Noise control for molecular computing

Tomislav Plesa[1], Konstantinos C. Zygalakis[2], David F. Anderson[3], Radek Erban[1]*

Abstract

Synthetic biology is a growing interdisciplinary field, with far-reaching applications, which aims to design biochemical systems that behave in a desired manner. With the advancement in nucleic-acid-based technology in general, and strand-displacement DNA computing in particular, a large class of abstract biochemical networks may be physically realized using nucleic acids. Methods for systematic design of the abstract systems with prescribed behaviors have been predominantly developed at the (less-detailed) deterministic level. However, stochastic effects, neglected at the deterministic level, are increasingly found to play an important role in biochemistry. In such circumstances, methods for controlling the intrinsic noise in the system are necessary for a successful network design at the (more-detailed) stochastic level. To bridge the gap, the *noise-control algorithm* for designing biochemical networks is developed in this paper. The algorithm structurally modifies any given reaction network under mass-action kinetics, in such a way that (i) controllable state-dependent noise is introduced into the stochastic dynamics, while (ii) the deterministic dynamics are preserved. The capabilities of the algorithm are demonstrated on a production-decay reaction system, and on an exotic system displaying bistability. For the production-decay system, it is shown that the algorithm may be used to redesign the network to achieve noise-induced multistability. For the exotic system, the algorithm is used to redesign the network to control the stochastic switching, and achieve noise-induced oscillations.

Keywords

Synthetic biology, biochemical engineering, molecular/DNA computing, chemical reaction networks, stochastic dynamical systems, intrinsic noise

[1] *Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, UK*
[2] *School of Mathematics, The University of Edinburgh, Maxwell Building, Peter Guthrie Tait Road, Edinburgh, UK*
[3] *Department of Mathematics, University of Wisconsin–Madison, Lincoln Drive, Madison Wi, USA*
*__Corresponding author__: erban@maths.ox.ac.uk

1. Introduction

Synthetic biology is an interdisciplinary field of science and engineering that aims to construct biochemical systems with prescribed behaviors [1, 2]. At the theoretical level, the synthetic systems may significantly enhance our understanding of biology. At the practical level, they may have broad applications, e.g. in medicine [3, 4, 5, 6, 7], industry [8, 9], and nanotechnology [10, 11]. The systems may also be of interest to space agencies for optimizing extraterrestrial explorations [12]. A proof-of-concept for synthetic biology is a synthetic oscillator called the repressilator, which was implemented in vivo [13]. The experimental advances since the repressilator range from isolated synthetic biochemical networks, to microorganisms containing partially, or even fully, synthetic DNA molecules (synthetic life) [14, 15, 16, 17]. Examples include microorganisms containing a synthetic bistable switch [18], and a cell-density controlling quorum sensor [19], microorganisms producing antimalarial drugs [6, 7], and synthetic systems designed for tumor detection, diagnosis and adaptive drug-response [4, 5].

The construction of biochemical networks in synthetic biology may be broken down into two steps: firstly, an abstract system is constructed, displaying prescribed properties, and taking the form of a chemical reaction network [20, 21, 22]. Secondly, the abstract network is mapped to a suitable physical network, which may then be integrated into a desired environment (e.g. a test-tube, a vesicle, or a living cell) [23, 24, 25, 26]. Let us note that the first step generally consists of a number of sub-steps, involving mathematical analyses and computational verifications, depending on the nature of the target physical network [23, 51] (see also Section 2.3 and supplementary material).

In the first step of network construction, the goal is to obtain an abstract network with desired dynamics. In this paper, we consider reaction networks under mass-action kinetics: it is assumed that each reaction fires at the rate proportional to the product of the concentrations of the underlying reacting species. In this setting, we consider two dynamical models of reaction networks [27, 22]: the deterministic model, and the stochastic model (see supplementary material for more details). The deterministic model takes the form of the reaction-rate equations, which are ordinary-differential equations governing the time-evolution of the species concentrations [22, 27]. The stochastic model takes the form of a Markov chain, which may be simulated using the Gillespie stochastic simulation algorithm [28]. The Gillespie algorithm generates random copy-number time-series, with the copy-

number distribution matching that obtained from the underlying chemical master equation [22, 27, 28, 29]. The stochastic model is more-detailed, taking into an account the discreteness of the species counts, and the stochastic nature of the dynamics, which may be particularly important in biochemistry, where reaction networks may contain low-abundance species [33, 34, 13, 18, 21, 30, 31, 32]. On the other hand, the deterministic model is less-detailed, and more appropriate when the species are in high-abundance, and the discreteness and stochasticity are negligible [35].

In the second step of network construction, the goal is to engineer a physical network whose dynamics match well the dynamics of a given abstract network, over a suitable time-interval. Engineering an appropriate physical network may proceed indirectly, by altering a preexisting physical network, or directly, by engineering a network, which involves a given set of physical species, from scratch. The advantage of the former approach is that a preexisting network may display (partially) desirable dynamical properties. However, such a network may involve DNA and RNA molecules, proteins, and metabolites [2], some of which may have complex biophysical properties. Consequently, the disadvantage is that the structure (and, thus, the dynamics) of such a network cannot generally be modified in an arbitrary manner. In the latter approach, one may choose the physical species, at the expense of having to build a network from scratch. This approach is followed in the subfield of nucleic-acid-based molecular computing. For example, in DNA computing, physical networks are engineered with chemical species consisting exclusively of DNA molecules, interacting via the toehold-mediated DNA strand-displacement mechanism [23]. DNA production is systematic and cost-effective, and due to the fact that DNA biophysics is relatively well-understood, one has more freedom in controlling the structure of corresponding physical networks. More precisely, an abstract network under mass-action kinetics may be mapped to a DNA-based physical network provided it consists of up to second-order reactions, with rate coefficients varying over up to six orders of magnitude. The resulting physical network has identical deterministic dynamics as the abstract network (in the asymptotic limit of some of the kinetic parameters [23]), up to a scaling of the dependent variables. A proof-of-concept for DNA computing is a synthetic oscillator called the displacillator, which was implemented in vitro [36]. Let us note that DNA-based networks may also be augmented with enzymes [26, 37]. Another emerging approach within nucleic-acid-based molecular computing is based on RNA strand-displacement [38] - a mechanism which is hypothesized to occur naturally within living cells [39].

The DNA-based reaction networks may involve only high-abundance species, mixed in a test-tube [23]. In such circumstances, it may be sufficient to construct the networks via the (less-detailed) deterministic model. However, recent experimental advancements, involving compartmentalized circuits [24, 25, 26], localized circuits [40, 41] and molecular robots [42, 43], may require reaction network construction via the (more-detailed) stochastic model. For example, in [24, 25, 26], the chemical mixture from a test-tube is split into a large number of cell-size vesicles (allowing for an experimental investigation of biochemistry in cell-like reactors). This corresponds to replacing a given reaction network, involving only high-abundance species, with a large number of topologically equivalent networks which, however, may involve species in a low-abundance, making the (intrinsic) noise an important part of the dynamics. The intrinsic noise may be controlled in two ways: it may be decreased (e.g. as in [34]), in order to reduce the differences between the stochastic and deterministic dynamics. On the other hand, it may be increased, in a state-dependent manner, in order to favorably change the stochastic dynamics. In the language of molecular computing, the latter approach corresponds to exploiting the proven computational power of the stochastic reaction networks [44], by reprogramming the underlying intrinsic noise. Let us note that exploitations of the noise for enhancing biological functions have been reported in applications [33, 32]. In this paper, we follow the latter approach, and present the noise-control algorithm (given as Algorithm 1) which maps an input reaction network to output networks whose stochastic dynamics have an additional controllable state-dependent noise. Importantly, the input and output networks have an identical deterministic model in appropriate limits of some of the parameters introduced by the algorithm. The algorithm may play a significant role in the biochemical network synthesis, allowing for a deterministic-stochastic hybrid approach. More precisely, when constructing abstract and physical networks, one may use the deterministic model to guide the construction [20, 21], and then apply the algorithm to favorably modify the intrinsic noise in the stochastic model, while preserving the desired deterministic dynamics. The algorithm may also be used to adjust the intrinsic noise to favorably interact with environment-induced effects (e.g. extrinsic noise).

The rest of the paper is organized as follows. In Section 2, we introduce Algorithm 1 by applying it to the test network (1), which at the deterministic level displays a globally attracting equilibrium point. We show that the algorithm can favorably modify the stationary probability distribution underlying (1) at arbitrary points of the state-space, without influencing the deterministic dynamics. For example, it is shown that the algorithm may be used to redesign (1) to achieve noise-induced multimodality (multistability). In Section 3, we apply Algorithm 1 to the exotic network (11), which at the deterministic level displays a bistability involving an equilibrium point and a limit cycle. The algorithm is used to redesign (11) to increase the stochastic switching between the two attractors, and to achieve noise-induced oscillations. Finally, in Section 4 we conclude with a summary and discussion. The notation used in the paper is introduced as needed, and is summarized at the beginning of supplementary material.

2. A One-species Regular System

Consider the one-species production-decay reaction network $\hat{\mathscr{R}}(s)$, given by (1).

$$\hat{\mathscr{R}}(s): \quad \varnothing \xrightarrow{k_1} s, \qquad s \xrightarrow{k_2} \varnothing, \quad (1) \qquad \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = k_1 - k_2\hat{x}, \qquad \hat{x}(0) = \hat{x}_0. \quad (2)$$

Species s from network (1) reacts according to the two reactions with rate coefficients $k_1, k_2 \in \mathbb{R}_{\geq}$, where $\mathbb{R}_{\geq}$ is the set of nonnegative real numbers, and $\varnothing$ is the zero-species (denoting species which are not of interest). In this paper, we assume reaction networks are under mass-action kinetics, with the reactions taking place in unit-volume reactors. Let us denote the concentration of species s from (1) at time $t \in \mathbb{R}_{\geq}$ by $\hat{x} = \hat{x}(t) \in \mathbb{R}_{\geq}$. The initial value problem for the deterministic model (also called the drift) for network (1) is given by system (2), with $\hat{x}_0 \geq 0$ (see also supplementary material). Since the deterministic model (2) has a globally attracting equilibrium point, given by k_1/k_2, network (1) is said to be regular [22].

Let us denote the copy-number of species s from (1) at time $t \geq 0$ by $\hat{X}(t) \in \mathbb{Z}_{\geq}$, where $\mathbb{Z}_{\geq}$ is the set of nonnegative integers. Under the stochastic model, $\hat{X}(t)$ is modelled as a continuous-time discrete-space Markov chain (see also supplementary material), whose realizations can be generated by using the Gillespie stochastic simulation algorithm [28]. Given $\hat{X}(t)$, there will be a mean interevent time until one of the reactions from (1) fires. The mean interevent time is given by $1/\hat{\alpha}(\hat{X}(t))$, and when the event takes place, the probability that the i-th reaction from (1) fires is $\hat{\alpha}_i(\hat{X}(t))/\hat{\alpha}(\hat{X}(t))$, for $i \in \{1,2\}$. Here, $\hat{\alpha}_1(x) = k_1$, and $\hat{\alpha}_2(x) = k_2 x$, are the so-called propensity functions of the first, and second, reactions from (1), respectively. The function $\hat{\alpha}(x) = k_1 + k_2 x$ is the total propensity function of network (1), i.e. the sum of propensity functions of all the underlying reactions.

We now wish to structurally modify network (1) in such a way that (i) the deterministic model from (2) is preserved, while (ii) a control is gained over the interevent time from the stochastic model. We accomplish this by, firstly, imposing a conservation law on the target species s from network (1), thereby truncating its state-space, $\hat{X}(t) \leq C$, where $C \in \mathbb{Z}_{>}$ is a conservation constant. The conservation law is imposed in such a way that the total propensity function of the resulting network, denoted by $\hat{\alpha}^C : [0,C] \cap \mathbb{Z}_{\geq} \to \mathbb{R}_{\geq}$, is given by $\hat{\alpha}^C(x) = k_1 + k_2 x$, i.e. it has the same form as the total propensity function of the original network (1), but is restricted to the bounded discrete domain $[0,C] \cap \mathbb{Z}_{\geq}$. With the restriction imposed, we furthermore embed appropriate reactions to the conservative network, so that an arbitrary nonnegative function, denoted by $g : [0,C] \cap \mathbb{Z}_{\geq} \to \mathbb{R}_{\geq}$, is added to $\hat{\alpha}^C$, i.e. the resulting total propensity function is given by $\alpha(x) = \hat{\alpha}^C(x) + g(x)$. This implies that the interevent time is controllably decreased for any desired state x, i.e. in a state-dependent manner. Equivalently, the two requirements imply that a controllable state-dependent noise is introduced into the stochastic dynamics. We have designed a three-step algorithm, given as Algorithm 1, which achieves such goals for arbitrary reaction networks under mass-action kinetics. Let us describe properties of the algorithm by applying it on network (1).

