of time values (t) to the value of $\theta^\star$ so that the algorithm sees at least two complete sinusoids at 10 or so different phases. Last, properly handling of the hard constraint $x_n(t) \geq 0$ is required so that gradient decent works effectively. We use relaxation [?], a technique that treats the hard constraints as soft constraints with large weights. This means that instead of using $L_n((\mathcal{P})$ in Eq. (10) as the loss function, we instead use $L_n^+(\mathcal{P})$:

$$L_n^+(\mathcal{P}) = \sum_t \left(x_n^\star(t) - \hat{x}_n(t)\right)^2 + \left(1_{\hat{x}_1(t)} w \hat{x}_1(t)\right)^2 + \left(1_{\hat{x}_2(t)} w \hat{x}_2(t)\right)^2 \qquad (11)$$

where 1_x is 0 if $x \geq 0$ and is 1 otherwise, and w is a large number. $L_n^+(\mathcal{P})$ has large gradients when $x_1(t) < 0$ or $x_2(t) < 0$ so that gradient descent moves away from regions of the search space in which $x_n(t) < 0$.

Observe that `designOscillator` could be implemented using simulation instead of the formulas in Table 1. However, doing so would require considerably more computational resources. First, simulating the reaction network is several orders of magnitude slower than evaluating the formulas in Table 1. Second, the simulation approach does not exploit the relationships between parameters that reduce the size of the search space (e.g., calculating k_d from $k_2, \theta^\star$). Thus, a simulation-based algorithm would have a search space consisting of 8 parameters (the k_i and the $x_n(0)$) in contrast to the 5 parameters used with the formulas based approach. This larger search space requires considerably more computational resources.

Results

We begin by investigating the accuracy of predictions made by our model as detailed by the formulas in Table 1. Fig. 3 displays simulations of the reaction network in Fig. 1 for four different values of the independent parameters. (Dependent parameters are calculated as described above.) Lines in the plots are simulation results, and the markers are model predictions. We see that in all cases, model predictions coincide with simulations results. We have done thousands of such simulations. In all

cases, model predications are identical with simulation results (within a tolerance of 10^{-13} to account for numerical errors). Although this does not *prove* the correctness of the model, it is strong confirmation. The ultimate proof is the correctness of our derivations as detailed in the supplemental material.

Next, we investigate the oscillator design algorithm in Fig. 2. Recall that the inputs to the algorithm are desired oscillation characteristics $\theta^\star, \alpha^\star, \phi^\star, \omega^\star$ that generate the sinusoidal concentrations $x^\star(t) = \alpha^\star sin(\theta^\star t + \phi^\star) + \omega^\star$. The design algorithm finds independent parameters $\hat{\mathcal{P}}(\theta^\star)$ that generate the concentrations $\hat{x}(t)$ for one of S_1, S_2.

We use the term **design error** to refer to deviations between $x^\star(t)$ and $\hat{x}(t)$. We consider several kinds of design errors. A **feasibility design errors** occurs if $\hat{x}_n(t) < 0$ for some n, t. An **amplitude design errors** is a deviation from $\alpha^\star$. Let $\hat{\alpha}$ be the amplitude achieved by $\hat{x}(t)$. The amplitude design error is calculated as $\frac{\hat{\alpha}-\alpha^\star}{\alpha^\star}$. Similarly, **phase design error** is the deviation from $\phi^\star$ of the phase (as a fraction of a cycle) achieved by $\hat{\mathcal{P}}(\theta)$. The phase design error is $\frac{\hat{\phi}-\phi^\star}{2\pi}$. Note that we do not consider errors in the frequency that result from the parameters returned by the design algorithm. This is because the algorithm uses Eq. (7), Eq. (3), and Eq. (4) to ensure that there are oscillating concentrations at the correct frequency, $\theta^\star$.

We simplify our studies by setting the desired DC offset $\omega^\star$ to the desired oscillation amplitude $\alpha^\star$. Our studies are conducted over 3 decimal orders of magnitude for both frequency (radians/sec) and amplitude: $\theta^\star, \alpha^\star \in [0.1, 100]$ in 8 increments. We consider four values of phase that are likely the most problematic because of challenges with calculating the inverse of the tangent function: $\phi^\star \in \{0, \frac{\pi}{2}, \pi, \frac{2\pi}{3}\}$. Throughout, we set the maximum value of the kinetic constants (k_i) to 1,000.

We begin with feasibility design errors. `designOscillator` is very robust to feasibility design errors. We have conducted several thousand simulations, and only a couple of them returned values of independent parameters that had negative values for the concentrations of S_1, S_2.

To analyze amplitude design errors, we first consider a variant of the algorithm in Fig. 2 in which the algorithm always returns $\hat{\mathcal{P}}_1(\theta)$, the parameter estimates obtained if S_1 is the oscillating species. Fig. 4 displays the results of these studies. There are four plots, one for each value of phase. Each plot is a heatmap with desired frequency ($\theta^\star$) as the horizontal axis and desired amplitude ($\alpha^\star$) as the vertical axis. Cells are colored by the magnitude of the design error, and they are annotated with the value of the design error followed by a letter. The letter "a" means that S_1 is the oscillating species; "b" indicates that S_2 is the oscillating species. Only the letter "a" appears in these plots since S_1 is always the oscillating species.