Firstly, in order to bound the domain of species s, an additional species $\bar{s}$ is introduced into network (1), in such a way that s and $\bar{s}$ satisfy a pairwise stoichiometric conservation law, formally written $s + \bar{s} = \text{constant}$. Secondly, to ensure the obtained enlarged network has the same deterministic model as the initial network (1), despite the added species $\bar{s}$, an auxiliary species I^1 is introduced. More precisely, applying the first two steps of the algorithm leads to network $\hat{\mathscr{R}}^1(s,\bar{s}) \cup \mathscr{R}^2_1(\bar{s})$ given by:

$$\begin{aligned} \hat{\mathscr{R}}^1(s,\bar{s}): \quad & \bar{s} + I^1 \xrightarrow{k_1} s + I^1, \\ & s \xrightarrow{k_2} \bar{s}, \\ \mathscr{R}^2_1(\bar{s}): \quad & \varnothing \xrightarrow{1/\mu} I^1, \\ & \bar{s} + I^1 \xrightarrow{1/\mu} \bar{s}. \end{aligned} \quad (3)$$

Species $s, \bar{s}, I^1$ from (3) react according to the four reactions with rate coefficients $k_1, k_2, 1/\mu \in \mathbb{R}_{\geq}$. Reaction network $\hat{\mathscr{R}}^1 = \hat{\mathscr{R}}^1(s,\bar{s})$, given in (3), is obtained from network $\hat{\mathscr{R}} = \hat{\mathscr{R}}(s)$, given by (1), in the following way: since the first reaction in $\hat{\mathscr{R}}$ *increases* the copy-number of s by one, $\bar{s}$ and I^1 are added to the reactants of the reaction, and I^1 is added to the products, leading to the first reaction in $\hat{\mathscr{R}}^1$. Since the second reaction in $\hat{\mathscr{R}}$ *decreases* the copy-number of s by one, $\bar{s}$ is added to the products, leading to the second reaction in $\hat{\mathscr{R}}^1$. This ensures that the desired conservation law, $s + \bar{s} = \text{constant}$, holds. The superscript in I^1 indicates that species I^1 is involved as a catalyst in a reaction of $\hat{\mathscr{R}}^1$ in which s is *increased by one*. The subscript in $\mathscr{R}^2_1 = \mathscr{R}^2_1(\bar{s})$ indicates that the network describes production and decay of I^1.

The initial value problem for the deterministic model of (3) is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= k_1(c-x)y - k_2 x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{1}{\mu}\left(1 - (c-x)y\right), \\ x(0) &= x_0, \\ y(0) &= y_0, \end{aligned} \quad (4)$$

where $x = x(t) \in [0,c] \cap \mathbb{R}_{\geq}$, and $y = y(t) \in \mathbb{R}_{\geq}$, are the concentrations of species s, and I^1, from (3), respectively, with $x_0, y_0 \in \mathbb{R}_{\geq}$ and $c \in \mathbb{R}_{>}$. We have used the kinetic conservation law $\bar{x}(t) = c - x(t)$, where $\bar{x}(t)$ is the concentration of species $\bar{s}$, and c is a finite time-independent conservation constant. Note that the conservation law truncates the state-space of x. Let us now describe relationships between systems (2) and (4), starting with the weak statement: for $c > k_1/k_2$, and for any $\mu > 0$, solutions of (2) and (4) are the same in the long-time limit $t \to \infty$. More precisely, the x-component of the equilibrium point of (4) is identical to the equilibrium

point of (2), and both are stable. In supplementary material, we justify the strong statement: for sufficiently large c, and for $0 < \mu \ll 1$, solutions of (2) and (4), with the same initial conditions, are approximately the same at each time $t \geq 0$. For these reasons, we call $\mathscr{R}_1^2$ a *drift-corrector network*. Let us note that we have assumed the rate coefficients appearing in subnetwork $\mathscr{R}_1^2(\bar{s})$ from (3) are identical for simplicity, and that this assumption may be relaxed. More precisely, if the rate coefficients of the first and second reactions in $\mathscr{R}_1^2(\bar{s})$ are $1/\mu_1$ and $1/\mu_2$, respectively, then the same conclusion from this paragraph holds, provided the rate coefficient k_1 from subnetwork $\hat{\mathscr{R}}^1(s,\bar{s})$ is replaced by $(\mu_1/\mu_2)k_1$.

2.1 Zero-drift Network $\mathscr{R}_{1,1}^3$

Having completed the first two steps, let us focus on the third (and final) step, in which we introduce arbitrary noise into the stochastic model of (3), without influencing the deterministic model (4). Let us start our consideration by embedding into (3) network $\mathscr{R}_{1,1}^3 = \mathscr{R}_{1,1}^3(s,\bar{s})$, which is given by

$$\mathscr{R}_{1,1}^3(s,\bar{s}): \qquad \begin{aligned} s+\bar{s} &\xrightarrow{k_{1,1}} 2s, \\ s+\bar{s} &\xrightarrow{k_{1,1}} 2\bar{s}. \end{aligned} \tag{5}$$

The subscript in $\mathscr{R}_{1,1}^3$ indicates that the underlying reactions have one molecule of s, and one of $\bar{s}$, as reactants. The two reactions in (5) preserve the conservation law from (3). Furthermore, the first and second reactions produce, and degrade, exactly one molecule of s, respectively, and they fire at the same rate. Consequently, embedding $\mathscr{R}_{1,1}^3$ into (3) does not affect the underlying deterministic model (4), and we call $\mathscr{R}_{1,1}^3$ a *zero-drift network*. Note that the deterministic dynamics are not preserved if the rate coefficients in (5) are different. However, $\mathscr{R}_{1,1}^3$ does affect the underlying stochastic model [45, 46, 47, 22]. To illustrate this, let us consider network $\mathscr{R}_{1,1}^3$ in isolation: the reactions from (5) fire when $X(t) \in (0,C)$, but not when $X(t) \in \{0,C\}$, so that $\mathscr{R}_{1,1}^3$ in isolation fires until $X(t)$ takes one of the extreme values $\{0,C\}$. Here, $X(t) \in \mathbb{Z}_{\geq}$, and $C \in \mathbb{Z}_{>}$, are the copy-number of species s appearing in (3) and (5) at time $t \geq 0$, and the finite conservation constant, respectively. Note that a possible biologically-relevant realization of network (5), aside from e.g. DNA strand-displacement mechanism, is a dimer version of the bifunctional histidine kinase/phosphatase reported in [48].

In supplementary material, we derive equation (S11) which describes the effective behavior of the Markov chain $X(t)$ from network $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$ in the limit $\mu \to 0$, and it follows that the effective total propensity function of the network, denoted $\alpha(x)$, satisfies

$$\alpha(x) \approx \hat{\alpha}^C(x) + 2K_{1,1}\beta_{1,1}(x), \quad \text{as } \mu \to 0, \tag{6}$$

$$\hat{\alpha}^C(x) = k_1 + k_2 x. \tag{7}$$

Function $\hat{\alpha}^C : [0,C] \cap \mathbb{Z}_{\geq} \to \mathbb{R}_{\geq}$ has the form of the total propensity of network (1), and $K_{1,1}\beta_{1,1}(x)$ is the propensity function of reactions in (5), with the scaled factors given by

$$K_{1,1} = \left(\frac{C}{2}\right)^2 k_{1,1}, \quad \beta_{1,1}(x) = \left(\frac{C}{2}\right)^{-2} x(C-x). \tag{8}$$

Function $\beta_{1,1}(x)$ is displayed in Figure 1(a), where one can notice its parabolic shape, arising from the underlying conservation law $X(t) + \bar{X}(t) = C$, which holds for all $t \geq 0$, where $\bar{X}(t) \in \mathbb{Z}_{\geq}$ is the copy-number of $\bar{s}$ at time $t \geq 0$. Comparing (6) and (7), it follows that, as $\mu \to 0$, the mean interevent time for network $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$ is lower than for network (1), in the regions of the common state-space where $\beta_{1,1}(x) \neq 0$, i.e. for $x \in (0,C)$. Coefficient $K_{1,1}$ controls by how much the interevent time is reduced. Equivalently, $\beta_{1,1}(x)$, and $K_{1,1}$, determine the support, and magnitude, respectively, of the state-dependent intrinsic noise which network (5) introduces into the dynamics of network (3).

In supplementary material, we rigorously formulate the following two approximate results (given as equations (S13), and (S17), respectively)

$$\lim_{K_{1,1}\to 0} p(x) \approx \begin{cases} \frac{1}{x!}\left(\frac{k_1}{k_2}\right)^x \exp\left(-\frac{k_1}{k_2}\right), & \text{if } x \in [0,C], \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

$$\lim_{K_{1,1}\to \infty} p(x) \approx \begin{cases} 1 - \frac{1}{C}\frac{k_1}{k_2}, & \text{if } x = 0, \\ \frac{1}{C}\frac{k_1}{k_2}, & \text{if } x = C, \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

where $p(x)$ is the stationary probability mass function (PMF) corresponding to network $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$ in the limit $\mu \to 0$, i.e. the probability that there are x molecules of species s as $\mu \to 0$ in the long-time limit $t \to \infty$. Let us interpret analytical results (9) and (10), and compare them with the numerically obtained counterparts. In Figure 1(b), we display numerically obtained stationary x-marginal PMFs for different values of $K_{1,1}$, with the rest of the (dimensionless) parameters fixed to $k_1 = 2.5$, $k_2 = 0.5$, $\mu = 10^{-3}$, and $C = 15$. It can be seen that, for $K_{1,1} = 0$, i.e. when the zero-drift network $\mathscr{R}_{1,1}^3$ does not fire, the PMF matches that of network (1), i.e. it is a Poissonian, as predicted by (9). Let us note that the matching of the PMFs of networks (1) and $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$ relies on choosing sufficiently large rate coefficients $1/\mu$ in the drift-corrector network $\mathscr{R}_1^2$. When $K_{1,1} = 5$, the PMF appears closer to a uniform distribution, than does the PMF when $K_{1,1} = 0$. Finally, for the larger value $K_{1,1} = 10^5$, i.e. when zero-drift network $\mathscr{R}_{1,1}^3$ fires much faster than network $\hat{\mathscr{R}}^1$, the PMF redistributes across the domain, accumulating at the boundary, and becoming bimodal. This is in qualitative agreement with (6), and in quantitative agreement with (10), which predicts $p(0) \approx 0.7$ and $p(15) \approx 0.3$. In Figure 1(c), a representative sample path is shown, obtained by applying the Gillespie algorithm on network $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$, when $K_{1,1} = 10^5$. Also shown is a trajectory obtained by numerically solving the deterministic model (4). Consistent with Figure 1(b), the sample path

switches between the boundary of the state-space, with a bias towards the left boundary point $x = 0$. This is in contrast to the deterministic trajectories, which are globally attracted to the equilibrium point $x = 5$.

2.2 General Zero-drift Networks $\mathscr{R}^3_{n,\bar{n}}$

The zero-drift network $\mathscr{R}^3_{1,1}(s,\bar{s})$, given by (5), involves a single molecule of s and $\bar{s}$ as reactants, and adds the noise at $x \in [1, C-1]$, i.e. in the interior of the state-space. Similar networks may be used to add the noise at any point in the state-space, without influencing the deterministic dynamics. In particular, in (15) and (16), we present general zero-drift networks $\mathscr{R}^3_{n,\bar{n}}(s,\bar{s})$, which involve n molecules of s, and $\bar{n}$ of $\bar{s}$, as reactants, and add the noise at $x \in [n, C-\bar{n}]$, where $n, \bar{n} \in \mathbb{N}_0$, and $(n+\bar{n}) \leq C$. Embedding a union of such networks, $\cup_{(n,\bar{n})}\mathscr{R}^3_{n,\bar{n}}(s,\bar{s})$, into (3), we arrive at the result similar to (6), with $K_{1,1}\beta_{1,1}(x)$ replaced by the linear combination $\sum_{(n,\bar{n})} K_{n,\bar{n}}\beta_{n,\bar{n}}(x)$. The scaled rate coefficient $K_{n,\bar{n}}$, and function $\beta_{n,\bar{n}}(x)$, are given in supplementary material as equations (S18), and (S19), respectively, where we also justify that an arbitrary nonnegative function, defined on a bounded discrete domain, may be represented by a suitable sum $\sum_{(n,\bar{n})} K_{n,\bar{n}}\beta_{n,\bar{n}}(x)$.

To illustrate general zero-drift networks, let us start with embedding into network (3), with the conservation constant $C = 15$, the zero-drift network $\mathscr{R}^3_{5,10}(s,\bar{s})$, satisfying (15) with $n = 5$ and $\bar{n} = 10$. In Figure 1(d), we show propensity function $\beta_{5,10}(x)$, which is nonzero only at $x = 5$. In (e), we show the numerically approximated stationary x-marginal PMFs underlying network $\hat{\mathscr{R}}^1 \cup \mathscr{R}^2_1 \cup \mathscr{R}^3_{5,10}$ for different values of $K_{5,10}$, with the rest of the parameters as in Figure 1(b). One can notice that, under the action of network $\mathscr{R}^3_{5,10}$, the PMF is gradually decreased to nearly zero at $x = 5$ (the deterministic equilibrium), and becomes bimodal, with the two noise-induced maxima at $x = 4$ and $x = 6$. In (f), we show a corresponding representative sample path.

In general, noise-induced multimodality may be achieved by a suitable combination of zero-drift networks. For example, let us synthetize noise such that the stationary PMF is trimodal, and nearly zero everywhere, except at $x \in \{1, 7, 11\}$. Such a task may always be achieved by a suitable combination of the basis zero-drift networks, i.e. those zero-networks that induce noise only at a single point in the state-space (e.g. subnetwork $\mathscr{R}^3_{5,10}$ with propensity function shown in Figure 1(d), see also supplementary material). In the present case, one could construct the thirteen basis zero-drift networks which add large enough noise at $x \in [0, 15] \setminus \{1, 7, 11\}$. Here, for simplicity, we achieve the task with only four zero-drift networks. In Figures 1(g)–(i), we consider network $\hat{\mathscr{R}}^1 \cup \mathscr{R}^2_1 \cup (\mathscr{R}^3_{0,15} \cup \mathscr{R}^3_{2,9} \cup \mathscr{R}^3_{8,5} \cup \mathscr{R}^3_{12,0})$. We denote $\beta(x) \equiv \beta_{0,15}(x) + \beta_{2,9}(x) + \beta_{8,5}(x) + \beta_{12,0}(x)$, and, for simplicity, take $K \equiv K_{0,15} = K_{2,9} = K_{8,5} = K_{12,0}$. The resultant propensity function $\beta(x)$ is shown in (g), while in (h) it can be seen that the PMF becomes trimodal for sufficiently large K, with the maxima at $x = \{1, 7, 11\}$. This is consistent with the corresponding representative sample path shown in blue in panel (i), which display tristability. Let us note that, while the stochastic dynamics display multistability in (c), (f) and (i), the corresponding deterministic dynamics, also shown in the plots, remain monostable.

2.3 Compilation Into DNA-based Networks

Chemical reaction networks, whose stochastic dynamics are controlled by Algorithm 1, may be mapped to the nucleic-acid-based ones. The mapping takes a different form depending on which molecular compiler is utilized and, in this section, we briefly outline two approaches. Firstly, the molecular compiler put forward in [23], based on 4-domain signal strands, requires that the input reaction network consists of up-to second-order reactions. On the other hand, let us note that it allows reactions with identical reactants (as is the case in zero-drift networks). Thus, one is generally required to apply a single pre-compiling step, where the higher-order reactions (i.e. reactions involving three or more reactants) are approximated by systems of up-to second-order ones [49, 50], before using the 4-domain DNA compiled. However, the 4-domain compiler has only been shown to preserve the deterministic dynamics when mapping an abstract network into a DNA-based one [23]. In supplementary material, we show that the stochastic dynamics are also preserved, making the compiler compatible with the noise-control algorithm. Furthermore, we apply the compiler to a network of the form (3) and (5), and briefly discuss the pre-compilation step, leaving the details for a future publication [50]. On the other hand, the 2-domain molecular compiler put forward in [51], and experimentally implemented in [52], can be used directly, without any pre-compilation, since it automatically handles higher-order reactions.

3. A Two-species Exotic System

Consider the two-species network $\tilde{\mathscr{R}}(s_1, s_2)$, given by

$$
\begin{aligned}
\tilde{\mathscr{R}}(s_1,s_2):\quad
&\varnothing \xrightarrow{k_1} s_1, && \varnothing \xrightarrow{k_7} s_2,\\
&s_1 \xrightarrow{k_2} 2s_1, && s_2 \xrightarrow{k_8} \varnothing,\\
&2s_1 \xrightarrow{k_3} 3s_1, && s_1+s_2 \xrightarrow{k_9} s_1+2s_2,\\
&s_1+s_2 \xrightarrow{k_4} s_2, && 2s_2 \xrightarrow{k_{10}} 3s_2,\\
&2s_1+s_2 \xrightarrow{k_5} s_1+s_2, && 3s_2 \xrightarrow{k_{11}} 2s_2,\\
&s_1+2s_2 \xrightarrow{k_6} 2s_1+2s_2, &&
\end{aligned}
\tag{11}
$$

where species s_1 and s_2 react according to the eleven reactions with rate coefficients $k_1, k_2, \ldots, k_{11} \geq 0$. We denote the copy-numbers of species s_1, and s_2, at time t by $X_1(t)$, and $X_2(t)$, respectively. It was established in [20] that, for particular choices of the rate coefficients, the deterministic model of reaction network (11), given in as equation (S34) in supplementary material, exhibits exotic dynamics: it undergoes a homoclinic bifurcation, and displays a bistability involving

a limit cycle and an equilibrium point. On the other hand, it is demonstrated in [21] that the stochastic model of (11) is not necessarily sensitive to the deterministic bifurcation, and may effectively behave in a monostable manner. The latter point is demonstrated in Figure 2(c), where we show in red numerically approximated x_1-solutions of equation (S34) from supplementary material, one initiated in the region of attraction of the equilibrium point, while the other of the limit cycle. For a comparison, we also show in blue a representative sample path generated by applying the Gillespie algorithm on (11). It can be seen that the stochastic solution spends significantly more time near the deterministic equilibrium point. To gain a clearer picture, we display in Figures 2(a), and (b), the joint, and the x_1-marginal, stationary PMFs, respectively, underlying network (11), which have been obtained numerically for the same parameter values as in Figure 2(c). In (b), one can notice that the PMF is bimodal, but the left peak, corresponding to the limit cycle, is significantly smaller than the right peak, which corresponds to the stable equilibrium point.

We now apply Algorithm 1 on network (11) to achieve two goals. Firstly, we balance the sizes of the two peaks of the stationary PMF from Figure 2(b), thereby forcing the stochastic system to spend comparable amounts of time at the two deterministic attractors. Secondly, we reverse the situation shown in Figure 2(b), by making the left PMF peak significantly larger than the right one, thereby forcing the stochastic system to spend most of the time near the limit cycle. We could achieve the goals by introducing species $\bar{s}_1, \bar{s}_2$ into (11), and using suitable basis zero-drift networks. We take a simpler approach, by mapping (11) to $\tilde{\mathscr{R}}^1(s_1,s_2,\bar{s}_2) \cup \mathscr{R}^2_1(\bar{s}_2) \cup (\mathscr{R}^3_{0,C_2-10}(s_2,\bar{s}_2) \cup \mathscr{R}^3_{30,0}(s_2,\bar{s}_2))$, which is given as equation (S35) in supplementary material. For our purposes, only one of $\bar{s}_1$, $\bar{s}_2$ is sufficient, since the stochastic dynamics of s_1 and s_2 are coupled. We have chosen $\bar{s}_2$ for convenience, since x_2-state-space may be truncated at a lower value, $C_2 = 180$, than x_1-state-space (see also Figure 2 (a)). The x_2-component of the deterministic limit cycle satisfies $x_2 \in (10,30)$. Correspondingly, we introduce two zero-drift networks: $\mathscr{R}^3_{0,C_2-10}(s_2,\bar{s}_2)$, and $\mathscr{R}^3_{30,0}(s_2,\bar{s}_2)$, which redistribute the PMF from $x_2 \in [0,10]$, and from $x_2 \in [30,C_2]$, respectively, to the limit cycle region, $x_2 \in (10,30)$. We fix the scaled rate coefficient K^2_{0,C_2-10} to a large value (so that the PMF is nearly zero for $x_2 \in [0,10]$), and vary the coefficient $K^2_{30,0}$, which redistributes the PMF from the deterministic equilibrium point to the limit cycle. Network $\mathscr{R}^2_1(\bar{s}_2)$ is necessary for the preservation of the deterministic dynamics of (11) under the application of Algorithm 1.

In Figures 2(d), and (e), we show the joint, and x_1-marginal, stationary PMFs for an intermediate value of $K^2_{30,0}$, when the PMF is partially redistributed from $x_2 \in [30,C_2]$ to $x_2 \in (10,30)$, so that the two peaks in (e) are of comparable sizes. In Figure 2(f), we show a representative sample path, obtained by applying the Gillespie algorithm on network (S35), together with the deterministic trajectories obtained by solving (S34). One can notice that the stochastic system now spends significantly more time near the limit cycle, when compared to (c). Let us note that stochastic switching between a coexisting equilibrium point and a limit cycle in a DNA-based reaction network has been observed experimentally [26]. In Figure 2(f)–(g), we show analogous plots, but for a sufficiently large value of $K^2_{30,0}$, when the PMF is almost completely redistributed from $x_2 \in [30,C_2]$ to $x_2 \in (10,30)$. Now, in contrast to Figure 2(a)–(c), the PMF becomes unimodal, and concentrated around the limit cycle. Let us note that the red trajectories from Figures 2(f) and (i) were generated by numerically solving the deterministic model (S34) from supplementary material. For our purposes, it is not necessary to solve the corresponding (stiff) deterministic model of network (S35) from supplementary material. The reason is that Algorithm 1 does not influence the deterministic equilibrium points of a given reaction network, regardless of the choice of the kinetic algorithm parameters. Thus, while the deterministic limit cycle is not necessarily preserved for the algorithm parameters chosen in Figure 2(i), the enclosed deterministic unstable focus is necessarily preserved. Consequently, the blue sample path shown corresponds to noise-induced oscillations either near a deterministic limit cycle, or near a deterministic unstable focus.

4. Discussion

In this paper, we have presented the noise-control algorithm, which is given as Algorithm 1. The algorithm maps an input chemical reaction network to output networks, all under mass-action kinetics, by introducing appropriate additional species and reactions, such that the output networks satisfy the following two properties. Firstly, the output networks have the same deterministic model as the input network, in appropriate limits of some of the parameters (rate coefficients) introduced by the algorithm. Secondly, controllable state-dependent noise is introduced into the stochastic model of the output networks. Thus, Algorithm 1 may be used to control the intrinsic noise of a given reaction network under mass-action kinetics, while preserving the deterministic dynamics. Let us note that the asymptotic conditions for the algorithm parameters are necessary for preservation of the time-dependent deterministic solutions. However, the time-independent deterministic solutions (the deterministic equilibrium points), which capture important features of the deterministic dynamics, are preserved under the algorithm even if the asymptotic conditions are not satisfied.

The algorithm has been applied to a test problem, taking the form of the one-species production-decay system given by (1). Using analytical and numerical methods, we have shown that the additional intrinsic noise, introduced by the algorithm, may be used to favorably modify the stationary probability mass function at arbitrary points in the state-space, as demonstrated in Figure 1. For example, in Figure 1(b), the noise is added to the whole interior of the state-space, while in (e) only at a single point, in both cases resulting in

noise-induce bimodality. On the other hand, in Figure 1(h), by adding the noise to specific points in the state-space, the network is redesigned to display noise-induced trimodality. As shown in Figures 1(c), (f), (i), the blue stochastic trajectories display multistability, while the red deterministic ones remain monostable.

The algorithm has also been applied to a more challenging problem, taking the form of the two-species system given by (11), which, for the parameters taken in this paper, at the deterministic level displays a bistability involving an equilibrium point and a limit cycle [20, 21]. At the stochastic level, the system is significantly more likely to be found near the equilibrium point, as demonstrated in Figures 2(a)–(c). We have used the algorithm to redesign network (11), so that the stochastic system spends comparable amounts of time near the two attractors, as demonstrated in Figures 2(d)–(f). The network was also redesigned to display noise-induced oscillations, which is shown in Figures 2(g)–(i). Put another way, one may view the dynamics shown in Figures 2(a)–(c) as being contaminated by the noise, which disrupts the oscillatory chemical computation. Algorithm 1 has been applied to address the disruption by appropriately reprogramming the noise. Such control is of practical relevance, since stochastic switching between an equilibrium and a limit cycle has already been observed experimentally in a DNA-based network [26].

The controllable state-dependent noise is generated by Algorithm 1 using the zero-drift networks (15) and (16). Any nonnegative function, defined on a bounded discrete domain, may be represented by a linear combination of propensity functions induced by an appropriate union of the zero-drift networks. Thus, choosing suitable zero-drift networks, the algorithm may control the intrinsic noise at arbitrary points in the state-space of the stochastic dynamics of reaction networks. The cost of such a precision in noise-control is a larger number of reactants in the underlying zero-drift networks. However, while the high-molecular reactions introduced by the algorithm are more expensive to synthetize, they do not limit applicability of Algorithm 1 to DNA computing. The reason for this is that such reactions may always be broken down into sets of up-to bimolecular reactions, with asymptotically equivalent dynamics [49, 50], as outlined in Section 2.3 and exemplified in supplementary material. Let us stress that the lower the molecular copy-numbers of a given reaction network are, the more important it becomes to control their stochastic behavior, and, fortunately, the *less* costly Algorithm 1 becomes (since the zero-drift networks involve less reactants).