Amplitude design error is in the range $[-1, \infty]$. A 0 means that there is no design error; a -1 means that the oscillating species always has a concentration of zero (and so there is no oscillation). Amplitude design error is mostly 0 in the plots, except when both amplitude and frequency are large. The reason is that in our studies, the maximum value of kinetic constants is 1,000. When both $\theta^\star$ and $\alpha^\star$ are large, much larger values of k_4, k_6 are required. Fig. 5 plots the results of studies in which either S_1 or S_2 may be the oscillating species, as is done in `designOscillator`. We see that there is a significant reduction in amplitude design error.

Fig. 6 displays the results for phase design error (in units of the fraction of an oscillation cycle). We see that phase design errors are mostly 0, although occasionally there is an error of 0.1 or -0.1.

Fig. 7 studies the distribution of parameter values for the foregoing studies. Values have an upper bound of 1,000 in our studies (to avoid unrealistically large values for kinetic constants). Note that $k_1 = 1$ by design since from Table 1, we know that k_1 does not influence the behavior of the reaction network.

The parameters k_2, k_3, and k_5 are mass action kinetic constants for reactions with a single reactant. We see that they mostly have small values, although there are some instances in which these constants exceed 500. On the other hand, the zeroth order kinetic constants k_4 and k_6 tend to be much larger. Because our studies restrict `designOscillator` to choose k_4, k_6 with values less than 1,000, there are often larger amplitude design errors for studies in which both the frequency and amplitude are large.

Discussion

This section explores in more detail the oscillating reaction network in Fig. 1 using the formulas in Table 1.

We begin by examining the parameters of the reaction network. Consider k_1, the kinetic constant for reaction R_1. k_1 does not appear in Table 1. Indeed, the only reference to k_1 is in Eq. (3) to calculate k_3. So, k_1 can be any non-negative number. If $k_1 = 0$, we have effectively eliminated reaction R_1. We have done many thousands of simulations in which $k_1 = 0$ with various values of the independent parameters and calculating the dependent parameters as described above. In all cases, we obtain the oscillating networks predicted by Table 1. From this we conclude, that R_1 is not required.

Next, consider k_2, the kinetic constant that controls the rate at which S_2 is converted to S_1. That is, because of reaction R_2, if S_2 is larger at time t, then S_1 will be larger at a future time, say $t + t_d$ for $t_d > 0$. In essence, k_2 controls the phase shift from S_2 to S_1. We can see this using the results in Table 1. If k_2 is large, then the two species have the same phases: $lim_{k_2 \to \infty} \phi_1 = \frac{k_4 - k_6}{x_1(0) + x_2(0)} = lim_{k_2 \to \infty} \phi_2$. This phase shift in turn determines the frequency, since $\theta = \sqrt{k_2 k_d}$. So, a large rate at which there is a change in phase in turn results in higher frequency oscillations.

Now consider k_3, k_5, the kinetic constants for reactions R_3 and R_5. R_3 is an autocatalysis reaction in which S_1 produces two copies of itself. This is a kind of positive feedback at the rate $k_3 S_1$. R_5 is degradation reaction in which S_2 is eliminated at the rate $k_5 S_1$. This is a kind of negative feedback in that a larger concentration of S_1 reduces the concentration of S_2, which in turn reduces the concentration of S_1 (because of R_2). From Eq. (4), we know that $k_5 > k_3$. Indeed, when $k_5 = k_3$, then the only eigenvalue of the system is 0. So, negative feedback must be larger than positive feedback in our oscillating network.

Reactions R_4, R_6 have zeroth order kinetics. These are essentially external tuning knobs that adjust oscillation characteristics in complex ways. An extreme case is in the DC offsets ω_n: if $k_4 = 0 = k_6$, then $\omega_1 = 0 = \omega_2$. Note that k_4, k_6 appear in every formula for the OCs in Table 1 except frequency θ.

Finally, we address the initial concentrations of S_1, S_2, which are denoted by $x_n(0)$. We see that initial concentrations affect amplitude and phase. However, initial concentrations do not impact frequency and DC offset.

Next we comment on related work in light of our methods and results. We start with claims related to nonlinearity. [?] claims that "the kinetic rate laws of the reaction mechanism must be sufficiently 'nonlinear' to destabilize the steady state." This claim is echoed by [?] as well. Our work disputes the need for nonlinearity, at least in theory, since we construct an oscillating reaction network with kinetics described by a linear ODE.

Another remark worthy of comment is that oscillations can result from a combination of positive and negative feedback [?]. Indeed, Fig. 1 has both positive and negative feedback. Reaction R_3 provides positive feedback at a rate $k_3 S_1$ through autocatalysis, and reaction R_5 provides negative feedback at the rate $k_5 S_1$ by degrading S_2. We add to these remarks the observation that the rate of negative feedback must be larger than the rate of positive feedback, at least in our reaction network, since by Eq. (4), $k_5 > k_3$.

Our final point is a bit more speculative. Disciplines such as electrical and mechanical engineering make extensive use of system identification and control theory in their designs and analysis. A key element of these techniques is the use of frequency analysis such as Bode plots. Frequency analysis requires the ability to generate sinusoids with specific oscillation characteristics, especially frequency, amplitude, and phase. While others have demonstrated the construction of biological clocks (e.g., [?, ?]), there is no technique for generating sinusoids with arbitrary oscillation characteristics. Our `designOscillator` algorithm provides a way to choose values of the parameters of the reaction network in Fig. 1 to achieve desired OCs. An actual biological implementation of a sinusoid signal generator may require enzyme engineering techniques such as those discussed in [?].

Conclusions

Oscillatory behavior is critical to many life sustaining processes, such as cell cycles, circadian rhythms, and notch signaling. We use the term oscillation characteristics (OCs) to refer to the frequency, amplitude, phase, and DC offset of oscillations. OCs affect many biological functions. Examples include: the timing of event initiations can be controlled by oscillation frequency; the strength of signaling can be regulated by the amplitude of oscillations; and the sequence of events can be determined by the phase of oscillations.

This paper is a theoretical study of an oscillating reaction network to better understand the relationships between OCs and network structure and parameter values. We develop a two species (S_1, S_2) reaction network whose dynamics can be described by a system of linear differential equations. We solve this system to obtain formulas for its oscillation characteristics. The predictions of these formulas match simulation results for thousands of studies that we have conducted.

We use the OC formulas to develop the `designOscillator` algorithm that finds values of parameters of our reaction network that achieve desired OCs. Our evaluations of the algorithm shows that it works well over a wide range of desired frequencies, amplitudes, and phases.

The formulas are also employed to analyze the roles of the reactions in the network. An insight here is that our network is still oscillatory if one of the reactions is eliminated, a reaction that converts S_1 to S_2.

The OC formulas also allow us to comment on and/or extend other studies of oscillatory reaction networks. One theme in the literature is that nonlinear dynamics are required to create oscillations in reaction networks. We refute this, at least in theory, since our oscillatory reaction network has linear dynamics. Another observation in the literature is that an oscillating network can be constructed by combining positive and negative feedback. This is certainly true in our network. Our formulas extend this insight by showing that the rate of negative feedback must exceed the rate of positive feedback in our network.

We are very interested in exploring how our results for designing an oscillating linear network might be applied to designing an oscillating nonlinear network. One direction here is to approximate the nonlinear network with our linear network. If we can develop an appropriate mapping between the two networks, we can tune the linear network in a desired way, and then map these adjustments back to the nonlinear network.

Declarations

Abbreviations and Notation
Abbreviations
- OC - Oscillation characteristic

Notation
- $\mathbf{A}$ - Jacobian matrix
- α_n - amplitude of oscillation for species n
- $\alpha_n^\star$ - desired amplitude
- $\hat{\alpha}_n$ - amplitude of the oscillation produced by `designOscillator`
- $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ - constants associated with the homogeneous solution
- Δ - $det\mathbf{A}$) (determinant)
- $\mathbf{F}(t)$ - fundamental matrix
- i - indexes kinetic constants
- k_i, k_d - non-negative constants for kinetic constants
- λ_n - eigenvalue
- n - indexes states (chemical species)
- N - number of species
- ω_n - DC offset of species n
- $\omega_n^\star$ - desired DC offset
- $\hat{\omega}_n$ - DC offset of the oscillation produced by `designOscillator`
- $\mathcal{P}$ - the set of parameters and associated values for the reaction network in Fig. 1
- $\mathcal{P}(\theta)$ - the set of parameters with k_d determined by θ and k_2
- ϕ_n - phase in radians for species n
- $\phi_n^\star$ - desired phase
- $\hat{\phi}_n$ - phase of the oscillation produced by `designOscillator`
- S_n - chemical species in the reaction network
- t - time
- τ - $tr(\mathbf{A})$ (trace)
- θ - frequency in radians
- $\theta^\star$ - desired frequency
- $\hat{\theta}$ - frequency of the oscillation produced by `designOscillator`
- $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ - forced input (kinetic constants for zeroth order rates)
- $\dot{\mathbf{x}}(t)$ - derivative w.r.t. time of $\mathbf{x}$

- $x_n(t)$ - time varying concentration of species n
- $x_n^\star$ - desired signal
- $\hat{x}_n$ - signal produced the oscillator constructed `designOscillator`

Availability of data and materials

Supplementary material is available at `https://github.com/ModelEngineering/Oscillators/blob/main/docs/supplemental_material/supplemental_material.pdf`, and the github project for this paper is at `https://github.com/ModelEngineering/Oscillators`.