Algorithm 1 constitutes a qualitatively novel scientific discovery which will facilitate the progress of nucleic-acid-based computing, such as DNA computing [23, 24, 25, 26]. In particular, we put forward a hybrid approach for constructing DNA-based reaction networks: the deterministic model may be used to guide the construction of reaction networks, and then Algorithm 1 may be applied to favorably reprogram the intrinsic noise in the stochastic model, while preserving the mean-field behavior. Put another way, the deterministic-stochastic hybrid approach allows one to reshape the probability distributions of target chemical species, while inheriting the fixed mean-field behavior. This provides a control over, not only the probability distributions of the chemical species, but also over their sample-paths, as e.g. demonstrated in Section 3 with a noisy limit cycle. Furthermore, the algorithm does not depend on the initial conditions of the underlying species, beyond the conservation laws. This is in contrast to the methods presented in [53, 54], which do not attempt to control the sample-path behavior, and whose performance depends strongly on the initial conditions of the underlying species, which may impose significant experimental challenges. The noise-control algorithm may be of critical importance when the synthetic networks involve species at low copy-numbers, since then the stochastic effects may play a significant role [24, 25, 26, 33, 34, 13, 18, 21, 30, 31, 32]. On the one hand, the algorithm may enhance our understanding of biology, via theoretical and experimental investigations of the role of intrinsic noise in both prebiotic and biotic chemical processes [24, 25, 26]. On the other hand, the algorithm may facilitate *in vivo* implementations of synthetic DNA-networks, where the reactions may take place at a cellular level. In such circumstances, without a control, the noise may contaminate the performance of the synthetic networks. Algorithm 1 provides a way to control the stochastic effects, enriching the DNA-based synthetic systems with novel, noise-induced functionalities. For example, one may envisage using the algorithm to design nucleic-acid-based circuits interacting favorably with gene-regulatory networks, where noise-induced multimodality is known to play a critical role [55, 56]. On the one hand, the algorithm could be used to induce multimodality in the probability distribution of an appropriate intracellular protein, resulting in cell phenotype diversity. On the other hand, the algorithm could also be utilized to make the protein distribution narrower around the single peak, thus inducing a cell phenotype robustness.

Author Contributions

All the authors have contributed to all the aspects of the paper.

Competing Interests

The authors declare that they have no competing interests.

Data Accessibility

Data supporting this paper are contained within the paper, and as the electronic supplementary material.

Acknowledgments

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme “Stochastic Dynamical Systems

in Biology: Numerical Methods and Applications", where work on this paper was undertaken. The authors would also like to thank John J. Tyson (Department of Biology, Virginia Polytechnic Institute and State University, United States) for a discussion on a possible realization of network (5) via a bifunctional histidine kinase/phosphatase from [48], and to Ibon Santiago (Department of Physics, University of Oxford, UK) for providing useful discussions and references about DNA nanotechnology.

Funding Statement

This work was supported by EPSRC grant no EP/K032208/1. This work was partially supported by a grant from the Simons Foundation. Konstantinos C. Zygalakis was supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1. David F. Anderson would like to acknowledge the NSF grant NSF-DMS-1318832, and Army Research Office grant W911NF-14-1-0401. Radek Erban would also like to thank the Royal Society for a University Research Fellowship.

References

[1] Endy D. Foundations for engineering biology. Nature, 2005;484:449–453.

[2] Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology, 2006;2:2006.0028.

[3] Abil Z, Xiong X, Zhao H. Synthetic biology for therapeutic applications. Molecular Pharmaceutics, 2015;12(2):322–331.

[4] Anderson JC, Clarke EJ, Arkin AP, Voigt CA. Environmentally controlled invasion of cancer cells by engineered bacteria. Journal of Molecular Biology, 2006;355(4):619–627.

[5] Benenson Y, Gil B, Ben-Dor U, Adar R, Shapiro E. Autonomous molecular computer for logical control of gene expression. Nature, 2004;429:423–429.

[6] Ro D, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature, 2006;440:940–943.

[7] Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, et al. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc. Natl Acad. Sci. USA, 2012;109:E111–E118.

[8] Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, et al. Engineering the salmonella type III secretion system to export spider silk monomers. Molecular Systems Biology, 2009;5(309).

[9] Sedlak M, Ho WY. Production of ethanol from cellulosic biomass hydrolysates using genetically engineered saccharomyces yeast capable of cofermenting glucose and xylose. Applied Biochemistry and Biotechnology, 2004;114(1):403–416.

[10] Ball P. Synthetic biology for nanotechnology. Nanotechnology, 2005;16:R1–R8.

[11] Jungmann R, Renner S, Simmel FC. From DNA nanotechnology to synthetic biology. Applied Biochemistry and Biotechnology, 2008;2(2):99–109.

[12] Menezes AA, Cumbers J, Hogan JA, Arkin AP. Towards synthetic biological approaches to resource utilization on space missions. Journal of The Royal Society Interface, 2015;12(102):20140715.

[13] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature, 2000;403:335–338.

[14] Deamer D. A giant step towards artificial life? Trends in Biotechnology, 2005;23(7):336–338.

[15] Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, et al. Essential genes of a minimal bacterium. Proceedings of the National Academy of Sciences, 2006;103(2):425–430.

[16] Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science's STKE, 2008;319(5867):1215–1220.

[17] Gibson DG. Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010;329(5987):52–56.

[18] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature, 2000;403:339–342.

[19] You L, Cox Ill RS, Weiss R, Arnold FH. Programmed population control by cell-cell communication and regulated killing. Nature, 2004;428:868–871.

[20] Plesa T, Vejchodský T, and Erban R. Chemical reaction systems with a homoclinic bifurcation: an inverse problem. Journal of Mathematical Chemistry, 2016;54(10):1884–1915.

[21] Plesa T, Vejchodský T, and Erban R. Test models for statistical inference: two-dimensional reaction systems displaying limit cycle bifurcations and bistability. Stochastic Dynamical Systems, Multiscale Modeling, Asymptotics and Numerical Methods for Computational Cellular Biology, 2017, available as https://arxiv.org/abs/1607.07738.

[22] Érdi P, Tóth J. Mathematical models of chemical reactions. Theory and applications of deterministic and stochastic models. Manchester University Press, Princeton University Press; 1989.

[23] Soloveichik D, Seeling G, Winfree E. DNA as a universal substrate for chemical kinetics. Proceedings of the National Academy of Sciences, 2010;107(12):5393–5398.

[24] Hasatani K, Leocmach M, Genot A J, Estévez-Torres A, Fujii T, Rondelez Y. High-throughput and long-term observation of compartmentalized biochemical oscillators. ChemComm, 2013;49:8090–8092.

[25] Weitz M, Kim J, Kapsner K, Winfree E, Franco E, Simmel F C. Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator. Nature Chemistry, 2014;6:295–302.

[26] Genot A J, Baccouche A, Sieskind R, Aubert-Kato N, Bredeche N, Bartolo J F, et al. High-resolution mapping of bifurcations in nonlinear biochemical circuits. Nature Chemistry, 2016;10.1038/nchem.2544.

[27] Anderson DF, Kurtz TG. Stochastic analysis of biochemical systems. Springer; 2015.

[28] Gillespie D. Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry, 1977;81(25):2340–2361.

[29] Van Kampen NG. Stochastic processes in physics and chemistry. Elsevier; 2007.

[30] Erban R, Chapman SJ, Kevrekidis I, Vejchodský T. Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model. SIAM Journal on Applied Mathematics, 2009;70(3):984–1016.

[31] Duncan A, Liao S, Vejchodský T, Erban R, Grima R. Noise-induced multistability in chemical systems: discrete vs continuum modelling. Physical Review E, 2015;91:042111.

[32] Yates C, Erban R, Escudero C, Couzin I, Buhl J, Kevrekidis I, et al. Inherent noise can facilitate coherence in collective swarm motion. Proceedings of the National Academy of Sciences, 2009;106(14):5464–5469.

[33] Vilar JMG, Kueh HY, Barkai N, Leibler S. Mechanisms of noise-resistance in genetic oscillators. Proceedings of the National Academy of Sciences of the United States of America, 2002;99(9):5988–5992.

[34] Dublanche Y, Michalodimitrakis K, Kummerer N, Foglierini M, Serrano L. Noise in transcription negative feedback loops: simulation and experimental analysis. Molecular Systems Biology, 2006;2(41):E1–E12.

[35] Kurtz TG. The relationship between stochastic and deterministic models for chemical reactions. Journal of Chemical Physics, 1972;57:2976–2978.

[36] Srinivas N, Parkin J, Seeling G, Winfree E, Soloveichik D. Enzyme-free nucleic acid dynamical systems. Science, 2017;358, eaal2052.

[37] Fujii T, Rondelez Y. Predator-prey molecular ecosystems. ACS Nano, 2013;7:27–34.

[38] Šulc P, Ouldridge TE, Romano F, Doye JPK, Louis AA. Modelling toehold-mediated RNA strand displacement. Biophysical Journal, 2015;108: 1238–1247.

[39] Homann M, Nedbal W, and Sczakiel G. Dissociation of long-chain duplex RNA can occur via strand displacement in vitro: biological implications. Nucleic Acids Research, 1996;24: 4395–4400.

[40] Teichmann M, Kopperger E, Simmel FC. Robustness of localized DNA strand displacement cascades. ACS Nano, 2014;8: 8487–8496.

[41] Chatterjee G, Dalchau N, Muscat RA, Phillips A, Seelig G. A spatially localized architecture for fast and modular DNA computing. Nature nanotechnology, 2017;12: 920–927.

[42] Wickham SFJ, Bath J, Katsuda Y, Endo M, Hidaka K, Sugiyama H, Turberfield AJ. A DNA-based molecular motor that can navigate a network of tracks. Nature Nanotechnology, 2012;7: 169–173.

[43] Thubagere AJ, Li W, Johnson RF, Chen Z, Doroudi S, Lee YL, Izatt G, Wittman S, Srinivas N, Woods D, Winfree E, Qian L. A cargo-sorting DNA robot. Science, 2017;357: eaan6558.

[44] Soloveichik D, Cook M, Winfree E, Bruck J. Computation with finite stochastic chemical reaction networks. Natural Computing, 2008;7(4):615–633.

[45] Ohkubo J, Shnerb N, Kessler DA. Transition phenomena induced by internal noise and quasi-absorbing state. Journal of the Physical Society of Japan, 2008;77:044002.

[46] Biancalani T, Dyson L, McKane AJ. Noise-induced bistable sates and their mean switching time in foraging colonies. Physical Review Letters, 2014;112:038101.

[47] Saito N, Kaneko K. Theoretical analysis of discreteness-induced transition in autocatalytic reaction dynamics. Physical Review Letters, 2015;91:022707.

[48] Subramanian K, Paul MR, Tyson JJ. Potential role of a bistable histidine kinase switch in the asymmetric division cycle of caulobacter crescentus. PLOS Computational Biology, 2013;9:e1003221.

[49] Wilhelm T. Chemical systems consisting only of elementary steps - a paradigma for nonlinear behavior. Journal of Mathematical Chemistry, 2000;27:71–88.

[50] Plesa T. Stochastic approximation of high-molecular by bi-molecular reactions. In preparation; 2018.

[51] Cardelli L. Two-domain DNA strand displacement. Math. Struc. in Comp. Science, 2013;23: 247–271.

[52] Chen YJ, Dalchau N, Srinivas N, Phillips A, Cardelli L, Soloveichik D, Seelig G. Programmable chemical controllers made from DNA. Nature Nanotechnology, 2013;8: 755–762.

[53] Fett B, Bruck J, Riedel MD. Synthesizing stochasticity in biochemical systems. In: 2007 44th ACM/IEEE Design Automation Conference, DAC 2007, pp. 640 − −645. IEEE.

[54] Cardelli L, Kwiatkowska M, Laurenti L. Programming Discrete Distributions with Chemical Reaction Networks. In: Rondelez Y., Woods D. (eds) DNA Computing and Molecular Programming. DNA 2016. Lecture Notes in Computer Science, vol 9818. Springer, Cham.

[55] Fraser D, Kaern M. A chance at survival: gene expression noise and phenotypic diversification strategies. Molecular microbiology, 2009;6: 1333–1340.

[56] Plesa T, Erban R, Othmer HG. Noise-induced Mixing and Multimodality in Reaction Networks. Submitted to European Journal of Applied Mathematics, 2018. Available as https://arxiv.org/abs/1801.09200.