Ethics approval and consent to participate

Not applicable.

Competing interests

The author declares that he has no competing interests.

Consent for publication

Not applicable.

Authors' contributions

JLH conceived of the ideas in this paper, did all derivations and analyses, and wrote the paper.

Author details

eScience Institute, University of Washington, Seattle, USA.

References

1. Riva, S., Cazzaniga, P., Nobile, M., Spolaor, S., Rundo, L., Besozzi, D., Tangherloni, A.: SMGen: A generator of synthetic models of biochemical reaction networks. Symmetry **14**, 119 (2022). doi:10.3390/sym14010119
2. Schaffter, T., Marbach, D., Floreano, D.: GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. Bioinformatics (Oxford, England) **27**, 2263–70 (2011). doi:10.1093/bioinformatics/btr373
3. Shen-Orr, S., Milo, R., Mangan, S., Alon, U.: Network motifs in the transcriptional regulation network of Escherichia coli. Nature genetics **31**, 64–8 (2002). doi:10.1038/ng881
4. Alon, U.: Alon, U.: Network motifs: theory and experimental approaches. nat. rev. genet. 8, 450. Nature reviews. Genetics **8**, 450–61 (2007). doi:10.1038/nrg2102
5. Ronen, M., Rosenberg, R., Shraiman, B., Alon, U.: Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. Proceedings of the National Academy of Sciences of the United States of America **99**, 10555–60 (2002). doi:10.1073/pnas.152046799
6. Loskot, P., Atitey, K., Mihaylova, L.: Comprehensive review of models and methods for inferences in bio-chemical reaction networks. Frontiers in Genetics (2019)
7. Christopoulos, A., Lanzafame, A., Mitchelson, F.: Allosteric interactions at muscarinic cholinoceptors. Clinical and Experimental Pharmacology and Physiology **25**(3-4), 185–194 (1998). doi:10.1111/j.1440-1681.1998.t01-4-.x. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1681.1998.t01-4-.x
8. Nussinov, R., Tsai, C.J., Liu, J.: Principles of allosteric interactions in cell signaling. Journal of the American Chemical Society **136** (2014). doi:10.1021/ja510028c
9. Liebermeister, W., Uhlendorf, J., Klipp, E.: Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. Bioinformatics (Oxford, England) **26**, 1528–34 (2010). doi:10.1093/bioinformatics/btq141
10. Rohwer, J., Hanekom, A., Hofmeyr, J.-H.: A universal rate equation for systems biology, pp. 175–187 (2007)
11. Liebermeister, W., Klipp, E.: Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints. Theoretical biology & medical modelling **3**, 41 (2006). doi:10.1186/1742-4682-3-41
12. Wagner, M.: Enzyme kinetics, behavior and analysis of rapid equilibrium and steady-state enzyme systems (Segel, Irwin H.). Journal of Chemical Education **53**(11), 472 (1976). doi:10.1021/ed053pA472. https://doi.org/10.1021/ed053pA472
13. Bisswanger, H.: Enzymkinetik: Theorie und Methoden. Wiley-Blackwell, German (2000)
14. Cornish-Bowden, A.: Fundamentals of Enzyme Kinetics. Portland Press Ltd., London (2004)
15. Snoep, J.: The silicon cell initiative: Working towards a detailed kinetic description at the cellular level. Current opinion in biotechnology **16**, 336–43 (2005). doi:10.1016/j.copbio.2005.05.003
16. Bakker, B., Michels, P., Opperdoes, F., Westerhoff, H.: Glycolysis in bloodstream form Trypanosoma brucei can be understood in terms of the kinetics of the Glycolytic enzymes. The Journal of biological chemistry **272**, 3207–15 (1997). doi:10.1074/jbc.272.6.3207
17. Helfert, S., Estévez, A.M., Bakker, B., Michels, P., Clayton, C.: Roles of triosephosphate isomerase and aerobic metabolism in Trypanosoma brucei. The Biochemical journal **357 Pt 1**, 117–25 (2001)
18. Lam, C.F., Priest, D.G.: Enzyme Kinetics: Systematic Generation of Valid King-Altman Patterns. Biophysical Journal **12**(3), 248–256 (1972). doi:10.1016/S0006-3495(72)86084-3
19. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U.: COPASI—a complex pathway simulator. Bioinformatics (Oxford, England) **22**, 3067–74 (2007). doi:10.1093/bioinformatics/btl485
20. Funahashi, A., Jouraku, A., Matsuoka, Y., Kitano, H.: Integration of cell designer and SABIO-RK. In silico biology **7**, 81–90 (2007)
21. Dräger, A., Zielinski, D., Keller, R., Rall, M., Eichner, J., Palsson, B., Zell, A.: SBMLsqueezer 2: Context-sensitive creation of kinetic equations in biochemical networks. BMC systems biology **9**, 68 (2015). doi:10.1186/s12918-015-0212-9

22. Keating, S., Waltemath, D., Helikar, T., Hucka, M., co-authors, a.: SBML level 3: an extensible format for the exchange and reuse of biological models (2020)

23. Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., Dumontier, M., Finney, A., Golebiewski, M., Hastings, J., Hoops, S., Keating, S., Kell, D., Kerrien, S., Lawson, J., Lister, A., Lu, J., Machne, R., Mendes, P., Le Novere, N.: Controlled vocabularies and semantics in systems biology. Molecular systems biology **7**, 543 (2011). doi:10.1038/msb.2011.77

24. Dräger, A., Hassis, N., Supper, J., Schroeder, A., Zell, A.: SBMLsqueezer: A CellDesigner plug-in to generate kinetic rate equations for biochemical networks. BMC systems biology **2**, 39 (2008). doi:10.1186/1752-0509-2-39

25. Le Novere, N., Finney, A., Hucka, M., Bhalla, U., Campagne, F., Crampin, E., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J., Spence, H., Wanner, B.: Minimum information requested in the annotation of biochemical models (MIRIAM). Nature biotechnology **23**, 1509–15 (2006). doi:10.1038/nbt1156

26. Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I., Müller, W.: SABIO-RK - database for biochemical reaction kinetics. Nucleic acids research **40**, 790–6 (2011). doi:10.1093/nar/gkr1046

27. Jupp, S., Burdett, T., Leroy, C., Parkinson, H.E.: A new ontology lookup service at EMBL-EBI. In: SWAT4LS (2015)

28. Bornstein, B.J., Keating, S.M., Jouraku, A., Hucka, M.: LibSBML: an API library for SBML. Bioinformatics **24**(6), 880–881 (2008)

29. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H.: Systems Biology Markup Language (SBML) Level 1: Structures and Facilities for Basic Model Definitions. California Institute of Technology. Available from COMBINE at http://identifiers.org/combine.specifications/sbml.level-1.version-2 (2003). `http://identifiers.org/combine.specifications/sbml.level-1.version-2`

30. Hucka, M., Bergmann, F.T., Dräger, A., Hoops, S., Keating, S.M., Le Novère, N., Myers, C.J., Olivier, B.G., Sahle, S., Schaff, J.C., Smith, L.P., Waltemath, D., Wilkinson, D.J.: Systems Biology Markup Language (SBML) Level 2 Version 5: Structures and facilities for model definitions. Journal of Integrative Bioinformatics **12**(2), 271 (2015). doi:10.2390/biecoll-jib-2015-271

31. Hucka, M., Bergmann, F.T., Chaouiya, C., Dräger, A., Hoops, S., Keating, S.M., König, M., Le Novère, N., Myers, C.J., Olivier, B.G., Sahle, S., Schaff, J.C., Sheriff, R., Smith, L.P., Waltemath, D., Wilkinson, D.J., Zhang, F.: The Systems Biology Markup Language (SBML): Language specification for Level 3 Version 2 Core Release 2. Journal of Integrative Bioinformatics **16**(2), 20190021 (2019). doi:10.1515/jib-2019-0021

32. Golightly, A., Wilkinson, D.: Bayesian inference for stochastic kinetic models using a diffusion approximation. Biometrics **61**, 781–8 (2005). doi:10.1111/j.1541-0420.2005.00345.x

33. Galagali, N., Marzouk, Y.: Bayesian inference of chemical kinetic models from proposed reactions. Chemical Engineering Science **123**, 170–190 (2015). doi:10.1016/j.ces.2014.10.030
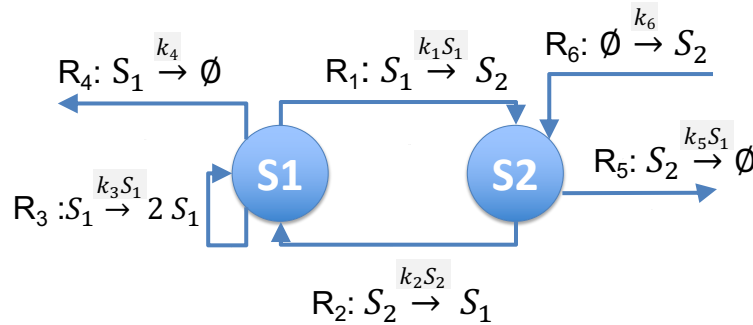
Figures

Figure 1 Reaction network that creates oscillations in the chemical species S_1, S_2. The reaction network is designed so that its time domain solution is a system of linear differential equations. The text describes constraints on the kinetic constants (k_i) and initial conditions of the chemical species to create an oscillator such that species concentrations are non-negative, a requirement for biological feasibility.

function designOscillator$(\theta^\star, \alpha^\star, \phi^\star, \omega^\star)$

1 Construct the reference signal: $x^\star(t) = \alpha^\star sin(t\theta^\star + \phi^\star) + \omega^\star$.

2 $\mathcal{P}(\theta) = \mathcal{P}$ with $k_d = \frac{(\theta^\star)^2}{k_2}$.