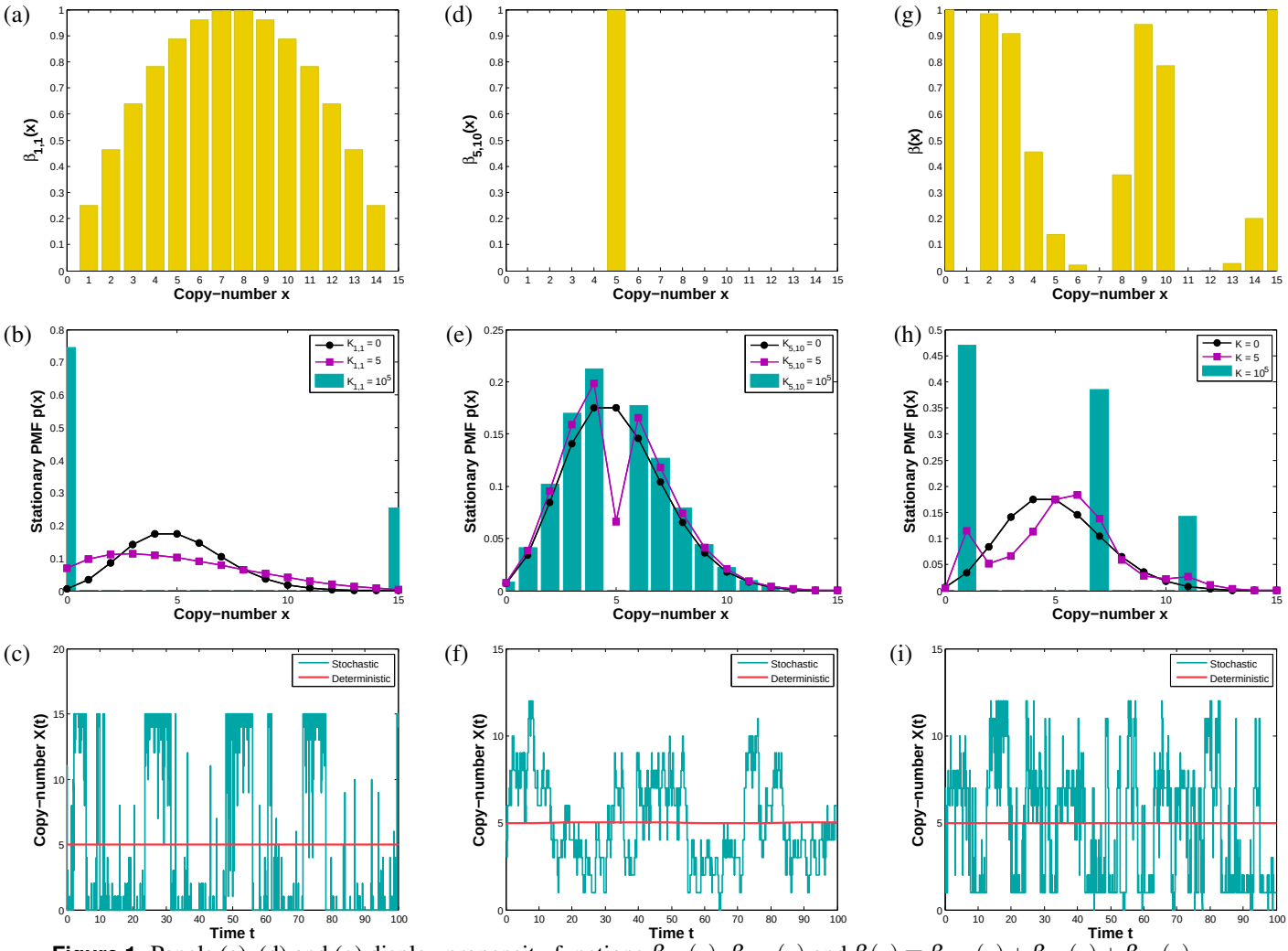

Figure 1. Panels (a), (d) and (g) display propensity functions $\beta_{1,1}(x)$, $\beta_{5,10}(x)$ and $\beta(x) \equiv \beta_{0,15}(x) + \beta_{2,9}(x) + \beta_{8,5}(x) + \beta_{12,0}(x)$, respectively. Panels (b), (e) and (h) display the stationary PMF of networks $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{1,1}^3$, $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup \mathscr{R}_{5,10}^2$ and $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2 \cup (\mathscr{R}_{0,15}^3 \cup \mathscr{R}_{2,9}^3 \cup \mathscr{R}_{8,5}^3 \cup \mathscr{R}_{12,0}^3)$, respectively, where $\hat{\mathscr{R}}^1 \cup \mathscr{R}_1^2$ is given by (3), while the rest of the (zero-drift) networks are as given in second step of Algorithm 1. In (h), $K \equiv K_{0,15} = K_{2,9} = K_{8,5} = K_{12,0}$. Panels (c), (f), and (i) display in blue the sample paths, corresponding to the PMFs shown as the blue histograms in (b), (e) and (h), respectively, and were obtained by applying the Gillespie algorithm on the underlying networks. Also shown in red are the deterministic trajectories, obtained by numerically solving the corresponding deterministic models. The dimensionless parameters are fixed to: $k_1 = 2.5$, $k_2 = 0.5$, $\mu = 10^{-3}$, $C = 15$, and the state-space for species I^1 is bounded in (b), (e) and (h) by 50. In (b) and (e), the two-species stationary chemical master equation (CME) was numerically solved, while in (h) the boundary zero-drift networks are taken in the asymptotic limits $\mu_{0,15}, \mu_{12,0} \to 0$. The blue and red trajectories from panel (i) were generated with $(\mu_{0,15})^{-1} M_{0,15} = (\mu_{12,0})^{-1} M_{12,0} = 10^7$. The trajectories from (c), (f) and (i) were all initiated at the deterministic equilibrium, $X(0) = 5$.

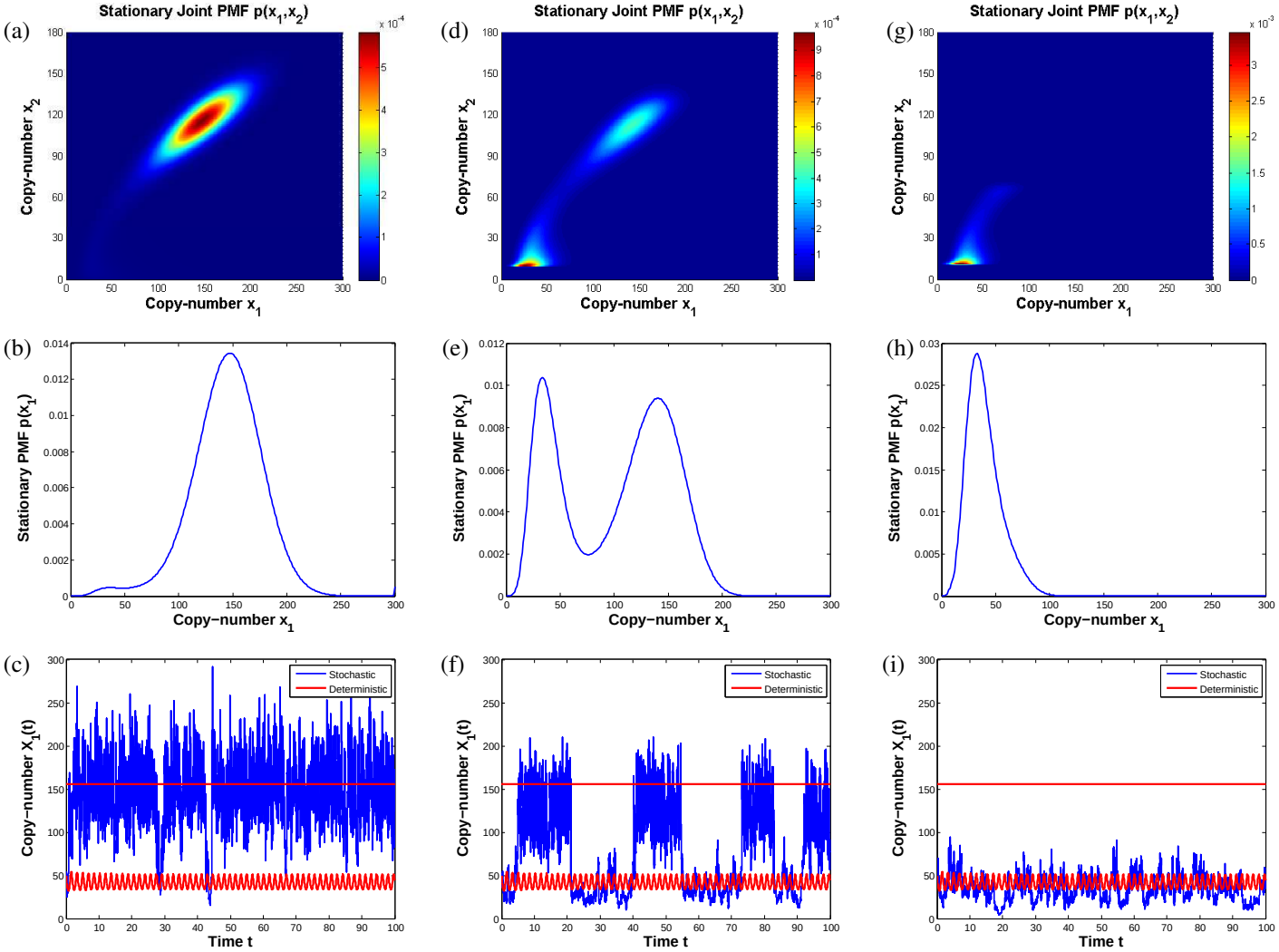

Figure 2. Panel (a) displays the joint stationary PMF of network (11), while (d) and (g) display the stationary PMFs of network (S35) from supplementary material, for $(K_{0,C_2-10}, K_{30,0}) = (10^{18}, 2\times 10^8)$ and $(K_{0,C_2-10}, K_{30,0}) = (10^{18}, 10^{18})$, respectively, with the rest of the parameters being the same. Panels (b), (e) and (h) display the x_1-marginal PMFs corresponding to (a), (b) and (c), respectively. Panels (c), (f) and (i) display in blue the sample paths, corresponding to the PMFs shown in (b), (e) and (h), respectively, and were obtained by applying the Gillespie algorithm on the underlying networks. Also shown in red are two deterministic trajectories, one initiated near the equilibrium point, while the other near the limit cycle, obtained by numerically solving equation (S34) from supplementary material. The dimensionless parameters are fixed to: $k_1 = 4$, $k_2 = 1.408$, $k_3 = 0.0518$, $k_4 = 0.164$, $k_5 = 3.1\times 10^{-3}$, $k_6 = 4.8\times 10^{-3}$, $k_7 = 4$, $k_8 = 8$, $k_9 = 0.16$, $k_{10} = 0.104$, $k_{11} = 2.1\times 10^{-3}$. In (a)–(b), (d)–(e) and (g)–(h), the stationary chemical master equation (CME) is numerically solved, with the state-space truncated to $(x_1, x_2) \in [0, C_1]\times[0, C_2]$, where $C_1 = 300$, $C_2 = 180$, and $\mu, \mu_{0,C_2-10}, \mu_{30,0} \to 0$. The blue sample paths from panels (f) and (i) were generated with $(\mu^{-1}, (\mu_{0,C_2-10})^{-1}M_{0,C_2-10}, (\mu_{30,0})^{-1}M_{30,0}) = (10^3, 10^{20}, 2\times 10^{10})$ and $(\mu^{-1}, (\mu_{0,C_2-10})^{-1}M_{0,C_2-10}, (\mu_{30,0})^{-1}M_{30,0}) = (10^3, 10^{20}, 10^{20})$, respectively. The blue trajectories from (c), (f) and (i) were all initiated near the deterministic limit cycle.

Input: Let the input reaction network be given by

$$\hat{\mathscr{R}}(s_1,\ldots,s_N):\ \sum_{i=1}^{N} c_{ij}s_i \xrightarrow{k_j} \sum_{i=1}^{N} c'_{ij}s_i,\ \ j\in\{1,\ldots,M\}, \tag{12}$$

where $s_1,\ldots,s_N$, are the species, k_j the reaction rate coefficients, and c_{ij}, c'_{ij} the stoichiometric coefficients.

(1) Step: Reaction network $\hat{\mathscr{R}}$, given by (12), is mapped to a *pairwise conservative network* $\hat{\mathscr{R}}^1$ given by

$$\begin{aligned}\hat{\mathscr{R}}^1(s_1,\ldots,s_N,\bar{s}_1,\ldots,\bar{s}_N):\ &\sum_{i=1}^{N}\Big(c_{ij}s_i+(\Delta x_{ij}\bar{s}_i+I_i^{\Delta x_{ij}})\times 1_{\mathbb{N}}(\Delta x_{ij})\Big)\xrightarrow{k_j}\\ &\sum_{i=1}^{N}\Big(c'_{ij}s_i-(\Delta x_{ij}\bar{s}_i)\times 1_{\mathbb{N}}(-\Delta x_{ij})+I_i^{\Delta x_{ij}}\times 1_{\mathbb{N}}(\Delta x_{ij})\Big),\ \ j\in\{1,\ldots,M\}.\end{aligned} \tag{13}$$

Here, $\bar{s}_i, I_i^{\Delta x_{ij}}$ are additional species, $\Delta x_{ij}=(c'_{ij}-c_{ij})$, and $1_{\mathbb{N}}(\cdot)$ is the indicator function of the natural numbers.

(2) Step: For each species $I_i^{\Delta x_{ij}}$, a *drift-corrector network* is constructed, $\mathscr{R}^2_{\Delta x_{ij}}(\bar{s}_i)=\mathscr{R}^2_{\Delta x_{ij}}(\bar{s}_i; I_i^{\Delta x_{ij}},\mu)$, given by

$$\begin{aligned}\mathscr{R}^2_{\Delta x_{ij}}(\bar{s}_i):\quad &\varnothing \xrightarrow{1/\mu} I_i^{\Delta x_{ij}},\\ &\Delta x_{ij}\bar{s}_i + I_i^{\Delta x_{ij}} \xrightarrow{1/\mu} \Delta x_{ij}\bar{s}_i.\end{aligned} \tag{14}$$

where $0\le\mu\ll 1$.