3 Given $\mathcal{P}(\theta)$, we calculate $x_n(t; \mathcal{P}(\theta))$, the concentrations of S_n using the formulas in Table 1. From this, we calculate the loss functions using relaxation (item (b) below):
    (a) $L_n(\mathcal{P}(\theta)) = \sum_t \delta(t) (x^\star(t) - x_n(t; \mathcal{P}(\theta)))^2$
    (b) where $\delta(t) = 1$ if $x_n(t; \mathcal{P}(\theta)) \geq 0$; otherwise, δ is a large number.

4 Find $\hat{\mathcal{P}}_n(\theta) = argmin_{\mathcal{P}(\theta)} L_n(\mathcal{P}(\theta))$

5 Select the independent parameters.
    (a) If $L_1(\hat{\mathcal{P}}_1(\theta)) < L_2(\hat{\mathcal{P}}_2(\theta))$, return $\hat{\mathcal{P}}_1(\theta)$
    (b) Else return $\hat{\mathcal{P}}_2(\theta)$

Figure 2 Algorithm for finding values of parameters of the reaction network that achieve desired oscillations. The function `designOscillator` takes as inputs the desired oscillator characteristics and returns the independent parameters $\mathcal{P}(\theta)$ for the reaction network: $k_2, k_4, k_6, x_1(0), x_2(0)$. In essence, `designOscillator` inverts $x_n(t)$ by finding the $\mathcal{P}$ that minimizes the squared error difference between the desired oscillations and $x_n(t)$ for either $n = 1$ or $n = 2$. In step 2, "relaxation" (via $\delta(t)$) is used to address the hard constraints that $x_n(t) \geq 0$.
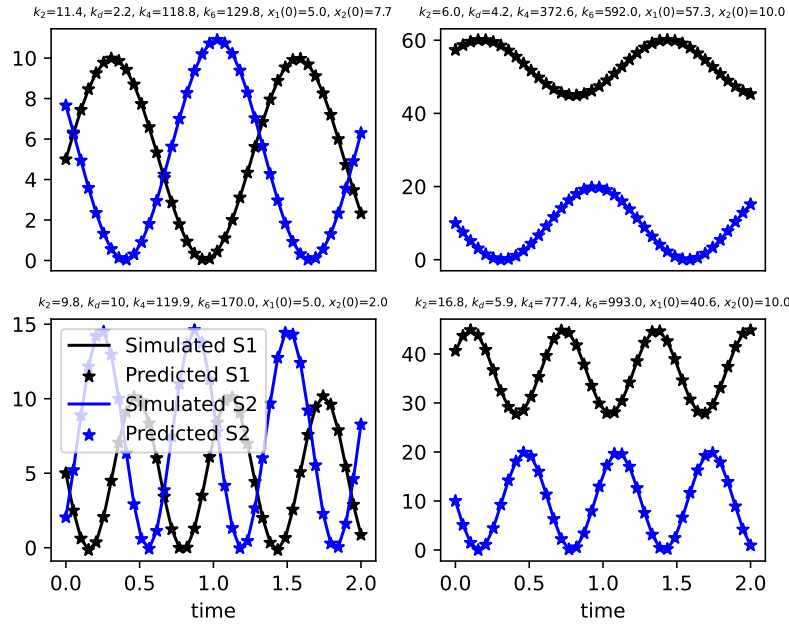
Figure 3 Evaluations of model accuracy. Plots of simulation results (solid lines) and model predictions (markers) using the formulas in Table 1 for four sets of parameter values of the reaction network in Fig. 1. In all cases, predictions coincide with the simulation results. The four cases have different values for the the independent parameters; dependent parameters are calculated as described in the text. The initial value of S_n is $x_n(0)$.

OCs	Formula
α_1	$\dfrac{\sqrt{\theta^2\left(k_2^2 x_1(0)+k_2^2 x_2(0)-k_2 k_4+k_2 k_d x_1(0)-k_4 k_d+\theta^2 x_2(0)\right)^2+\left(k_2^2 k_4-k_2^2 k_6+k_2 k_4 k_d+k_2\theta^2 x_1(0)-k_6\theta^2+k_d\theta^2 x_1(0)\right)^2}}{\theta^2(k_2+k_d)}$
α_2	$\dfrac{\sqrt{\theta^2\left(k_2 x_1(0)+k_2 x_2(0)-k_6+k_d x_1(0)\right)^2+\left(k_2 k_4-k_2 k_6+k_4 k_d-\theta^2 x_2(0)\right)^2}}{\theta^2}$
θ	$\sqrt{k_2 k_d}$
ϕ_1	$\tan^{-1}\left(\dfrac{k_2^2 k_4-k_2^2 k_6+k_2 k_4 k_d+k_2\theta^2 x_1(0)-k_6\theta^2+k_d\theta^2 x_1(0)}{\theta\left(k_2^2 x_1(0)+k_2^2 x_2(0)-k_2 k_4+k_2 k_d x_1(0)-k_4 k_d+\theta^2 x_2(0)\right)}\right)+\pi_1$
ϕ_2	$\tan^{-1}\left(\dfrac{k_2 k_4-k_2 k_6+k_4 k_d-\theta^2 x_2(0)}{\theta(k_2 x_1(0)+k_2 x_2(0)-k_6+k_d x_1(0))}\right)+\pi_2$
ω_1	$\dfrac{-k_2^2 k_4+k_2^2 k_6-k_2 k_4 k_d+k_6\theta^2}{\theta^2(k_2+k_d)}$
ω_2	$\dfrac{k_2 k_4-k_2 k_6+k_4 k_d}{\theta^2}$

Table 1 Formulas for oscillator characteristics (OCs) in terms of the kinetic constants k_i of the reaction network in Fig. 1 and the initial concentrations of the chemical species, $x_n(0)$ for species S_n. The formulas are obtained by solving the system of equations for the reaction network. The oscillator characteristics (OCs) are: amplitude (α_n), frequency (θ), phase (ϕ_n), and DC offset (ω_n). The terms π_1, π_2 are defined in the text, and reflect technical details related to the inverting the tangent function.
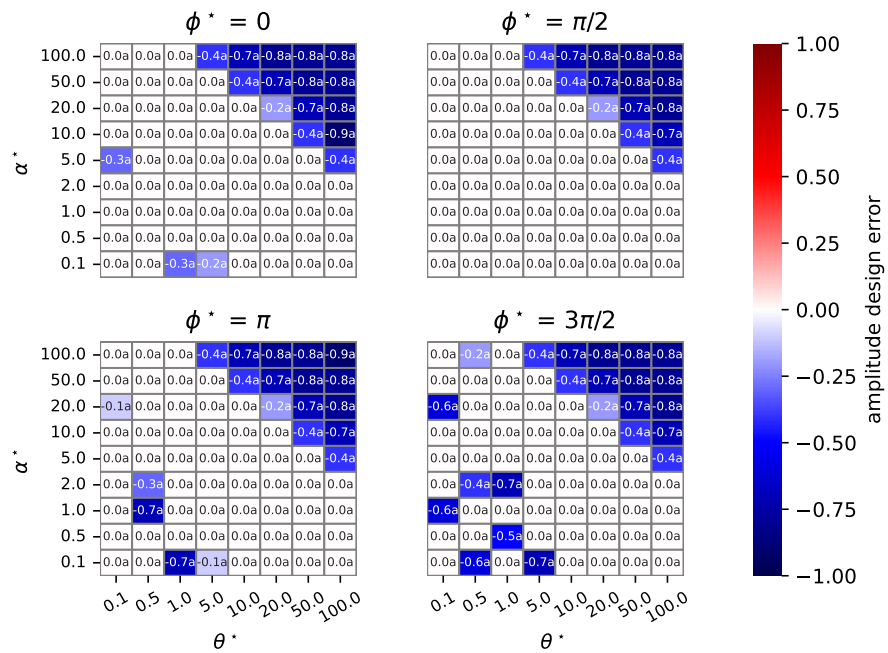
Figure 4 Amplitude design errors when `designOscillator` in Fig. 2 is modified so that S_1 is always the oscillating species. The heatmaps display amplitude design errors for four phases ($\phi^\star$). In each heatmap, the horizontal axis is desired frequency ($\theta^\star$), and the vertical axis is desired amplitude ($\alpha^\star$). Cell colors indicate the magnitude of the amplitude design error, and cells are annotated with the actual value. The letter "a" indicates that species S_1 is the oscillating species.
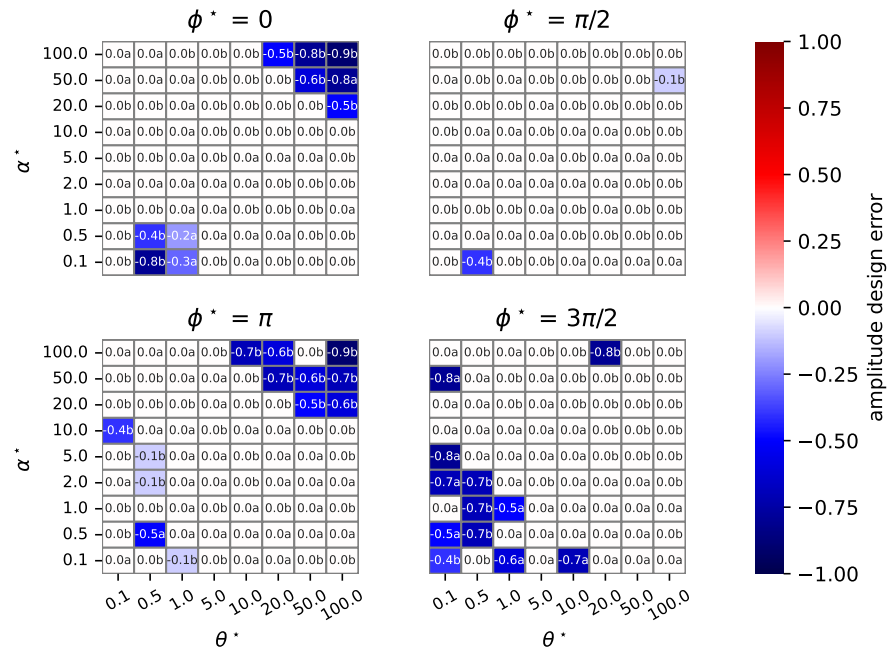
Figure 5 Amplitude design errors are reduced when `designOscillator` is used as-is so that either S_1 or S_2 can be the oscillating species. The heatmaps are structured as in Fig. 4. The letter "a" indicates that species S_1 is the oscillating species, and "b" indicates that S_2 is the oscillating species.
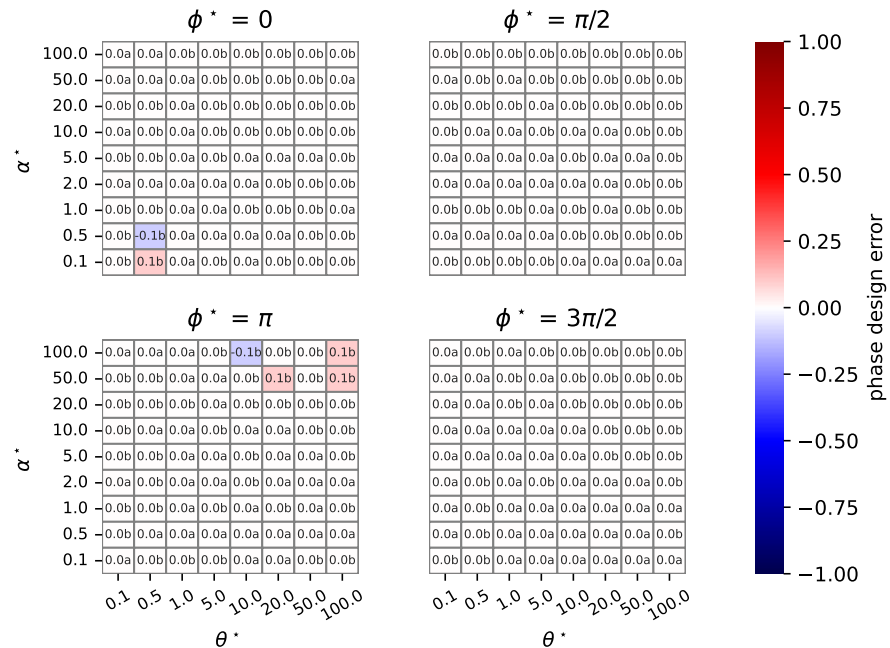