(3) Step: For each species $\bar{s}_i$, a union of *zero-drift networks* may be constructed. Let $n,\bar{n}\in\mathbb{N}_0$, and $(n+\bar{n})\le C_i$. Network $\mathscr{R}^3_{n,\bar{n}}(s_i,\bar{s}_i)=\mathscr{R}^3_{n,\bar{n}}(s_i,\bar{s}_i;k^i_{n,\bar{n}})$, with $n,\bar{n}\neq 0$, is given by

$$\begin{aligned}\mathscr{R}^3_{n,\bar{n}}(s_i,\bar{s}_i):\ & ns_i+\bar{n}\bar{s}_i \xrightarrow{k^i_{n,\bar{n}}} (n+1)s_i+(\bar{n}-1)\bar{s}_i,\\ & ns_i+\bar{n}\bar{s}_i \xrightarrow{k^i_{n,\bar{n}}} (n-1)s_i+(\bar{n}+1)\bar{s}_i.\end{aligned} \tag{15}$$

Network $\mathscr{R}^3_{0,\bar{n}}(s_i,\bar{s}_i)=\mathscr{R}^3_{0,\bar{n}}(s_i,\bar{s}_i;\ B_i,k^i_{0,\bar{n}},\mu_{0,\bar{n}})$, with $\bar{n}\neq 0$, is given by

$$\begin{aligned}\mathscr{R}^3_{0,\bar{n}}(s_i,\bar{s}_i):\ & \bar{n}\bar{s}_i \xrightarrow{k^i_{0,\bar{n}}} s_i+(\bar{n}-1)\bar{s}_i,\\ & C_is_i+B_i \xrightarrow{k^i_{0,\bar{n}}} (C_i-1)s_i+\bar{s}_i+B_i,\\ & \bar{n}\bar{s}_i \xrightarrow{1/\mu_{0,\bar{n}}} \bar{n}\bar{s}_i+B_i,\\ & C_is_i+B_i \xrightarrow{1/\mu_{0,\bar{n}}} C_is_i,\end{aligned} \tag{16}$$

where $0\le\mu_{0,\bar{n}}\ll 1$, and B_i is an additional species. Network $\mathscr{R}^3_{n,0}=\mathscr{R}^3_{0,n}(\bar{s}_i,s_i;\bar{B}_i,k^i_{n,0},\mu_{n,0})$.

Output: An output reaction network $\mathscr{R}$ is given by

$$\mathscr{R}=\hat{\mathscr{R}}^1\cup\mathscr{R}^2\cup\mathscr{R}^3, \tag{17}$$

where $\mathscr{R}^2=\cup_i\cup_{\Delta x_{ij}}\mathscr{R}^2_{\Delta x_{ij}}(\bar{s}_i)$, and $\mathscr{R}^3=\cup_i\cup_{(n,\bar{n})}\mathscr{R}^3_{n,\bar{n}}(s_i,\bar{s}_i)$.

Algorithm 1. *The noise-control algorithm.*

Supplementary material for Noise control for molecular computing

Plesa et al., Journal of the Royal Society Interface

Notation. Set $\mathbb{R}$ is the space of real numbers, $\mathbb{R}_{\geq}$ the space of nonnegative real numbers, and $\mathbb{R}_{>}$ the space of positive real numbers. Similarly, $\mathbb{Z}$ is the space of integer numbers, $\mathbb{Z}_{\geq}$ the space of nonnegative integer numbers, and $\mathbb{Z}_{>}$ the space of positive integer numbers. Given two real numbers $a,b\in\mathbb{R}$, open interval $\{x\in\mathbb{R}|a<x<b\}$ is denoted by (a,b), while closed interval $\{x\in\mathbb{R}|a\leq x\leq b\}$ by $[a,b]$. Given sets $\mathscr{R}_1$ and $\mathscr{R}_2$, their union is denoted by $\mathscr{R}_1\cup\mathscr{R}_2$, while their intersection by $\mathscr{R}_1\cap\mathscr{R}_2$. *Support* of function $f:\mathbb{Z}\to\mathbb{R}$ is defined by $\text{supp}(f)=\{x\in\mathbb{Z}|f(x)\neq 0\}$.

1. Dynamical models of chemical reaction networks

Let us consider the mass-action reaction network $\mathscr{R}$ given by

$$\mathscr{R}(s_1,\ldots,s_N):\ \sum_{i=1}^{N}c_{ij}s_i\xrightarrow{k_j}\sum_{i=1}^{N}c'_{ij}s_i,\ \ j\in\{1,\ldots,M\}, \tag{S1}$$

where $s_1,\ldots,s_N$ are the reacting species, k_j the reaction rate coefficients, and c_{ij},c'_{ij} the stoichiometric coefficients. Let us denote by $\mathbf{c}_j,\mathbf{c}'_j\in\mathbb{N}_0^N$ the vectors of the stoichiometric coefficients of reaction j, and let $\Delta\mathbf{x}_j=\mathbf{c}'_j-\mathbf{c}_j$.

The *deterministic model* of reaction network (S1) is given by the following system of ordinary-differential equations (ODEs), known as the reaction-rate equations [1]:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}=\sum_{j=1}^{M}k_j\mathbf{x}^{\mathbf{c}_j}\Delta\mathbf{x}_j,\quad i\in\{1,\ldots,N\}. \tag{S2}$$

Here, $\mathbf{x}=\mathbf{x}(t)\in\mathbb{R}_{\geq}^N$ is the vector of species concentrations, i.e. $x_i(t)$ is the concentration of species s_i at time t, and $\mathbf{x}^{\mathbf{c}_j}=\prod_{l=1}^{N}x_l^{c_{lj}}$, with the convention $0^0=1$.

The *stochastic model* of reaction network (S1) is given by the following system of partial difference-differential equations, known as the chemical master equation (CME) [1, 2]:

$$\frac{\partial}{\partial t}p(\mathbf{x},t)=\mathscr{L}p(\mathbf{x},t)=\sum_j(E_{\mathbf{x}}^{-\Delta\mathbf{x}_j}-1)\big(\alpha_j(\mathbf{x})p(\mathbf{x},t)\big). \tag{S3}$$

Here, $p(\mathbf{x},t)$ is the probability mass function (PMF), i.e. the probability that the vector of copy-numbers $\mathbf{X}=\mathbf{X}(t)\in\mathbb{Z}_{\geq}^N$ of species $s_1,\ldots,s_N$ at time t is given by $\mathbf{x}$. Linear operator $\mathscr{L}$ is called the forward operator, and step operator $E_{\mathbf{x}}^{-\Delta\mathbf{x}_j}$ is such that $E_{\mathbf{x}}^{-\Delta\mathbf{x}_j}p(\mathbf{x},t)=p(\mathbf{x}-\Delta\mathbf{x}_j,t)$. Function $\alpha_j(\mathbf{x})$ is the propensity function [1] of the j-th reaction from (S1), and is given by

$$\alpha_j(\mathbf{x})=k_j\mathbf{x}^{\underline{\mathbf{c}_j}}=k_j\prod_{l=1}^{N}x_l^{\underline{c_{lj}}}, \tag{S4}$$

where $x_l^{\underline{c_{lj}}}$ denotes a falling factorial of x_l, i.e. $x_l^{\underline{c_{lj}}}=x_l(x_l-1)\ldots(x_l-c_{lj}+1)$, with the convention $x^{\underline{0}}=1$ for all $x\in\mathbb{Z}_{\geq}$.

2. Analysis of dynamical models of network $\hat{\mathscr{R}}^1\cup\mathscr{R}_1^2\cup\mathscr{R}_{1,1}^3$

In what follows, we analyse the deterministic and stochastic models of the output reaction network $\hat{\mathscr{R}}^1\cup\mathscr{R}_1^2\cup\mathscr{R}_{1,1}^3$, consisting of the subnetworks given by (3) and (5) in the paper, in the asymptotic limit $\mu\to 0$ (when the drift-corrector network $\mathscr{R}_1^2$ fires infinitely fast). We also analyse the stochastic model in the asymptotic limits $K_{1,1}\to 0$, and $K_{1,1}\to\infty$ (when the zero-drift network $\mathscr{R}_{1,1}^3$ does not fire, and when it fires infinitely fast, respectively). The obtained limiting dynamics of the output network are compared with the dynamics of the input network $\hat{\mathscr{R}}$, given by (1) in the paper.

2.1 The deterministic model in the limit $\mu\to 0$

Let us analyse equation (4), from the paper, in the asymptotic limit $\mu\to 0$. It follows from the Tikhonov theorem [3] that the ODE for y, given by second equation in (4), reduces to the algebraic equation $y=(c-x)^{-1}$ as $\mu\to 0$. Substituting the algebraic equation into (4) results in

$$\begin{aligned}&\frac{\mathrm{d}x}{\mathrm{d}t}=k_1-k_2x,\\ &x(0)=x_0,\ \text{as}\ \mu\to 0.\end{aligned} \tag{S5}$$

Initial value problems (2), from the paper, and (S5) have the same form, and let us denote their solutions by $\hat{x}(t;\hat{x}_0)$ and $x(t;x_0)$, respectively. Then, choosing the conservation constant $c\geq\max_{t\geq 0}\hat{x}(t;\hat{x}_0)<\infty$, and $x_0=\hat{x}_0$, ensures that the concentration of auxiliary species $\bar{s}$ is nonnegative, $\bar{x}(t)=c-x(t)\geq 0$, and that the solutions of (2) and (4) are asymptotically equivalent in the limit $\mu\to 0$.

2.2 The stochastic model in the limit $\mu\to 0$

CME induced by network $\hat{\mathscr{R}}^1\cup\mathscr{R}_1^2\cup\mathscr{R}_{1,1}^3$ is given by

$$\frac{\partial}{\partial t}p(x,y,t)=\left(\mathscr{L}^1+\frac{1}{\mu}\mathscr{L}_1^2+K_{1,1}\mathscr{L}_{1,1}^3\right)p(x,y,t), \tag{S6}$$

where $x(t),y(t)\in\mathbb{Z}_{\geq}$ are copy-numbers of species s,I^1 from (3), respectively, with

$$\begin{aligned}\mathscr{L}^1&=k_1(E_x^{-1}-1)\left((C-x)y\right)+k_2(E_x^{+1}-1)x,\\ \mathscr{L}_1^2&=(E_y^{-1}-1)+(C-x)(E_y^{+1}-1)y,\\ \mathscr{L}_{1,1}^3&=(E_x^{-1}+E_x^{+1}-2)\beta_{1,1}(x),\end{aligned} \tag{S7}$$

and $K_{1,1},\beta_{1,1}(x)$ given in equation (8), in the paper. Operators $\mathscr{L}^1,\mathscr{L}_1^2,\mathscr{L}_{1,1}^3$ are induced by subnetworks $\hat{\mathscr{R}}^1,\mathscr{R}_1^2,\mathscr{R}_{1,1}^3$, respectively.

Let us analyse system (S6) in the limit $\mu \to 0$, and consider the following perturbation series:

$$p(x,y,t) = p_0(x,y,t) + \mu p_1(x,y,t) + \dots + \mu^i p_i(x,y,t) + \dots, \tag{S8}$$

with $i \geq 2$. Substituting (S8) into (S6), equating terms of equal powers in μ, and writing $p_0 = p_0(x,y,t)$, $p_1 = p_1(x,y,t)$, the following system of equations is obtained:

$$\begin{aligned} &\mathscr{O}\left(\frac{1}{\mu}\right):\ \mathscr{L}_1^2 p_0 = 0, \\ &\mathscr{O}(1):\ -\mathscr{L}_1^2 p_1 = (\mathscr{L}^1 + K_{1,1}\mathscr{L}_{1,1}^3 - \frac{\partial}{\partial t})p_0. \end{aligned} \tag{S9}$$

Order $1/\mu$ *equation*. Let us write $p_0(x,y,t) = p_0(y|x)p_0(x,t)$, where $p_0(y|x)$ is the stationary PMF of y conditional on x, while $p_0(x,t)$ is the marginal PMF of x. Substituting $p_0(x,y,t) = p_0(y|x)p_0(x,t)$ into the first equation in (S9), with t and x fixed, leads to $-\mathscr{L}_1^2 p_0(y|x) = 0$. It follows that $p_0(y|x)$ is the Poisson distribution with parameter $(C-x)^{-1}$, so that the zero-order PMF is given by

$$p_0(x,y,t) = \left(\frac{1}{y!}\left(\frac{1}{(C-x)}\right)^y \exp\left(-\frac{1}{(C-x)}\right)\right) p_0(x,t). \tag{S10}$$