Figure 6 Phase design error. These heatmaps are organized in the same way as Fig. 4, but cell values are phase design errors, the fraction of a cycle that the phase of the designed network differs from the phase of the desired oscillations. `designOscillator` almost always produces a phase design error of 0.
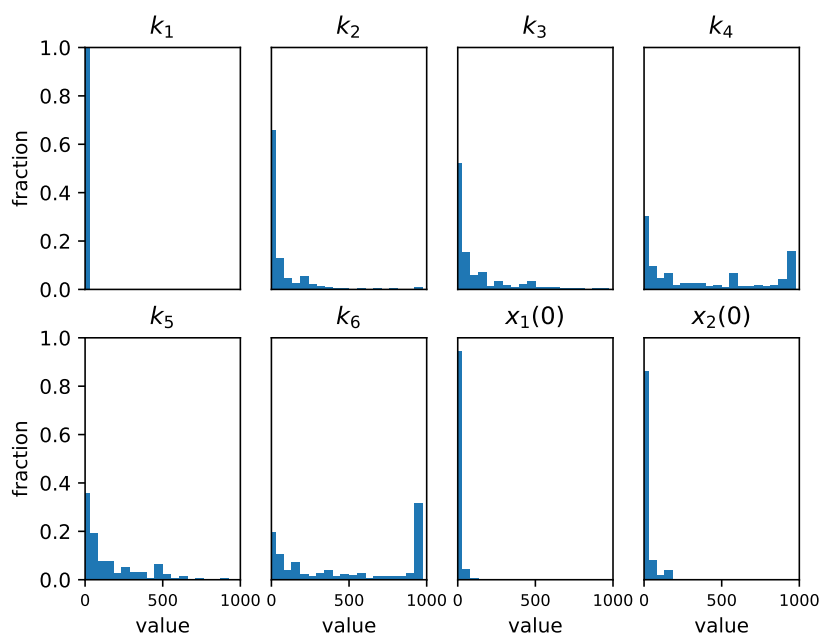
Figure 7 Histograms of values of parameter estimates in the numerical studies. The algorithm limits parameter values to the range $[0, 1000]$. Only the parameters associated with the boundary reactions R_4 (k_4) and R_6 (k_6) have values close to the upper bound of this range. These larger values are needed to construct oscillators that produce large amplitudes and high frequencies.