Order 1 *equation*. Substituting (S10) into the second equation in (S9), summing over all the possible states $y \in \mathbb{Z}_{\geq}$, using (S7), and equalities $\sum_y y p_0(y|x) = (C-x)^{-1}$ and $\sum_y p_0(y|x) = 1$, one obtains the *effective CME*, given by

$$\frac{\partial}{\partial t} p_0(x,t) = \left(\mathscr{L} + K_{1,1}\mathscr{L}_{1,1}^3\right) p_0(x,t), \tag{S11}$$

where $\mathscr{L}$ is the forward operator corresponding to network (1), and has the following form

$$\mathscr{L} = k_1(E_x^{-1} - 1) + k_2(E_x^{+1} - 1)x. \tag{S12}$$

2.2.1 Limit $K_{1,1} \to 0$

Setting the left-hand side to zero, and taking $K_{1,1} = 0$, in (S11), and assuming C is fixed to a sufficiently large value, it follows that the stationary PMF is a Poisson distribution with parameter k_1/k_2 [2]:

$$p_0(x) = \begin{cases} \frac{1}{x!}\left(\frac{k_1}{k_2}\right)^x \exp\left(-\frac{k_1}{k_2}\right), & \text{if } x \in [0,C], \\ 0, & \text{otherwise.} \end{cases} \tag{S13}$$

2.2.2 Limit $K_{1,1} \to \infty$

Let us substitute the perturbation series

$$p_0(x) = f_0(x) + \frac{1}{K_{1,1}} f_1(x) + \dots + \left(\frac{1}{K_{1,1}}\right)^i f_i(x) + \dots, \tag{S14}$$

with $i \geq 2$, into (S11) with the left-hand side set to zero, and consider the limit $K_{1,1} \to \infty$. Then, equating terms of equal powers in $1/K_{1,1}$, one obtains:

$$\begin{aligned} &\mathscr{O}(K_{1,1}):\ \mathscr{L}_{1,1}^3 f_0(x) = 0, \\ &\mathscr{O}(1):\ -\mathscr{L}_{1,1}^3 f_1(x) = \mathscr{L} f_0(x). \end{aligned} \tag{S15}$$

Order $K_{1,1}$ *equation*. The solution to the first equation in (S15) is given by

$$f_0(x) = \begin{cases} 1 - \frac{a}{C}, & \text{if } x = 0, \\ \frac{a}{C}, & \text{if } x = C, \\ 0, & \text{otherwise,} \end{cases} \tag{S16}$$

where $a \in \mathbb{R}_{\geq}$ is an arbitrary constant.

Order 1 *equation*. Multiplying the second equation in (S15) by x, and summing over $x \in \mathbb{Z}_{\geq}$, with the convention that $f_0(x) = 0$ and $\beta_{1,1}(x) = 0$ for $x \notin [0,C]$, one obtains the solvability condition $0 = \sum_{x=0}^{\infty} x \mathscr{L} f_0(x)$, which implies $a = k_1/k_2$. Substituting a into (S16) leads to the zero-order approximation of the stationary PMF:

$$f_0(x) = \begin{cases} 1 - \frac{1}{C}\frac{k_1}{k_2}, & \text{if } x = 0, \\ \frac{1}{C}\frac{k_1}{k_2}, & \text{if } x = C, \\ 0, & \text{otherwise.} \end{cases} \tag{S17}$$

3. Zero-drift networks $\mathscr{R}_{n,\bar{n}}^3$

The propensity function of reactions underlying $\mathscr{R}_{n,\bar{n}}^3(s,\bar{s})$, $n,\bar{n} \in \mathbb{Z}_{\geq}$, and $(n+\bar{n}) \leq C$, is given by $K_{n,\bar{n}}\beta_{n,\bar{n}} : [0,C] \to \mathbb{R}_{\geq}$, with

$$K_{n,\bar{n}} = M_{n,\bar{n}} k_{n,\bar{n}}, \tag{S18}$$

and

$$\beta_{n,\bar{n}}(x) = (M_{n,\bar{n}})^{-1} x^{\underline{n}} (C-x)^{\underline{\bar{n}}}, \tag{S19}$$

where the scaling factor $M_{n,\bar{n}}$ is introduced to approximately normalize $\beta_{n,\bar{n}}(x)$, and is given by

$$M_{n,\bar{n}} = \left(\frac{n}{n+\bar{n}}C\right)^{\underline{n}} \left(\frac{\bar{n}}{n+\bar{n}}C\right)^{\underline{\bar{n}}}. \tag{S20}$$

Function $\beta_{n,\bar{n}}(x)$ is nonzero on the interval $[n, C-\bar{n}]$, with the unique maximum approximately at $Cn/(n+\bar{n})$.

Interior zero-drift networks. Zero-drift network $\mathscr{R}_{n,\bar{n}}^3(s,\bar{s})$, with $n,\bar{n} \neq 0$, satisfies equation (15) in the paper, and the propensity function of its reactions, which is proportional to (S19), is nonzero only in the interior of the state-space. Since the propensity function of $\mathscr{R}_{n,\bar{n}}^3(s,\bar{s})$, with $n,\bar{n} \neq 0$, attains its maximum in the interior of the domain, we call the network an interior zero-drift network.

Boundary zero-drift networks. Network $\mathscr{R}_{0,\bar{n}}^3(s,\bar{s})$, satisfying equation (16) in the paper, is a zero-drift network in the limit $\mu_{0,\bar{n}} \to 0$. Furthermore, in the same limit, the

first two reactions from (16) have the same propensity function, which is proportional to (S19) with $n=0$, and which is nonzero at the left boundary point, $x=0$. Similarly, network $\mathscr{R}^3_{n,0}=\mathscr{R}^3_{0,n}(\bar{s},s;\bar{B},k_{n,0},\mu_{n,0})$ is a zero-drift network as $\mu_{n,0}\to 0$, and its first two reactions have the same propensity function, which is nonzero at the right boundary point, $x=C$. Since networks with $n=0$ (respectively, $\bar{n}=0$) generate propensity functions with the maximum values at the left (respectively, right) boundary point, we call such networks left (respectively, right) boundary zero-drift networks.

Basis zero-drift networks. Stoichiometric coefficients $n,\bar{n}$ control the support of the intrinsic noise, which network $\mathscr{R}^3_{n,\bar{n}}$ introduces into the stochastic dynamics, via the control of support of function (S19). The larger the sum $(n+\bar{n})$ is, with $(n+\bar{n})\leq C$, the smaller the support of (S19), and hence one obtains a more precise noise-control. In the special case when $n+\bar{n}=C$, the propensity function (S19) is nonzero only at a single point in the state-space, $x=n$. We call networks $\mathscr{R}^3_{n,\bar{n}}(s,\bar{s})$, with $n+\bar{n}=C$, basis zero-drift networks, and the corresponding propensity functions *basis propensity functions*. Any nonnegative function, defined on a bounded discrete domain, may be represented by a suitable linear combination of the basis propensity functions.

4. Stochastic DNA Compiler

In this section, we analyze the 4-domain DNA compiler from [4], which has been shown to preserve the deterministic dynamics when mapping suitable abstract networks to the DNA-based ones. We show that the compiler also preserves the stochastic dynamics. Furthermore, we apply the compiler to a reaction network designed using the noise-control algorithm, and briefly outline the pre-compiling step: how to approximate higher-order reactions with up-to second order ones. Let us note that we change the notation slightly in this section, in order to match it with the notation from [4].

First-order reactions. Consider an arbitrary input first-order reaction

$$X_m \xrightarrow{k} \sum_{i=1}^{N} c_i X_i, \tag{S21}$$

involving abstract species $X_1,X_2,\ldots,X_N$, and the following output DNA-based second-order network

$$\begin{aligned} X_m+G &\xrightarrow{\varepsilon k} O,\\ O+T &\xrightarrow{q} \sum_{i=1}^{N} c_i X_i, \end{aligned} \tag{S22}$$

where $X_1,X_2,\ldots,X_N$ represent target single-stranded DNA molecules, O is an intermediate single-stranded DNA, and G and T are double-stranded DNAs called a gate and a translator, respectively [4]. Here, we assume rate coefficients k and q are of order one, $k,q=\mathscr{O}(1)$, with respect to the asymptotic parameter $0<\varepsilon\ll 1$. Furthermore, we assume that the copy-numbers of the reservoir species G and T are initially given by $G(0)=T(0)=1/\varepsilon\equiv N_{\max}\in\mathbb{Z}_{>}$, where we abuse the notation slightly by denoting identically the species and their corresponding copy-numbers. It has been shown that reaction (S21) and network (S22) are approximately deterministically identical over suitable time-intervals for sufficiently small ε. We now show that this is also true at the stochastic level.

To this end, let us firstly assume ε is chosen small enough, so that species G and T remain approximately fixed over a desired time-interval $[0,\Omega]$, $\Omega\geq 0$, i.e. $G(t)\approx 1/\varepsilon$ and $T(t)\approx 1/\varepsilon$ for $t\in[0,\Omega]$, and so that network (S22) may then be approximated by

$$\begin{aligned} X_m &\xrightarrow{k} O,\\ O &\xrightarrow{q/\varepsilon} \sum_{i=1}^{N} c_i X_i. \end{aligned} \tag{S23}$$

Denoting the copy-numbers of species X_i and O by x_i and o, respectively, and defining new coordinates $\bar{x}_i=(x_i+c_i o)$ for $i\in\{1,2,\ldots,N\}$, the CME induced by (S23) reads as

$$\frac{\partial}{\partial t}p(\bar{\mathbf{x}},o,t)=\left(\frac{1}{\varepsilon}\mathscr{L}_0+\mathscr{L}_1\right)p(\bar{\mathbf{x}},o,t), \tag{S24}$$

where

$$\begin{aligned} \mathscr{L}_0&=(E_o^{+1}-1)qo,\\ \mathscr{L}_1&=(E_{\bar{\mathbf{x}}}^{-\Delta\bar{\mathbf{x}}}E_o^{-1}-1)k(\bar{x}_m-c_m o), \end{aligned}$$

and with $\Delta\bar{x}_i=c_i$ for $i\neq m$, and $\Delta\bar{x}_m=c_m-1$. Substituting the perturbation series $p(\bar{\mathbf{x}},o,t)=p_0(\bar{\mathbf{x}},o,t)+\varepsilon p_1(\bar{\mathbf{x}},o,t)+\ldots$ into (S24), one obtains the hierarchy of equations given by

$$\begin{aligned} &\mathscr{O}\left(\frac{1}{\varepsilon}\right):\ \mathscr{L}_0 p_0(\bar{\mathbf{x}},o,t)=0,\\ &\mathscr{O}(1):\ -\mathscr{L}_0 p_1(\bar{\mathbf{x}},o,t)=(\mathscr{L}_1-\frac{\partial}{\partial t})p_0(\bar{\mathbf{x}},o,t). \end{aligned} \tag{S25}$$

The solution to the first equation from (S25) is given by $p_0(\bar{\mathbf{x}},o,t)=p_0(\bar{\mathbf{x}},t)\delta_{o,0}$, where $\delta_{i,j}$ is the Kronecker-delta function ($\delta_{j,j}=1$, and $\delta_{i,j}=0$ for $i\neq j$). Thus, it follows that o converges to zero in probability, implying that $\bar{\mathbf{x}}\to\mathbf{x}$. Summing the second equation from (S25) over variable o, one finally obtains the effective CME: $\frac{\partial}{\partial t}p_0(\mathbf{x},t)=(E_{\mathbf{x}}^{-\Delta\mathbf{x}}-1)kx_m$. Since this matches the CME of the input reaction (S21), we have established that (S21) and (S22) match stochastically in a weak sense.

Second-order reactions. Consider now an arbitrary input second-order reaction

$$X_m+X_n \xrightarrow{k} \sum_{i=1}^{N} c_i X_i, \tag{S26}$$

and the following output DNA-based second-order network

$$\begin{aligned} X_m + L &\underset{q}{\overset{k}{\rightleftharpoons}} H + B, \\ X_n + H &\xrightarrow{q} O, \\ O + T &\xrightarrow{q} \sum_{i=1}^{N} c_i X_i, \end{aligned} \tag{S27}$$

where $X_1, X_2, \ldots, X_N$ represent target single-stranded DNA molecules, B and O are auxiliary single-stranded DNAs, while L, H and T are auxiliary double-stranded DNAs [4]. Here, $\rightleftharpoons$ denotes a reversible reaction (which is equivalent to two irreversible reactions). We assume that the copy-numbers of the reservoir species L, B and T are initially given by $L(0) = B(0) = T(0) = 1/\varepsilon = N_{\max} \in \mathbb{Z}_{>}$. Furthermore, for simplicity, let us also assume that the rate coefficient q is chosen to be significantly larger than k, i.e. $k = \mu q$, with $0 < \mu \ll 1$. Such an assumption may be relaxed by appropriately rescaling the rate coefficients appearing in (S27) [4]. We now show that reaction (S26) and network (S27) are approximately dynamically identical over suitable time-intervals at the stochastic level for sufficiently small ε, and $\mu/\varepsilon \equiv \lambda = \mathcal{O}(1)$.

Assume ε is chosen small enough, so that $L(t) \approx 1/\varepsilon$, $B(t) \approx 1/\varepsilon$, $T(t) \approx 1/\varepsilon$ over the desired time-interval $t \in [0, \Omega]$, and network (S27) is then approximately given by

$$\begin{aligned} X_m &\underset{q/\varepsilon}{\overset{\lambda q}{\rightleftharpoons}} H, \\ X_n + H &\xrightarrow{q} O, \\ O &\xrightarrow{q/\varepsilon} \sum_{i=1}^{N} c_i X_i. \end{aligned} \tag{S28}$$

Let us denote the copy-numbers of species X_i, H and O by x_i, h and o, respectively, and assume that species X_m and X_n are distinct, i.e. that $m \neq n$. Defining new coordinates $\bar{x}_i = x_i + c_i o$ for $i \neq m$, and $\bar{x}_m = x_m + c_m o + h$, and a slow time-scale $\tau = \mathcal{O}(1)$ by $\tau = \varepsilon t$, and substituting into the resulting CME the perturbation series $p(\bar{\mathbf{x}}, h, o, \tau) = p_0(\bar{\mathbf{x}}, h, o, \tau) + \varepsilon p_1(\bar{\mathbf{x}}, h, o, \tau) + \varepsilon^2 p_2(\bar{\mathbf{x}}, h, o, \tau) + \ldots$, one obtains the following system of equations

$$\begin{aligned} \mathcal{O}\left(\frac{1}{\varepsilon^2}\right) &: \ \mathscr{L}_0 p_0(\bar{\mathbf{x}}, h, o, \tau) = 0, \\ \mathcal{O}\left(\frac{1}{\varepsilon}\right) &: \ -\mathscr{L}_0 p_1(\bar{\mathbf{x}}, h, o, \tau) = \mathscr{L}_1 p_0(\bar{\mathbf{x}}, h, o, \tau), \\ \mathcal{O}(1) &: \ -\mathscr{L}_0 p_2(\bar{\mathbf{x}}, h, o, \tau) = \mathscr{L}_1 p_1(\bar{\mathbf{x}}, h, o, \tau) \\ &\qquad - \frac{\partial}{\partial \tau} p_0(\bar{\mathbf{x}}, h, o, \tau). \end{aligned} \tag{S29}$$

Here, operators $\mathscr{L}_0$ and $\mathscr{L}_1$ are given by

$$\begin{aligned} \mathscr{L}_0 &= (E_h^{+1} - 1) q h + (E_o^{+1} - 1) q o, \\ \mathscr{L}_1 &= (E_h^{-1} - 1) \lambda q (\bar{x}_m - c_m o - h) \\ &\quad + (E_{\bar{\mathbf{x}}}^{-\Delta\bar{\mathbf{x}}} E_h^{+1} E_o^{-1} - 1) q (\bar{x}_n - c_n o) h, \end{aligned}$$

where $\Delta\bar{x}_i = c_i$ for $i \notin \{n, m\}$ and $\Delta\bar{x}_i = c_i - 1$ for $i \in \{n, m\}$.

The solution to the first equation from (S29) is given by $p_0(\bar{\mathbf{x}}, h, o, \tau) = p_0(\bar{\mathbf{x}}, \tau)\delta_{h,0}\delta_{o,0}$. A solution to the second equation is given by $p_1(\bar{\mathbf{x}}, h, o, \tau) = p_0(\bar{\mathbf{x}}, \tau)\delta_{o,0} p_1(h)$, where the third factor satisfies $h p_1(h) = \lambda E_h^{-1}(\bar{x}_m - h)\delta_{h,0}$. Let us note that o and h approach zero as $\varepsilon \to 0$, implying that $\bar{\mathbf{x}} \to \mathbf{x}$. Finally, substituting the obtained solutions into the third equation from (S29), and summing over variables h and o, one obtains the effective CME $\frac{\partial}{\partial t} p_0(\mathbf{x}, t) = (E_{\mathbf{x}}^{-\Delta\mathbf{x}} - 1) k x_m x_n p_0(\mathbf{x}, t)$, which is identical to the CME of the input reaction (S26) with $m \neq n$. Applying similar reasoning when $m = n$ also leads to the correct effective CME: $\frac{\partial}{\partial t} p_0(\mathbf{x}, t) = (E_{\mathbf{x}}^{-\Delta\mathbf{x}} - 1) k x_m (x_m - 1) p_0(\mathbf{x}, t)$, with $\Delta x_i = c_i$ for $i \neq m$, and $\Delta x_m = c_m - 2$.

4.1 An Example

Let us now apply the molecular compiler on the reaction network

$$\begin{aligned} \mathscr{R}^1 : \quad & \bar{s} \xrightarrow{k_1} s, \\ & s \xrightarrow{k_2} \bar{s}, \\ \mathscr{R}^3_{1,1} : \quad & s + \bar{s} \xrightarrow{k_{1,1}} 2s, \\ & \bar{s} + s \xrightarrow{k_{1,1}} 2\bar{s}. \end{aligned} \tag{S30}$$

Network (S30) has been obtained by applying Algorithm 1 from the paper on the conservative input network $\bar{s} \xrightarrow{k_1} s$, $s \xrightarrow{k_2} \bar{s}$, so that the second step of the algorithm may be omitted. It is similar to the network consisting of subnetworks given by (3) and (5) in the paper, and is chosen here for simplicity. Replacing s by X_1, and $\bar{s}$ by X_2, to match the notation from [4], a DNA-based implementation of (S30) is given by

$$\begin{aligned} \bar{\mathscr{R}}^1 : \quad & X_2 + G_1 \xrightarrow{k_1/N_{\max}} O_1, \\ & O_1 + T_1 \xrightarrow{q} X_1, \\ & X_1 + G_2 \xrightarrow{k_2/N_{\max}} O_2, \\ & O_2 + T_2 \xrightarrow{q} X_2, \\ \bar{\mathscr{R}}^3_{1,1} : \quad & X_1 + L_3 \underset{q}{\overset{k_{1,1}}{\rightleftharpoons}} H_3 + B_3, \\ & X_2 + H_3 \xrightarrow{q} O_3, \\ & O_3 + T_3 \xrightarrow{q} 2X_1, \\ & X_2 + L_4 \underset{q}{\overset{k_{1,1}}{\rightleftharpoons}} H_4 + B_4, \\ & X_1 + H_4 \xrightarrow{q} O_4, \\ & O_4 + T_4 \xrightarrow{q} 2X_2, \end{aligned} \tag{S31}$$

where we require that the initial copy-numbers of the auxiliary species G_i, T_i, for $i = 1, \ldots, 4$, and L_i, B_i, for $i = 1, 2$, are all initially set to a sufficiently large value $N_{max} \in \mathbb{Z}_{>}$, and that q

is set to a sufficiently large value, ensuring $k_{1,1} \ll q$ (note that this condition may be relaxed [4]).

If, instead of the zero-drift network $\mathscr{R}^3_{1,1}$, we embed e.g. the zero-drift network $\mathscr{R}^3_{2,2}$ into network $\mathscr{R}^1$ from (S30), which is given by

$$\begin{aligned} \mathscr{R}^3_{2,2}: \qquad & 2s+2\bar{s} \xrightarrow{k_{2,2}} 3s+\bar{s}, \\ & 2s+2\bar{s} \xrightarrow{k_{2,2}} s+3\bar{s}, \end{aligned} \tag{S32}$$

then a pre-compilation step is performed. In particular, we first approximate (S32) by a set of up-to bimolecular reactions e.g. as follows [5, 6]:

$$\begin{aligned} \bar{\mathscr{R}}^3_{2,2}: \qquad & s+\bar{s} \underset{1/\varepsilon}{\overset{K_1}{\rightleftharpoons}} Q_1, \\ & s+Q_1 \underset{1/\varepsilon}{\overset{K_2}{\rightleftharpoons}} Q_2, \\ & \bar{s}+Q_2 \xrightarrow{K_3} s+Q_2, \\ & \bar{s}+Q_2 \xrightarrow{K_3} 2\bar{s}+Q_1, \end{aligned} \tag{S33}$$

where Q_1 and Q_2 are auxiliary species. Provided $\varepsilon^2 K_1 K_2 K_3 = k_{2,2}$, $\varepsilon K_i \ll 1$, for $i = 1,2,3$, and $0 < \varepsilon \ll 1$, networks (S32) and (S33) are approximately dynamically identical. As opposed to the original fourth-order network (S32), the approximating second-order network (S33) can be mapped to a DNA-based one using the compiler from [4].

In this section, we have verified one way to structurally compile an input abstract network, involving low molecular copy-numbers, into a DNA-based output network. Provided the target species copy-numbers are sufficiently low, so that the noise introduced by the zero-drift networks is significant without requiring too large rate coefficients, it may be achievable to constrain all the rate coefficients within six orders of magnitude. Thus, in principle, low copy-number networks designed using the noise-control algorithm are experimentally realizable.

5. Network $\tilde{\mathscr{R}}$

The deterministic model of reaction network $\tilde{\mathscr{R}}$, given by equation (11) in the paper, is given by

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= k_1 + k_2 x_1 + k_3 x_1^2 - k_4 x_1 x_2 - k_5 x_1^2 x_2 + k_6 x_1 x_2^2, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= k_7 - k_8 x_2 + k_9 x_1 x_2 + k_{10} x_2^2 - k_{11} x_2^3, \end{aligned} \tag{S34}$$

where $x_1 = x_1(t), x_2 = x_2(t)$ are the concentrations of species s_1, s_2, respectively, at time t.

5.1 An application of the noise-control algorithm

Network $\tilde{\mathscr{R}}^1(s_1,s_2,\bar{s}_2) \cup \mathscr{R}^2_1(\bar{s}_2) \cup (\mathscr{R}^3_{0,C_2-10}(s_2,\bar{s}_2) \cup \mathscr{R}^3_{30,0}(s_2,\bar{s}_2))$, arising from an application of the noise-control algorithm on network $\tilde{\mathscr{R}}$, is given by

$$\begin{aligned} \tilde{\mathscr{R}}^1(s_1,s_2,\bar{s}_2): \qquad & \varnothing \xrightarrow{k_1} s_1, \\ & s_1 \xrightarrow{k_2} 2s_1, \\ & 2s_1 \xrightarrow{k_3} 3s_1, \\ & s_1+s_2 \xrightarrow{k_4} s_2, \\ & 2s_1+s_2 \xrightarrow{k_5} s_1+s_2, \\ & s_1+2s_2 \xrightarrow{k_6} 2s_1+2s_2, \\ & \bar{s}_2+I_2^1 \xrightarrow{k_7} s_2+I_2^1, \\ & s_2 \xrightarrow{k_8} \bar{s}_2, \\ & s_1+s_2+\bar{s}_2+I_2^1 \xrightarrow{k_9} s_1+2s_2+I_2^1, \\ & 2s_2+\bar{s}_2+I_2^1 \xrightarrow{k_{10}} 3s_2+I_2^1, \\ & 3s_2 \xrightarrow{k_{11}} 2s_2+\bar{s}_2, \\ \mathscr{R}^2_1(\bar{s}_2): \qquad & \varnothing \xrightarrow{1/\mu} I_2^1, \\ & \bar{s}_2+I_2^1 \xrightarrow{1/\mu} \bar{s}_2, \\ \mathscr{R}^3_{0,C_2-10}(s_2,\bar{s}_2): \qquad & (C_2-10)\bar{s}_2 \xrightarrow{k^2_{0,C_2-10}} s_2+(C_2-11)\bar{s}_2, \\ & C_2 s_2+B_2 \xrightarrow{k^2_{0,C_2-10}} (C_2-1)s_2+\bar{s}_2+B_2, \\ & (C_2-10)\bar{s}_2 \xrightarrow{1/\mu_{0,C_2-10}} (C_2-10)\bar{s}_2+B_2, \\ & C_2 s_2+B_2 \xrightarrow{1/\mu_{0,C_2-10}} C_2 s_2, \\ \mathscr{R}^3_{30,0}(s_2,\bar{s}_2): \qquad & 30s_2 \xrightarrow{k^2_{30,0}} 29s_2+\bar{s}_2, \\ & C_2\bar{s}_2+\bar{B}_2 \xrightarrow{k^2_{30,0}} s_2+(C_2-1)\bar{s}_2+\bar{B}_2, \\ & 30s_2 \xrightarrow{1/\mu_{30,0}} 30s_2+\bar{B}_2, \\ & C_2\bar{s}_2+\bar{B}_2 \xrightarrow{1/\mu_{30,0}} C_2\bar{s}_2. \end{aligned} \tag{S35}$$

References

[1] Anderson DF, Kurtz TG. Stochastic analysis of biochemical systems. Springer; 2015.

[2] Van Kampen NG. Stochastic processes in physics and chemistry. Elsevier; 2007.

[3] Klonowski W. Simplifying principles for chemical and enzyme reaction kinetics. Biophysical Chemistry, 1983;18(3):73–87.

[4] Soloveichik D, Seeling G, Winfree E. DNA as a universal substrate for chemical kinetics. Proceedings of the National Academy of Sciences, 2010;107(12):5393–5398.

[5] Wilhelm T. Chemical systems consisting only of elementary steps - a paradigma for nonlinear behavior. Journal of Mathematical Chemistry, 2000;27:71–88.

[6] Plesa T. Stochastic approximation of high-molecular by bi-molecular reactions. In preparation; 2018.